Crowdsourcing Mental Health and Emotional Well-Being

by

Robert R. Morris

MSc. Media Arts and Sciences, MIT (2009)
A.B. Psychology, Princeton University (2003)

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Media Arts and Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2015

Author_____
Program in Media Arts and Sciences
September 12, 2014

Certified by_____
Rosalind W. Picard
Professor of Media Arts and Sciences
Thesis Supervisor

Accepted by_____
Prof. Pattie Maes
Associate Academic Head
Program in Media Arts and Sciences

Crowdsourcing Mental Health and Emotional Well-Being

by

Robert R. Morris

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
on September 12, 2014, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Media Arts and Sciences

Abstract

More than 30 million adults in the United States suffer from depression. Many more meet the diagnostic criteria for an anxiety disorder. Psychotherapies like cognitive-behavioral therapy can be effective for conditions such as anxiety and depression, but the demand for these treatments exceeds the resources available. To reach the widest possible audience, mental health interventions need to be inexpensive, anonymous, always available, and, ideally, delivered in a way that delights and engages the user.

Towards this end, I present *Panoply*, an online intervention that administers emotion-regulatory support anytime, anywhere. In lieu of direct clinician oversight, *Panoply* coordinates support from crowd workers and unpaid volunteers, all of whom are trained on demand, as needed. *Panoply* incorporates recent advances in crowdsourcing and human computation to ensure that feedback is timely and vetted for quality. The therapeutic approach behind this system is inspired by research from the fields of emotion regulation, cognitive neuroscience, and clinical psychology, and hinges primarily on the concept of cognitive reappraisal. Crowds are recruited to help users think more flexibly and objectively about stressful events.

A three-week randomized controlled trial with 166 participants compared *Panoply* to an active control task (online expressive writing). *Panoply* conferred greater or equal benefits for nearly every therapeutic outcome measure. Statistically significant differences between the treatment and control groups were strongest when baseline depression and reappraisal scores were factored into the analyses. *Panoply* also significantly outperformed the control task on all measures of engagement (with large effect sizes observed for both behavioral and self-report measures). This dissertation offers a novel approach to computer-based psychotherapy, one that is optimized for accessibility, engagement and therapeutic efficacy.

Thesis Supervisor: Rosalind W. Picard

Title: Professor of Media Arts and Sciences

Crowdsourcing Mental Health and Emotional Well-Being

by

Robert R. Morris

Thesis Reader_____

Rosalind W. Picard
Professor of Media Arts and Sciences
Massachusetts Institute of Technology

Thesis Reader_____

Stephen M. Schueller
Research Assistant Professor, Feinberg School of Medicine
Northwestern University

Thesis Reader_____

Sepandar Kamvar
LG Career Development Professor of Media Arts and Sciences
Massachusetts Institute of Technology

Thesis Reader_____

Robert C. Miller
Associate Professor of Computer Science
Massachusetts Institute of Technology

Acknowledgements

I am extremely grateful for all the help I received during my time at MIT. This work would not have been possible without the love and support from all my family, friends, and colleagues.

First, I would like to thank my committee members for helping me bring this work to fruition. I want to thank Professor Rosalind Picard, in particular, for believing in this work and for always encouraging me to think bigger and bolder. Roz, I am so thankful for all your brilliant insights and I am eternally grateful to have had you as an adviser. Professor Stephen Schueller, thank you for your invaluable feedback and for your enthusiasm for this work. Professor Sep Kamvar, thank your creative and wonderful perspectives on this project. Professor Rob Miller, thank you for generously sharing your wonderful insights and for introducing me to all your exceptional students and colleagues.

From the crowdsourcing and collective intelligence community, I would like to thank Professor Tom Malone for providing valuable feedback on early versions of this work and for serving on my general exam committee. I would also like to thank Drs. Mira Dontcheva, Elizabeth Gerber, and Joel Brandt for all the wonderful collaborations over the years. Working with you three was a wonderful education and I'm grateful for the many ways you helped inform my work. Thanks also to Drs. Michael Bernstein and Greg Little. Thank you both for all your wonderful comments and suggestions.

From the psychology community, I would like to thank Professor James Gross for serving on my general exam committee and for providing feedback on early versions of

this work. Thanks also to Sam Moulton, Phil Enock, and Laura Ramsey for providing suggestions with respect to experimental design and statistical analyses.

I would also like to thank the members of the Affective Computing group at MIT, especially my officemates Daniel McDuff and Javier Rivera. Working alongside you both was an invaluable education. My fondest memories of MIT will always include the two of you.

Thanks also to the many wonderful undergraduate research assistants that helped shape this work over the years. In particular, I'd like to thank Blake Elias, Zoe Redstone-Rothstein, Kevin Mustelier, Rebecca Corcillo, and Veronica Newlin.

Last, but not least, I would like to express my sincere thanks to my mom, my dad, and my lovely wife Molly. I could not have asked for more love and support over the years.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"What really frightens or dismays us is not external events themselves, but the way in which we think of them." – Epictetus, Manual for Living

"It's hard to wrestle with your own consciousness, because you have no tool in this battle except your consciousness itself." – Andrew Solomon, The Noonday Demon

1.1 Motivation

Some of the most powerful software applications and services from the past decade have been powered by collective intelligence. Google, Wikipedia, and Quora, to name just a few, work by intelligently coordinating the collective behaviors of thousands, if not millions, of people on the Internet. One particularly powerful collective intelligence platform, though perhaps lesser known to the general public, is *StackOverflow* – an online question-and-answer (Q&A) site catered to software engineers and widely praised for the speed and the quality of its content (Mamykina, Manoim, Mittal, Hripcsak, &

Hartmann, 2011). This dissertation owes a great debt to *StackOverflow*. Indeed, this work may have never come to fruition without it.

Before starting this project, I had no formal computer science training. I managed to learn the basics of computer programming on my own, but my quest to build a large, data-driven application required more than a simple mastery of "for loops." I quickly found myself drowning in the arcana of back-end systems, frantically googling ungodly things like 'memcached memory-based frameworks.' Fortunately, anytime I had a question, I could always turn to *StackOverflow*. For nearly every question I had, I found that someone had already posted a smart, sensible answer on *StackOverflow*. And when my question hadn't yet been answered, all I had to do was post it to the site and wait a few minutes. Usually I'd receive a carefully crafted response within an hour.

And yet, while *StackOverflow* was great for fixing bugs in my code, it was not particularly great for fixing bugs in my thinking. For me, the biggest challenge of learning how to code was not the bugs themselves, but the way in which I thought about them. Often, when I got stuck, I would curse myself and consider my broken code as proof of my total incompetence as a computer programmer. *StackOverflow*, alas, was not designed to help me through this kind of self-doubt and intense frustration.

What is needed is a *StackOverflow* for the mind, a platform that leverages collective intelligence to help people (such as myself) reframe negative, distorted thinking. Such a system would be powerful, not just for disgruntled programmers, but for anyone battling negative, depressive thoughts. Excessive and distorted negative thinking is a hallmark of many affective disorders, including depression and anxiety. Disorders like depression and anxiety are among the leading causes of disability in the United States. According to epidemiological studies, more than 30 million adults in the United States suffer from depression (Kessler et al., 2003). Many more meet the diagnostic criteria for an anxiety disorder (Kessler, Chiu, Demler, & Walters, 2005). While inflation-adjusted income in the United States has nearly tripled since World War II, average rates of happiness appear to have remained stable. We live in a blooming, buzzing world of technological wonders and material wealth and yet, in spite of all this, millions of people are wracked with emotional distress.

Psychotherapies like cognitive-behavioral therapy (CBT) can be very effective for conditions like anxiety and depression, but the demand for these treatments exceeds the resources available. There are simply not enough clinicians. Access is further limited by cost, stigma, and the logistics of scheduling and traveling to appointments. To reach the widest possible audience, we need mental health interventions that are inexpensive, anonymous, always available, and, ideally, delivered in a way that delights and engages the user. Computer-based mental health interventions can solve many of these problems. To date, online interventions have been developed for depression and anxiety, many of which employ evidence-based techniques drawn from therapies like CBT. Unfortunately, while these systems have been shown to be successful in clinical trials, they frequently fail to sufficiently engage users outside the auspices of a controlled experiment. When these applications are openly available online, adherence can drop precipitously, down to around 1% in some studies (Christensen, Griffiths, Korten, Brittliffe, & Groves, 2004; Farvolden, Denisoff, Selby, Bagby, & Rudy, 2005). Unlike a platform like *StackOverflow,* most of these systems are not social or interactive and they are not designed to respond to the unique situations of the user. Addressing these shortcomings is the primary aim of this thesis. *The goal is to develop mental health technologies that are as engaging and personalized as person-to-person therapy without sacrificing the benefits that come from purely self-administered approaches – namely, accessibility, anonymity, and scalability.* One of the better ways to do this is to leverage state-of-the art crowdsourcing techniques drawn from the fields of social computing and human computation.

1.2 A Crowd-Powered System for Mental Health

In this dissertation, I present *Panoply*[1], an online intervention that administers emotion-regulatory support anytime, anywhere. In lieu of direct clinician oversight, *Panoply*

---

[1] The term *panoply* has multiple meanings, all of which describe the multiple aims of this project: it is both a psychological defense and an extensive array of resources (www.oed.com).

coordinates support from crowd workers who are trained on demand, as needed. The system accommodates three distinct labor pools: (1) for-pay workers from labor markets like Amazon's Mechanical Turk Service (MTurk), (2) unpaid volunteers who may not need support themselves but find pleasure in giving it to others, and (3) individuals experiencing emotional distress who are moved to help others as part of their own therapeutic process. *Panoply* incorporates recent advances in crowdsourcing and human computation to ensure that feedback is timely and vetted for quality.

The therapeutic approach behind this system is inspired by research from the fields of emotion regulation, cognitive neuroscience, and clinical psychology, and hinges primarily on the concept of cognitive reappraisal. Users are taught to think more flexibly and objectively about the stressful events that upset them. They learn these techniques experientially, in relation to their own, day-to-day problems and negative self-beliefs. They also learn by acting as respondents in the system and applying these techniques to other people. The platform is part assistive device, part training aid. While users may elect to stay in the system as crowd volunteers, the hope is that many will outgrow the assistive components of the platform.

This thesis has both technological and therapeutic aims. The technological evaluation focuses on the user experience of the system. Is it accessible? Easy to use? Fun to use? What is the usage rate? Answers to these questions will help guide the design of future crowd-powered assistive devices, both for mental health interventions and for other health applications. The therapeutic evaluation will focus on whether repeated use of the platform produces significant changes in preferred emotion regulatory strategies, depressive symptoms and positive emotions. This research will help us better understand how crowdsourced approaches might be used in the treatment and prevention of affective disorders, such as depression and anxiety.

It should be emphasized that this work is not meant to replace traditional models of psychotherapy. Rather, the proposed technology is designed to supplement traditional approaches, to help serve the needs of individuals who are currently unable or unwilling to get help from a qualified mental health professional. It could also be used to complement and reinforce traditional psychotherapeutic techniques for those already in treatment.

The remainder of this introduction provides a roadmap for the rest of the dissertation, along with a summary of the main findings and contributions.

1.3   Dissertation Outline

- The next chapter (Chapter 2) reviews the background literature that informed the design and evaluation of this dissertation. Studies from experimental, clinical and social psychology are integrated with research in human-computer interaction, particularly recent work from the fields of crowdsourcing and human computation.

- Chapter 3 highlights the design process involved in testing and prototyping early versions of the *Panoply* platform. Two user studies are described that address the feasibility of crowdsourcing reappraisals from paid crowd workers.

- Chapter 4 describes the final *Panoply* system in its entirety. It begins with an overview of the various design principles that were used throughout the development of the system. Next, I present a detailed description of how the *Panoply* system works, both from the perspective of registered site users and paid crowd workers.

- Chapters 5-7 present an evaluation of the system along three lines of inquiry. A randomized controlled trial is used to examine: (1) the therapeutic efficacy of the platform, (2) its ability to engage users, and (3) its overall performance as a system, with particular emphasis on speed, cost, and quality.

- Chapter 8 takes a closer look at some of the more revealing personal interactions that took place on the platform throughout the experimental trial. These interactions shed light on the efficacy of the platform and its potential to engage

users, as well as some of the challenging issues that can arise with systems such as these.

- Lastly, Chapter 9 concludes the dissertation with a survey of how this work could be extended in the future. Particular emphasis is placed on ways to expand the *Panoply* platform to incorporate additional therapeutic techniques.

1.4 Summary of Design
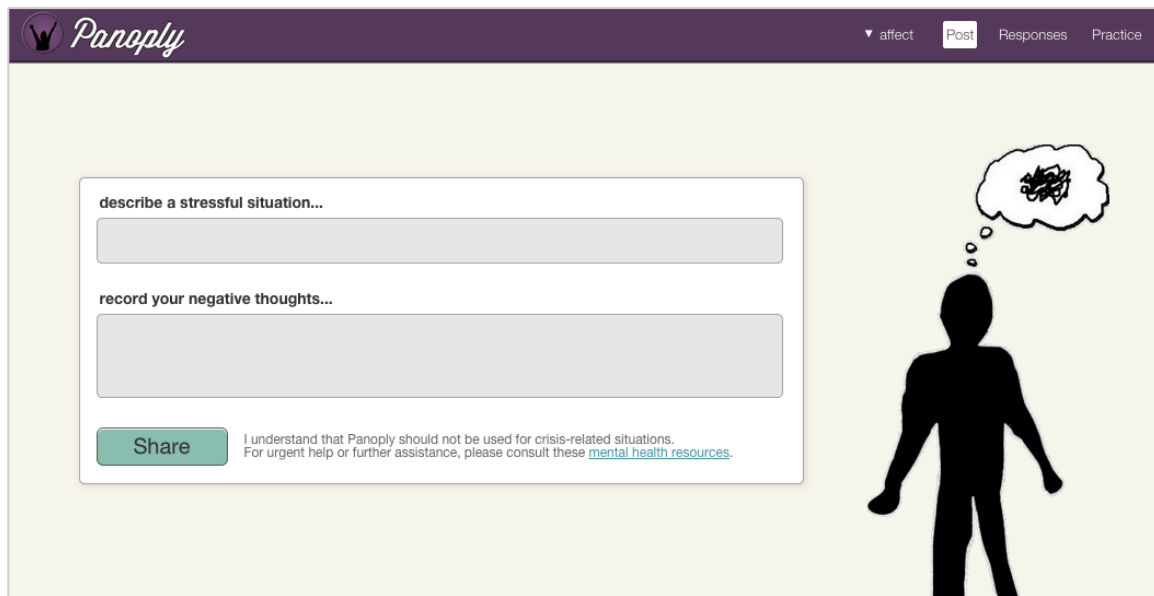
Figure 1.1: A screenshot of the *Panoply* platform, illustrating how users post content to the site.

A thorough description of the system is presented in Chapter 3, but a quick preview of how it works will provide some much needed context for the remainder of the dissertation. *Panoply* is a social system, with all the trappings you might typically find in a Q&A site. Users can post content, respond to others, and get feedback on their performance. The system distinguishes itself from other Q&A sites in that its primary
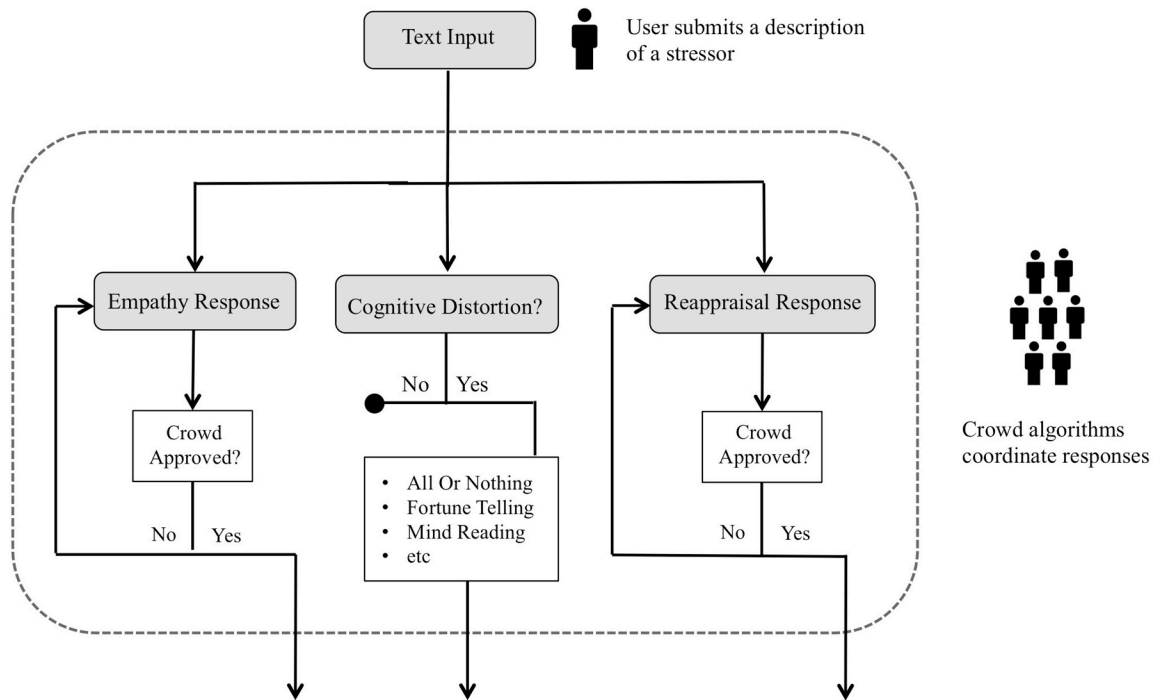
Figure 1.2: Multiple sets of crowd workers are coordinated to compose and curate responses on the system. The cognitive distortions are identified and labeled, but not subject to crowd review.

objective is to address negative thinking. It also involves considerably more crowdsourcing complexity than you might find on typical Q&A sites. Whenever someone makes a post on *Panoply*, a complex ballet of crowd work is set into motion.

First, posts are vetted for appropriateness and then they are sent to three sets of crowd workers, each of which is trained to respond using a specific evidence-based therapeutic technique. Various crowd workers are taught to (1) offer empathy, (2) identify cognitive distortions, and (3) help users reframe negative situations in ways that are more positive and adaptive (see Figure 1.2). Responses on the site are vetted by other crowd workers before being returned to the user. If a response is deemed inappropriate or abusive, it is immediately discarded. Offending respondents are placed on a 'blacklist' and are barred from contributing until further notice.

As with most Q&A sites, *Panoply* users can compose responses as well as receive them. Indeed, the crowd workers on the system are a mixture of paid workers from MTurk and unpaid volunteers who've registered on the site and want to help others. The

platform encourages site users to respond to others, not only in the service of doing a good deed, but also as a way to take advantage of a powerful learning opportunity; *Panoply* users gain expertise in techniques like cognitive restructuring and reappraisal by applying them to others as well as to themselves.

1.5   Summary of Findings

A randomized controlled trial, conducted with 166 participants over the course of two months, examined whether repeated use of the *Panoply* system confers any mental health benefits. Compared to a modified expressive writing task - an intervention that often produces positive psychological outcomes on its own   (Baikie & Wilhelm, 2005), *Panoply* yielded significantly greater changes in cognitive reappraisal – an adaptive emotion regulatory technique that is associated with well-being and lowered incidences of depression. For individuals with elevated levels of depression at baseline, *Panoply* produced greater reductions in depression symptoms and perseverative thinking. *Panoply* was especially helpful for individuals who don't typically make use of reappraisal strategies. Individuals with low reappraisal scores at baseline experienced significantly greater reductions in negative affect, depression symptoms and perseverative thinking. Across all analyses, *Panoply* outperformed the control task on every outcome measure.

The two interventions were also compared with respect to measures of engagement. Across all measures, *Panoply* proved to be significantly more engaging than the writing task. It was rated significantly higher on all self-report measures of user experience and it was also used far more regularly. By objective measures (e.g., total word count and average session duration), *Panoply* attracted more than twice as much usage as the writing task.

The system itself was fast and cost-effective. The median response time was 8 minutes and the average cost per user for three weeks of use was $3.11. It was also discovered that responses from unpaid volunteers were rated significantly higher than responses from MTurk workers.

Self-reports from study participants also point to the overall value of the *Panoply* system. Participants reported benefiting from the crowdsourced responses and they

valued the accessibility and anonymity of the system. Most users did not find the speed of the system particularly important, suggesting that future versions need not rely on paid workforces in order to get responses back as quickly as possible.

1.6 Summary of Contributions

The contributions of this dissertation span several disciplines, including clinical psychology, experimental psychology, and human-computer interaction. The main contributions are as follows:

- I present the first scalable, crowd-powered system for mental health.

- I evaluate the therapeutic merits of this system in a randomized controlled trial with 166 participants. I find that *Panoply* confers several significant psychological benefits compared to an active control intervention.

- I evaluate the system's ability to engage its users. Behavioral and self-report measures show that our system garners considerably more usage than a matched control task.

- In addition to these immediate, practical benefits, the system also provides a new research platform. It provides a controlled, but natural setting for psychologists and other researchers to study individual differences in interpersonal emotion regulation. Future studies conducted on the *Panoply* platform could also yield important new insights for the treatment of affective disorders.

- Finally, this work offers technological contributions that are of interest to the crowdsourcing community. In particular, it highlights the value of blending both paid and unpaid workforces into a single crowd-powered system. It also

exemplifies a new, more equitable model of crowdsourcing – one that provides great value to both requesters and crowd workers.

Chapter 2

Background and Related Work

This chapter begins with a review of the psychological foundations that motivate this work. In particular, I examine cognitive-based interventions for mental health from the perspectives of experimental psychology, cognitive neuroscience, and clinical psychology. I also address the mechanisms through which social support is thought to protect against depression and other mood disorders. Next, I discuss crowdsourcing and review existing work on crowd-powered assistive interfaces. I will briefly describe the state of the art in computer-based psychotherapy and online peer support tools. Lastly, I address how the proposed thesis technology fills important unmet needs in these domains.

2.1 Psychological Foundations

2.1.1 Philosophical Antecedents

For millennia, we humans have sought ways to arm ourselves against the stressors of daily life. To fight famine, we created agriculture. To vanquish disease, we developed modern medicine. Even that most pernicious of stressors – other people – has been mitigated somewhat, thanks to the development of modern nation states and judicial

systems (Pinker, 2012). And yet despite this progress and despite the relative tranquility enjoyed in many first-world western cultures, many of us remain stubbornly distressed. Unfortunately, stress will always be part of life. No matter how far technology progresses, there will always be a thousand shocks that flesh is heir to.

This is not necessarily as lamentable as it sounds, however, as one could convincingly argue that a life without stress is hardly a life at all. Rather, the difficulty arises when negative stress *chronically* overwhelms a person's natural coping skills. Excessive, chronic stress can lead to cardiovascular disease, maladaptive coping behaviors (e.g., addiction) and various affective disorders (Sapolsky, 2007).

Over the centuries, humans have developed innumerable strategies to manage stress and negative emotions. Some of the earliest recorded approaches were developed by philosophers and religious figures. One insight from these practices is to adopt a posture of acceptance and try to openly embrace the stressors that befall us. As Longfellow wrote, sometimes "the best thing one can do when it is raining is let it rain." This approach has long been integral to Eastern philosophies and is an important component of modern mindfulness practices and acceptance and commitment therapy. It rests on the notion that negative emotions thrive on struggle; they hate pacifists and will slink away at the slightest sign of open-armed acceptance. This insight has some empirical evidence, most notably from studies showing how attempts to suppress unwanted thoughts can paradoxically make them more accessible and harder to subdue (Wegner, Erber, & Zanakos, 1993).

Another approach, drawn largely from Western traditions, involves reframing the meaning of a situation in order to change one's emotional experience. This technique is based on the idea that the content of our thoughts is what shapes our emotional reactions. To quote Hamlet (one of literature's most famously introspective figures), "there is nothing either good or bad, but thinking makes it so."

Cognitive strategies such as these have their roots in ancient philosophical teachings, most notably those espoused by Roman Stoic philosophers such as Epictetus and Seneca the Younger. The stoics were keen students of emotion and they had an exceptionally nuanced grasp of the way cognitions influence feelings. In his *Meditations*, Marcus Aurelius writes, "If you are distressed by anything external, the pain is not due to the

thing itself, but to your estimate of it; and this you have the power to revoke at any moment." Like many of his stoic contemporaries, he believed that changing how we think can have a profound effect on how we feel and that cultivating this practice can have important therapeutic benefits.

2.1.2 Cognitive-Behavioral Therapy

Many centuries later, William James revisited these ideas and incorporated them into his psychological and philosophical works. Among his most well-known maxims is the notion that, "the greatest weapon against stress is our ability to choose one thought for another." Yet, despite James's influence and intellectual clout, this idea did not immediately lead to new psychotherapeutic techniques. It took almost another century for cognitive approaches to be operationalized and utilized in psychotherapeutic practice, in the form of therapies like cognitive-behavioral therapy (CBT) and Rational Emotive Behavior Therapy (REBT).

CBT uses a variety of techniques to help patients address maladaptive cognitions and behaviors. The theoretical basis for these practices rests on the notion that cognitions play a causal role in the emotional disturbances that characterize disorders such as depression and anxiety (Beck, 1979). Patients are taught to identify and question distorted thinking, in the hopes that more objective and flexible thinking will reduce negative feelings and promote adaptive coping behaviors. Behavioral approaches are also applied and can involve relaxation techniques, exposure therapy and behavioral activation. Meta-analyses of CBT have shown large effect sizes for unipolar depression, generalized anxiety disorder, posttraumatic stress disorder, and childhood depressive and anxiety disorders (Butler, Chapman, Forman, & Beck, 2006). With CBT, effect sizes for depression and anxiety are comparable to pharmacological interventions, while relapse rates are far lower (Butler et al., 2006).

2.1.3 Cognitive Reappraisal

In the past decade, considerable research attention has been given to the cognitive mechanisms that underlie therapies like CBT. Experimental psychologists in the field of emotion regulation have placed a particular emphasis on *cognitive reappraisal* − a technique that involves changing the meaning of a thought or situation to change an emotional experience. When contrasted with other emotion regulatory techniques, such as expressive suppression (i.e., stifling all outward signs of emotional expression) reappraisal techniques are thought to be particularly adaptive. Individuals who use reappraisal habitually show lower incidence of depression (Garnefski & Kraaij, 2006), better interpersonal functioning (Gross & John, 2003), and higher levels of well-being (Gross & John, 2003) than those who use other, less adaptive regulatory strategies. In the laboratory, reappraisal is associated with adaptive subjective and physiological responses to stressful situations. When faced with lab-induced stressors (e.g., unpleasant images, public speaking challenges, anger inductions), participants who are taught reappraisal techniques report less emotional distress and show reduced psychophysiological reactivity (Gross, 1998; Hofmann, Heering, Sawyer, & Asnaani, 2009; Mauss, Cook, Cheng, & Gross, 2007).

Reappraisal techniques also seem to promote adaptive response patterns in the brain. Reappraisal strategies are associated with increased activation of prefrontal regions devoted to executive functioning and cognitive control, such as the dorsolateral prefrontal cortex (Ochsner & Gross, 2005). This, in turn, appears to down-regulate limbic regions devoted to emotion processing, such as the amygdala. While the activation patterns vary somewhat across different studies, there is a consistent pattern of top-down, prefrontal modulation of emotion processing regions (Ochsner & Gross, 2005). Taken together, these studies offer compelling neuroscientific evidence to suggest that changing how you think can in fact change how you feel.

2.1.4 Peer-Supported Emotion Regulation

While intrapersonal factors like cognition and behavior are clearly important for mental health, one must not forget interpersonal factors as well, such as an individual's social support network. Indeed, there is a wealth of evidence showing a striking inverse relationship between social support and incidence of depression (Cohen & Wills, 1985). However, the precise mechanisms behind this relationship are not fully understood. Recently, researchers have begun to theorize that reappraisal may be an important mediating factor (Marroquín, 2011; Panzarella, Alloy, & Whitehouse, 2006). Reappraisal requires cognitive effort and creativity, faculties that often elude us when we are under stress. Peers, and even strangers, can lend us these resources in times of stress, and help us consider alternative appraisals of stressful or depressing situations. A particularly beautiful example of this is recounted in Viktor Frankl's "Man's Search for Meaning." He writes:

> Once, an elderly general practitioner consulted me because of his severe depression. He could not overcome the loss of his wife who had died two years before and whom he had loved above all else. Now, how could I help him? What should I tell him? Well, I refrained from telling him anything but instead confronted him with the question, "What would have happened, Doctor, if you had died first, and your wife would have had to survive you?" "Oh," he said, "for her this would have been terrible; how she would have suffered!" Whereupon I replied, "You see, Doctor, such a suffering has been spared her, and it was you who have spared her this suffering - to be sure, at the price that now you have to survive and mourn her." He said no word but shook my hand and calmly left my office. In some way, suffering ceases to be suffering at the moment it finds a meaning, such as the meaning of a sacrifice. - (Frankl, 2006)

Peer-supported reappraisal isn't typically this profound, but this example serves to illustrate what is possible. It demonstrates the way creative reinterpretations from peers can yield tremendous emotional change. Research into these kinds of interpersonal exchanges is still in its infancy, but this dissertation will help examine some of these ideas more deeply. The *Panoply* platform recruits crowd workers to help guide users toward more flexible and adaptive explanations of negative situations. In so doing, users are taught valuable coping skills that they can learn to apply on their own.

2.2   Crowd-Powered Applications

Crowdsourcing is a fairly new discipline and, while it presents many exciting opportunities, it is not without its challenges. Unlike outsourcing, crowd workers are not bound by contractual or managerial imperatives; they can come and go as they please, and it is often up to designers to find clever ways to recruit and retain this kind of ad hoc workforce. Further, since the crowd is typically diverse, both in terms of skill set and motivation, there is often a lot of variance in the quality of the work that gets done. To ensure high quality work, various techniques need to be employed (e.g., administering qualification tests to pre-screen workers, creating gold-standard questions to detect cheating, and filtering results *post hoc*, either through statistical techniques or through crowd-based adjudication).

While crowdsourcing is a large area of research, considerable attention has been given to microtask crowd work, an approach characterized by: (1) decomposing complex tasks into simple subtasks and (2) routing labor to many workers, either in parallel or across a series of iterative steps.   Much of the microtask crowdsourcing research has been conducted on online labor markets like Amazon's Mechanical Turk service (MTurk). MTurk is an attractive resource because it has its own application programming interface (API), allowing programmers to algorithmically coordinate human labor as needed. Frameworks like TurKit and Crowdforge go beyond the MTurk API, offering new ways to create more advanced workflows (Kittur, Smus, Khamkar, & Kraut, 2011; Little, Chilton, Goldman, & Miller, 2010). For example, Crowdforge has tools to automatically subdivide tasks, route them to different workers, and then aggregate the results. Using tools such as these, researchers have built a number of intriguing crowd-powered applications, including systems to compose encyclopedia articles (Kittur et al., 2011), plan travel itineraries (Zhang et al., 2012), edit word documents (Bernstein et al., 2010), and assist individuals with visual impairments (Bigham et al., 2010). By leveraging human intelligence, these technologies provide services that go far beyond what can be offered by automated processes alone. Since workforces like MTurk are available on demand, day or night, many of these technologies operate fairly quickly, despite the time required to hire and train a human for each task. There is almost always a worker

somewhere available to perform a task, provided he or she is properly incentivized to complete it. Recent work by Bernstein, Brandt, Miller, and Karger (2011) has shown ways to reduce the latency of these systems even further, by placing workers on retainer, so they are ready to work as soon as the user needs assistance.

Much of this dissertation is inspired by recent work on crowd computing and crowd-powered applications (e.g., Bernstein et al., 2010; Bigham et al., 2010; Zhang et al., 2012a). However, there are a few differences that distinguish this work from previous work on crowd-powered applications. First, multiple labor sources are used, each of which requires a unique set of instructions and incentives. Second, users are respondents themselves and specific efforts are made to facilitate this bidirectional interaction. Third, returning workers develop a reputation over time, which changes the way responses get vetted. Responses from well-reputed workers require less oversight and voting stages, which helps speed up the system response. Finally, as will be discussed later on, *Panoply* requires workers to respond tactfully – a 'soft' skill that is not easy to teach in crowdsourced settings. Innovations are required to ensure that workers not only complete the tasks as instructed, but also with tact and poise.

2.3 Computer-Based Psychological Interventions

Computer-based interventions offer several advantages over traditional, in-person therapy. Unlike in-person treatments, computer-based approaches are not burdened by cost, stigma, or the logistics of traveling and scheduling appointments (Cartreine, Ahern, & Locke, 2010). Also, there is evidence that patients are more willing to disclose private information to computers than to human interviewers (Kobak, Greist, Jefferson, & Katzelnick, 1996; Locke et al., 1990). Knowing more about a patient's issues, simply because she is more forthcoming, could help facilitate treatment.

To date, most computer-based psychotherapeutic approaches have come in one of three forms: (1) self-guided treatments that use interactive text and video modules to teach therapeutic concepts (2) distance therapy, wherein patients are connected with therapists remotely, via email, chat, phone, or webcam, and (3) hybrid approaches that

combine self-guided treatments with distance therapy, connecting users with clinicians on an as-needed basis (Cartreine et al., 2010).

Self-guided treatments have been applied to individuals with various conditions, including depression, anxiety, eating disorders, and substance abuse. While assessments of computer-based therapies have generally not reached the methodological equivalent of other evidence-based therapies or pharmacotherapies (Kiluk et al., 2011), there is now considerable evidence that these approaches can be efficacious. Recent meta-analyses and systematic reviews have shown positive results for several affective disorders, most notably depression and anxiety (Reger & Gahm, 2009; Spek et al., 2007). The effect sizes vary across studies, but there is evidence that some of these treatments can be as effective as traditional, in-person psychotherapy (Kaltenthaler et al., 2006).

One of the most oft-cited virtues of computer-based interventions is their supposed accessibility. Since they do not require clinician oversight, they can, in theory, be used at any time, as often as one likes. In practice, though, this advantage is not always seized by the consumer. A recent review of web-based interventions found a median completion rate of 56% (Donkin et al., 2013). When online interventions are open to the public and are not part of an experimental study, adherence drops even further (Christensen, Griffiths, & Farrer, 2009). Accessibility is not a valuable resource if no one wants to access the content in the first place. Ideally, users should be intrinsically motivated to engage with the technology on their own, without prompting from a clinician or an experimenter. Engaging interventions will provide higher doses of therapeutic exposure and are more likely to be revisited later (for instance, for individuals requiring 'booster' sessions).

There are several reasons why many existing computer-based interventions fail to provide compelling user experiences. Schueller, Muñoz, and Mohr (2013) describe many of the problems that currently plague the field. They note the need for collaborations spanning multiple disciplines, including psychology, design, and engineering. Further efforts are also needed to refine the design process itself. Computer-based interventions should incorporate lean, agile design strategies to help iteratively refine the user experience. Also, while it is important that new interventions are based on sound psychological principles, they needn't be perfect digital replicas of treatments that are

Figure 2.1: Examples of psychological skeuomorphs, taken from the *eCBT Mood* application. This app features therapists and waiting room clipboards, just as you might find in a clinic. This design approach is questionable. Many users download mental health apps precisely because they want to avoid these associations.

delivered in the clinic. Indeed, many computer-based interventions fall prey to 'psychological skeuomorphism'– a tendency to retain clinical elements within the digital realm, even at the expense of user experience (Schueller et al., 2013). Clinical components such as 'weekly sessions', 'paper and pencil' worksheets, and 'therapists' are sometimes blindly incorporated into apps without a careful consideration of their effects on the user (see Figure 2.1).

2.4 Online Emotional Support Systems

In addition to computer-based mental health interventions, there is a wide range of online social support services. Anonymous, timely social support could be extremely helpful for anyone, regardless of clinical diagnosis. To serve this need, several companies are building systems to utilize peer support, in lieu of professional counseling. *SpillNow* offers students feedback on stressful, school-related situations. Respondents were originally a handful of trained volunteers, but the service has now opened up to anyone who wants to help. *OverTheLine*, a service from MTV, in collaboration with the MIT Media Lab, offers individuals an anonymous forum to discuss issues related to bullying

and other teen-related stressors. Systems such as these can reach many people, but it is unclear how effective they are at promoting long-term emotional recovery, let alone short-term emotional relief. In fact, contrary to our intuitions, a large body of experimental research suggests that socio-affective support, by itself, does not necessarily promote emotional recovery of stressful events (see Rimé, 2009). Unless the appraisal of an emotionally charged event is modified or challenged in a structured way, it may remain destabilizing. These systems also fail to disseminate therapeutic techniques that might benefit users offline. Thus, online support groups that offer only empathy and shared experiences may provide short-term emotional relief, but little long-term emotional recovery.

At the same time, these platforms are very popular. Horrigan (2001) reports that over 28% of Internet users have visited an online support group at least once. Further, many people report it easier to share their problems online than through face-to-face interactions (Kummervold et al., 2002). Despite their popularity, there remains a distinct lack of rigorous, controlled studies on the efficacy of Internet peer support groups (Griffiths, Calear, & Banfield, 2009). Future work is needed to determine whether these platforms are as helpful as they claim to be. For some individuals, unmoderated Internet support platforms may actually be detrimental. Kaplan et al. (2011) showed that individuals who participated frequently on unstructured, online mental health forums reported greater psychological distress over time. Interestingly, this study also found that those who reported more positive experiences on these forums showed more distress than those who reported less positive experiences.

2.5 *Panoply*'s Role

This dissertation occupies an important middle ground between self-guided computer-based approaches and online peer support networks (see Figure 2.2). Unlike self-guided computer-based approaches, *Panoply* offers timely, personalized feedback. Core therapeutic concepts are learned experientially, in relation to the user's unique issues. Social dynamics and persuasive design strategies are also employed to heighten user engagement.

Panoply can also be distinguished from Q&A sites and online peer support networks where feedback is typically unstructured and often unmoderated. Unlike these platforms, the feedback in *Panoply* is guided by evidenced-based psychotherapeutic principles, such as cognitive reappraisal and cognitive restructuring. It is also vetted by crowd computing algorithms, to help improve the overall quality of the responses. In addition, *Panoply* does not penalize users for posting too frequently and even the most quotidian dilemmas can be submitted to the system. By contrast, many other online support networks actively discourage flooding or posting too frequently.

Social Support Systems   *Panoply*   Self-guided Interventions
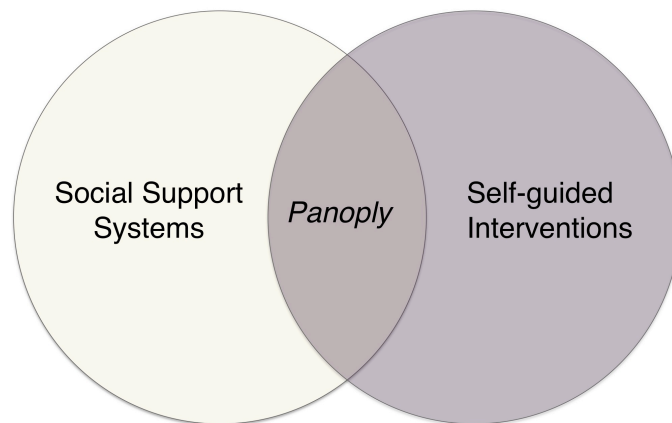
Figure 2.2: *Panoply* combines the personalized interactivity of social support systems with the evidence-based techniques of self-guided computer-based mental health interventions.

Chapter 3

Prototype Assessment

The complete *Panoply* platform did not come together in one fell swoop. Many of the system components were prototyped and tested separately before being integrated into the final platform. This chapter describes this iterative design process and reviews findings from two large-scale feasibility experiments. The feasibility experiments were conducted to test some of the core assumptions upon which the entire project rests – namely, that crowd workers can be trained, on demand, to provide therapeutically useful assistance.

3.1 Usability Testing

3.1.1 MTurk Testing

From the very start, MTurk was the most important prototyping tool for this dissertation. Almost all elements of the *Panoply* system, even those that would not ultimately involve crowd workers, were tested on MTurk. Using this approach, many aspects of the system could be tested, redesigned, and tested again, all within the span of several hours. In essence, MTurk acted as a stand-in for an active user base, allowing iterative design and hypothesis-driven experimentation on a large scale with hypothetical users. Unfortunately, the gains in speed and scale that MTurk offers were often offset by losses

in quality and consistency. Workers on this platform tended to rush through tasks or misunderstand instructions, making some experimental results very hard to interpret. Also, while it is possible to ask workers for their feedback, it is difficult to conduct detailed interviews with this population to find out precisely why some design approaches are flawed. To compensate for these limitations, several in-person usability tests were conducted in the laboratory.

3.1.2 Laboratory Testing

Throughout the development process, MIT undergraduates were recruited to help assess the usability of the various *Panoply* features. After providing informed consent, participants were given access to the software and were asked to create a user account, post content, or compose responses. An experimenter was present at all times and the sessions were moderated using techniques such as 'concurrent thinking aloud', 'retrospective thinking aloud', and 'retrospective probing.' (see Rubin & Chisnell, 2008). This approach was not as fast and convenient as MTurk, but it provided considerable insight into some of the more subtle (but nonetheless important) aspects of the *Panoply* user experience. For example, the laboratory studies helped identify subtle points of confusion in the task instructions that could never be observed through interactions with MTurk workers. Oftentimes, simple changes in the wording of the instructions resolved these issues. Lab participants also helped make the site more usable by identifying points of confusion around site navigation and other user interface components (e.g., buttons, links). Thus, the lab studies offered more scalpel-like precision when evaluating various design components. MTurk, by contrast, was more of a blunt instrument; it was fast and it helped identify some of the most striking, obvious design blunders, but it could not be counted on to assess the more nuanced aspects of the user experience.

3.2 Feasibility Experiments

In practice, many of the usability tests were small in scale and fairly ad hoc. Quite often, only a few interactions between MTurk workers or lab participants were required to test

small updates to the system. However, some of the larger, core assumptions behind the technology could not be evaluated quite so quickly or easily. Before devoting considerable time and resources towards the development of *Panoply*, two large-scale feasibility studies were conducted to help validate some of the key premises on which the whole project rests. In particular, these studies examined (1) whether crowd workers could be taught to compose empathetic reappraisal responses and (2) whether crowd workers could be taught to identify distorted negative thoughts. If it could be shown that crowd workers can be trained to complete these tasks reasonably well and reasonably quickly, then further development of the *Panoply* platform could be justified.

3.2.1 Experiment One

Much of the way *Panoply* is designed rests on the assumption that crowd workers do not naturally compose high quality cognitive reappraisals and empathetic statements on their own; they need to be trained and their work needs to be vetted for quality. But perhaps this is an overly pessimistic view of crowd workers. If most crowd workers naturally generate good responses, then human computation algorithms would not be needed to guide workers or control for quality. The system could simply ask anyone from the crowd to use whatever means at their disposal to help cheer a person up[2].

To explore these ideas further, I conducted a study that directly compared responses from two separate conditions: an *unstructured* condition, in which workers were simply asked to help the user feel better and a *structured* condition, in which workers were given structured guidance to provide empathetic statements and cognitive reappraisals. Responses from both conditions were examined to see whether they would be appropriate for use in a larger system like *Panoply*.

Method

Participants on MTurk were asked to respond to three short descriptions of negative situations (see Table 3.1).

---

[2] This 'open-ended' approach is common in many of today's emotional support applications. Respondents are often not given any specific instructions as to how to support or respond to the negative situations of other users.

Michael says, *"I have been working on a blog and have made many mistakes. I'm feeling really stressed."*
Sarah says, *"My boyfriend did not call me this morning, like he said he would. I'm feeling really angry"*
Jack says, *"Yesterday my dad drank the last of the coffee and didn't make more. I'm feeling really irritated!"*

Table 3.1: The three stressful situations used in experiment 1. All were contributed by MTurk workers. The names have been changed to protect the identities of the workers.

After accepting the task, 102 MTurk participants were randomly assigned to the unstructured or structured condition. In the unstructured condition, participants were asked to help the target feel better about his/her situation. They were asked to limit their response to six sentences or less. In the structured condition, participants were asked to first empathize with the target, using no more than three sentences. They were then asked to reframe the target's thoughts to make the situation seem less distressing. For the reappraisal component, responses were also limited to three sentences. As such, the total length of the structured and unstructured responses was balanced and limited to six sentences in both conditions.

Next, 70 MTurk workers were recruited to rate the responses. These raters saw a random mixture of 34 structured and unstructured responses. Four decoy responses were also included, two of which were off-topic, and two of which were overtly rude and uncaring. Five raters failed to give poor scores to the decoy responses and these individuals were not included in the overall ratings scores.

For each response, workers were asked to rate the extent to which they agreed or disagreed with the following two statements:

1) This response is empathetic. The respondent seems to sympathize with this individual's situation.

2) This response offers a positive way to think about this situation.

Ratings were made using a 7-point Likert scale, with endpoints labeled as 1="strongly disagree" and 7="strongly agree." We used data from the first and second Likert questions as scores for empathy and reappraisal, respectively.

Results

A two-way multivariate analysis of variance (MANOVA) was conducted to examine the difference between the structured and unstructured responses, with response structure (structured vs. unstructured) as a between-subjects factor. Empathy and reappraisal scores were used as dependent variables, and the type of input stressor was included as a covariate in our analyses.

The results showed that empathy scores were significantly higher in the structured condition ($M = 5.71$, $SD = .62$) compared to the unstructured condition ($M = 4.14$, $SD = 1.21$), [$F_{1,99} = 73.02$, $p < .005$]. Similarly, the structured condition had significantly higher reappraisal scores ($M = 5.45$, $SD = .59$) than the unstructured condition ($M = 4.41$, $SD = 1.11$), [$F_{1,99} = 34.90$, $p < .005$] The covariate analysis showed no significant effect of input statement on either the empathy scores [$F(1, 99) = .387$, $p > .54$] or reappraisal scores [$F_{1,99} = .194$, $p > .66$], suggesting that the type of stressful situation did not produce differential effects across the different conditions.

Discussion

These results support the hypothesis that, with guidance, most crowd workers respond to strangers with significantly better empathetic responses and significantly higher quality reappraisals. However, perhaps the most illustrative finding from this experiment was the consistent missteps observed from crowd workers in the unstructured condition. This group relied heavily on problem-focused coping methods, often providing specific prescriptions for how to fix the problem (as contrasted to the structured response group, which generally provided advice on how to *rethink* the problem). Without intimate knowledge of the person or the situation, the problem-focused advice often seemed rash or inappropriate (e.g., "you should dump your boyfriend since he clearly has no respect for you"). Many of these respondents seemed well intentioned, but their advice extended far beyond the context they were given.

By contrast, the results from the structured condition were more encouraging. It was surprising how well individuals were able to learn concepts like empathy and reappraisal and apply them to strangers' situations, despite knowing very little context. Of course, the data overall still showed considerable room for improvement. Even though all the workers were purportedly from English-speaking countries, some of the responses were poorly written and suggested poor comprehension of the task. Also, while some of the respondents wrote extremely sensitive reappraisals, others seemed somewhat tactless and hurried. Finally, even in the structured condition, many respondents still made attempts to help the person solve the problem. Before deploying the full *Panoply* system in a large user study, these issues would need to be solved. The steps that were taken to deal with these problems are described in detail in the system design section (see section 4.3.3).

3.2.2 Experiment Two

In addition to providing empathy and reappraisal support, the *Panoply* system was also conceived to help users identify and dispute cognitive distortions. This process, known as 'cognitive restructuring', is itself a form of reappraisal, but it involves reframing thoughts rather than reframing situations.

In cognitive therapy, cognitive distortions are defined as logical fallacies within negative statements (Beck, 1979). For example, consider the following statement: "*I'm going to flunk out of school and I'll never get a job, I know it!*" This statement would be classified as distorted because it makes unrealistic assumptions about the future that no one could know. There is no way this person could know that s/he will flunk out and be perpetually unemployed. By contrast, a statement like "*There is construction on my street and I didn't get much sleep last night,*" is not distorted because it does not contain any illogical assumptions, predictions, or conclusions.

The *Panoply* system assumes that crowd workers can reliably identify cognitive distortions with very little training. The second feasibility study tests this assumption directly by asking workers to classify a set of input statements as either distorted or undistorted.

Method

73 participants were recruited from MTurk. Participants were each given a short tutorial to help train them on the concept of cognitive distortions. The training was extremely short and involved (1) a brief definition of cognitive distortions, using simple nontechnical language and (2) several quick examples of distorted and undistorted statements. After completing the training, participants were asked to classify a set of 32 statements. Each statement was negatively valenced and included a one-to-three sentence description of an emotion-eliciting thought or situation. Half of the statements were distorted in some way, and each participant saw a random mixture of distorted and undistorted statements. The stimuli set included real stressors described by MTurk workers. Distorted thoughts were also taken from online resources and cognitive therapy literature (Burns, 1999), and were chosen to capture a wide variety of distorted thoughts (see Table 3.2).

Classification	Input Statements
Distorted	*"My son acted up at church. Everyone must think I have no control over him and that I'm a terrible parent."* *"I forgot my lines in the play and really made a fool of myself."*
Undistorted	*"My best friend doesn't call me as much as she used to."* *"My car needs to be repaired, but I'd rather use that money to pay my rent!"*

Table 3.2: Examples of the distorted and undistorted statements workers were asked to classify.

Results

A confusion matrix was used to plot MTurk classifications against the ground truth. Accuracy was calculated by dividing the number of correct classifications (true positives

and true negatives) by the total number of all recorded classifications. On average, workers correctly classified 89% (*SD*=7%) of the input statements (see Figure 3.1).
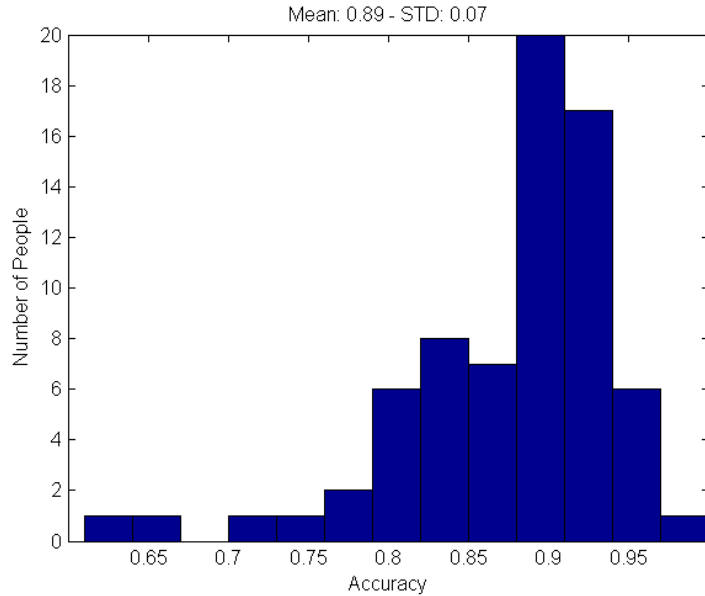
Figure 3.1: A histogram of classification accuracy from the sample of MTurk workers.

Discussion

Based on the analysis from experiment 2, it seems that MTurk workers can correctly identify most cognitive distortions within short, one-to-three sentence descriptions. With minimal instructions, MTurk workers seemed to understand the concept quite well. It seems reasonable to conclude that, with additional training, workers could also be recruited to label these distortions as well and classify them further as examples of 'all-or-nothing thinking', 'fortune-telling', etc.

Findings from both experiments were encouraging and suggest that crowds can in fact be trained to help identify and reconstrue negative, distorted thinking. Unfortunately, these early studies and preliminary prototypes could not answer the question of whether repeated exposure to this kind of crowdsourced feedback might be therapeutic over time. To answer this question, the full *Panoply* system needed to be built and deployed in a randomized controlled trial. In the next section, I describe the final design of the *Panoply* system. I also discuss the design philosophy that guided the development of the platform

and I describe some of the challenging trade-offs that were negotiated when designing for both engagement and therapeutic efficacy.

Chapter 4

System Design

Panoply was built to satisfy the twin aims of efficacy and engagement. Throughout the design process, there was a constant need to negotiate tensions between providing evidence-based content on the one hand and delivering a high quality user experience on the other. To date, most computer-based interventions for affective disorders are optimized for efficacy and accessibility, but not necessarily for user experience. Yet, interactive technologies offer exciting new pathways for engagement, many of which can be implemented without sacrificing clinical efficacy. To make *Panoply* as engaging as possible, while still providing sound therapeutic principles, several design principles were used that are commonly employed in successful online social systems. This section discusses four design principles that were especially important for *Panoply* and describes how they were implemented in the system.

4.1 Design Principles

4.1.1 Accommodate Short Bursts of Attention

The way people absorb application content has changed dramatically over the past few years. Users are now more likely to 'snack' on applications, visiting them frequently, but in short bursts (see Oulasvirta, Tamminen, Roto, & Kuorelahti, 2005; Vaish, Wyngarden,

Chen, Cheung, & Bernstein, 2014). Therefore, instead of bundling application content into long, once-weekly sessions that require lengthy periods of sustained attention, computer-based mental health interventions should accommodate multiple levels of commitment. Ideally, intervention exercises should be accessible for people who have a long afternoon free, but also for those can only spare five minutes here and there during their commute or as they wait in line for their coffee. Yochai Benkler, a scholar of peer-production systems, calls this design principle 'heterogeneous granularity' and argues that it helps sustain participation on commons-based peer production platforms (Benkler, 2007). When possible, the *Panoply* system follows this design strategy; it offers tutorials and other didactic content, but all the content is self-contained in short, bite-sized chunks. Everything can be absorbed in piece-meal, if necessary, without requiring extended time commitments on the part of the user.

4.1.2 Autonomy and Competence

Several researchers in human-computer interaction (HCI) have applied Self-Determination Theory to explain the intrinsic appeal of successful online games and learning experiences (Ryan, Rigby, & Przybylski, 2006). In short, Self-Determination Theory argues that human motivation is closely linked to core fundamental psychological needs, such as autonomy and competence (Ryan & Deci, 2000). Motivation is thought to increase when these innate psychological needs are met. Computer-based interventions can increase autonomy by offering users more choices in the types of exercises that are offered. Competence can be supported by offering positive feedback on task exercises. Unfortunately, many existing digital applications for mental health offer limited choice and are not sufficiently interactive to provide feedback to help the user feel competent. *Panoply,* by contrast, offers its users a variety of intervention options and there are considerable sources of positive feedback within the site, most notably from crowd workers and other registered users.

4.1.3 Speed

Feedback from interactive applications should be quick. This seems to be particularly the case for social-based Q&A sites. A recent exploration of *StackOverflow* identified speed as one of the key factors in its early success; users were delighted to have their questions answered in incredibly short periods of time (Mamykina et al., 2011). *Panoply* uses MTurk workers to help provide feedback at any time, day or night. Quick response times from the crowd are hypothesized to enhance the overall user experience of the system.

4.1.4 Persuasive Design

Ideally, computer-based mental health systems should be used habitually, at least for as long as the treatment is required. Proponents of persuasive design argue that habits are best formed when the following three elements are in place: (1) *motivation* to complete a behavior, (2) *triggers* to remind a person to do it, and (3) the *ability* to perform the behavior when desired (Fogg, 2009a). Interestingly, persuasive design strategies emphasize the latter two and suggest these should be optimized before one attempts to increase motivation in a user. These principles should be regarded as heuristics, not empirically derived laws, but they were nonetheless useful guidelines to consider when developing the *Panoply* platform.

Motivation

Motivation, generally speaking, was piqued by appealing to intrinsic impulses whenever possible. Without being too overbearing, the platform repeatedly reminds users that repeated use of the system might help them learn to better manage stress. Users were also reminded that helping others can be an end in itself, insofar as it helps boost positive emotions.

Triggers

Panoply also has many built-in triggers that remind people to use the system. The site provides email notifications whenever a user receives new responses or new feedback

from other users. These emails act as triggers, prompting users to return to the site. Users who might not initially be motivated to return to the site may find themselves there anyway, after clicking the links in the email notifications. It is worth noting that this method of 'push notification' is widely used in many successful online platforms such as Facebook and Twitter (and indeed, persuasive design advocates suggest that imitation of successful similar systems is not a bad design approach, e.g., Fogg, 2009b).

Accessibility

Panoply is also extremely accessible. Should someone want to visit the platform at 3am, there will always be respondents. Users can interact with the system at any time; they do not need to wait for weekly sessions to start, as is common in some online interventions (e.g., *Big White Wall*).

4.2   Technology Stack

To best incorporate all the design principles mentioned thus far, *Panoply* was built to be a data-driven web application, containing all the features one might normally expect from most Q&A sites (e.g., user registration, user profiles, email notifications).

The platform was built using Django, a python-based web framework for building data-driven applications. The site was hosted on *Heroku*, a cloud application platform that supports web frameworks such as Django. In addition, I used PostgresSQL for back-end database storage, *New Relic* for real-time application monitoring, *SendGrid* for email services, and the Heroku Scheduler for running cron tasks. Additional Django libraries were used to facilitate common tasks such as database migrations and user registration. Amazon S3 was used to host static assets, such as javascript and html files.

On the front-end, various javascript libraries were used to build the interactive tutorials, handle data visualization, and to help pass user data asynchronously to and from the server. I used jQuery, Highcharts.js, and various other libraries found within the *Zurb* front-end framework (www.zurb.com).

To handle S3 storage and to facilitate connections to and from MTurk, I used Boto, a python interface to Amazon Web Services. Data from MTurk was stored on Amazon's

servers and the Heroku PostgresSQL database. By storing data on PostgresSQL, I could make use of Django's 'signal dispatcher' framework, which allowed the platform to automatically post new crowdsourcing tasks as soon as MTurk data was uploaded to our server. This method was preferred over a 'pull' method, in which a cron task polls the MTurk server at automated intervals to determine if new data has been stored.

The front-end makes limited use of 'responsive design patterns' (Bryant & Jones, 2012), but the interface design of all the components (panels, tutorials, etc) was built to scale for tablet devices. The site was tested on all major browsers as well as Android tablets and iPads.

As the site was being developed, front-end design practices migrated from skeuomorphic techniques to 'flat' design approaches. The skeueomorphic approach utilizes effects like shadows and gradients to simulate contours and physical shapes that might be found in the real world. The flat design approach emphasizes the simplicity and reduced cognitive load of unembellished shapes and colors. Because it was developed throughout this transition from skeuomorphism to flat design, *Panoply* straddles both philosophies – while some of the buttons have shading and depth, there is limited use of gradients and other skeuomorphic elements.

4.3   Panoply Platform

4.3.1 Overview

The overall architecture of *Panoply* is similar to many Q&A sites, such as *StackOverflow*, *Quora*, or *Whisper*. Users on *Panoply* have the option of both posting content or responding to others. They can review responses they have received and they are given feedback about their own performance on the site. Unlike typical Q&A sites, however, responses on *Panoply* are subdivided into three categories ('support', 'debug', and 'reframe'). Responses are also groomed for quality using crowd-computing techniques.

The remainder of this chapter surveys the main components of the *Panoply* platform. I discuss how users post content, how they respond to content, how they review responses, and how they receive feedback from other site users. There is also a description of the

features that are specific to MTurk users. I begin with a description of how users post content on the site.

4.3.2 Posting Content

Onboarding

Before users can post content and gain access to the full *Panoply* site, they must first go through an 'onboarding' process. This helps orient users to the site's objective and its primary features. It also helps establish norms of behavior and walks users through the process of creating their first post. The *Panoply* onboarding sequence is depicted in Figure 4.1.
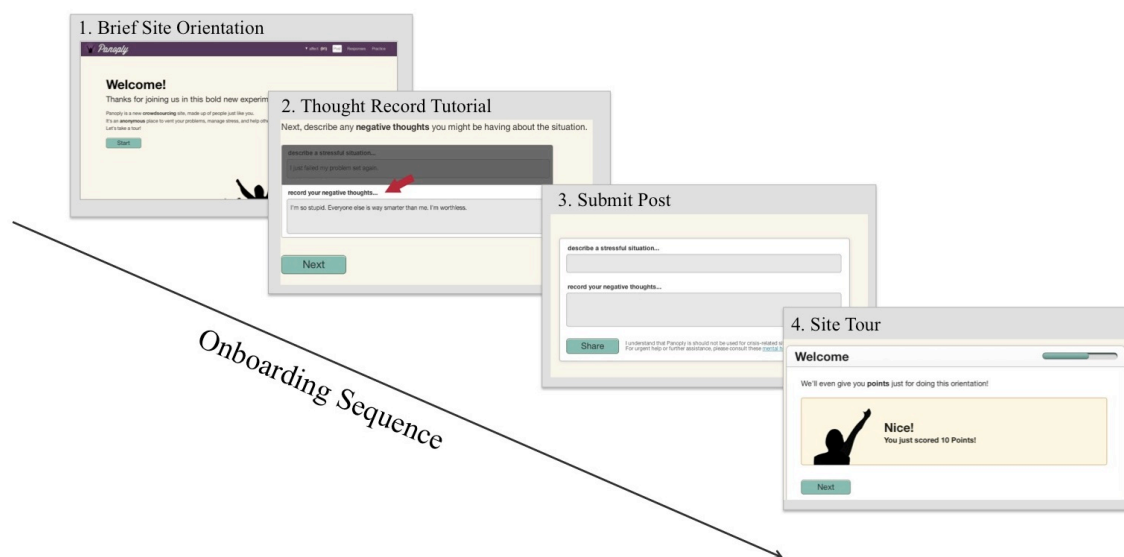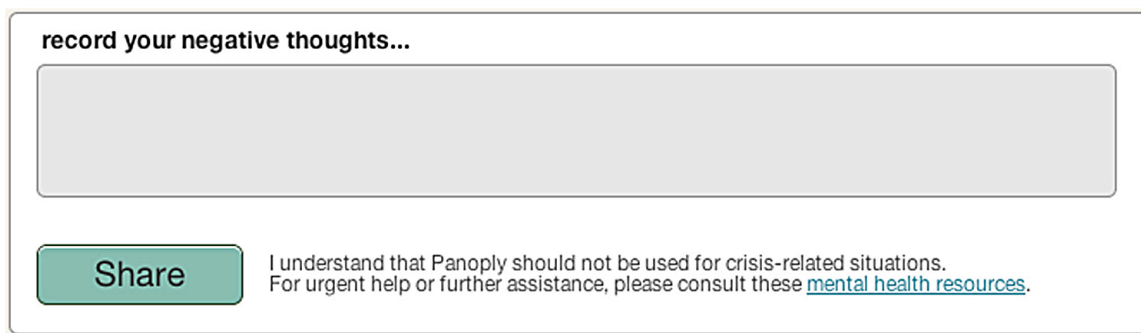
Figure 4.1: When users first activate their accounts, they move through a brief onboarding sequence that orients them to the site. They are given: (1) a brief overview, (2) an introduction to thought records, (3) a chance to compose a thought record, and (4) a tour of the site's various features and therapeutic aims.

An important objective of the onboarding process is to get users to submit a post as quickly and effortlessly as possible. Once a post is submitted to the system, responses are returned to the user, creating a natural incentive for further exploration of the site. Thus, the onboarding process not only helps educate users about the site content, it also

provides an opportunity to 'hook' the user, drawing them into site interactions that may continue to evolve over several days.

Post Interface – Version One

In the early prototypes, users were given little instruction about how to compose a post on *Panoply*. The initial hope was to simply present a text box and let people submit negative thoughts as quickly as possible, without complicating matters by including long, drawn-out tutorials and instruction pages. Many apps currently on the market do precisely this, letting users vent any kind of negative thoughts, without providing any instructions for how to do so (e.g., *eCBT Mood*, *Emotional Bag Check)*. The first design is shown in Figure 4.2.

record your negative thoughts...

Share   I understand that Panoply should not be used for crisis-related situations. For urgent help or further assistance, please consult these mental health resources.

Figure 4.2: A screenshot of the original post interface. The original design was minimalistic and did not offer much guidance. Users were free to write about anything that was troubling them.

Users were given examples of previous posts and a maximum character count, but little additional guidance. This design was simple and elegant, but the posts that people submitted were often not appropriate for the system. These minimal instructions did not help people produce useful descriptions of negative situations and thoughts. People composed detailed descriptions of their situation, but there was little indication of how it affected them emotionally or what kinds of negative thoughts it brought to the surface. For *Panoply* to be most effective, the posts should feature some indication of how the user is interpreting the stressful situation. If one of the goals of *Panoply* is to help people

systematically reinterpret stressors, it is important they record their initial interpretations of the situation.

Post Interface – Version Two

To address these shortcomings, the posting interface was updated to include more instructions. Before composing a response, users now had to first complete a short, one to two-minute tutorial. The tutorial introduces the concept of a CBT-style thought record and its importance for well-being and stress management. As with all written instructions on the *Panoply* site, the tutorial language was refined repeatedly over the course of several usability experiments (using both MTurk workers and participants in the lab). Throughout the tutorial, repeated assurances of anonymity are provided, to help remind users that anything they post will not be linked to any personal identifiers, such as their name or email address.

When posting, users are asked to first describe a stressful situation, using one to two sentences. Next, they are asked to record any automatic negative thoughts they might have about the situation. As users click within the interface, they are shown additional hints and are given links to examples submitted by previous users.

Figure 4.3: The final version of the posting interface includes embedded hints and additional, 'just-in-time' instructions. To keep the design elegant and uncluttered, the embedded hints do not appear until the user clicks on the textbox.

It is worth noting that there is a limit to how much content a user can include in their posts. Unlike other peer support sites (e.g., *StudentSpill)*, *Panoply* forces users to condense their submissions into tiny chunks. This is an important design decision for several reasons. First, it alleviates tremendous burden on the part of the respondents. Rather than wade through a complex, multi-paragraph depiction of multiple issues, respondents can focus their attention on a concise 1-3 sentence description. Second, condensing a negative thought or situation requires some detachment on the part of the user and could be therapeutic in itself. Finally, recording short negative thoughts is common practice in CBT-based thought records. To prevent users from flooding the system and posting repeatedly, the system bars users from posting more than four times in 24 hours.

Once users submit their first post, they are congratulated and reminded that other *Panoply* users will review their post and send them something back in return. They are then given a brief summary of the rest of the platform (stage 4 in the onboarding process, see Figure 4.1).

Site Orientation

When creating the brief on-screen synopsis of *Panoply*, great care was taken to emphasize the potential therapeutic elements of the site without raising expectations too high. This is a challenging balance to achieve. On the one hand, it is important to highlight the possible benefits of *Panoply*, to help encourage greater use of the site. Doing so also enhances any placebo effects that might come from the mere expectation of improvement – a technique that is encouraged by many researchers (see Linde, Fässler, & Meissner, 2011). On the other hand, it is important to set realistic expectations about the platform. Users were reminded that the site is still an experiment and that any feedback they get on the site will come from peers like them, many of whom are still learning. The entire text is included in Appendix B.

To better illustrate the platform going forward, let us consider a specific user scenario, drawn from real data collected on *Panoply*. To preserve anonymity, aliases are used in place of real names and usernames. When appropriate, usernames are redacted in the screenshots.

Let's consider Jenny, a 28-year-old graduate student at MIT. She activates an account on *Panoply*, chooses the username 'FortuneCookie' and logs into the system. She is welcomed to the site and is introduced to the concept of a thought record. After reviewing the instructions, she recalls a recent stressful situation and writes:

> Ended a 2.5 year relationship with my boyfriend I'm not getting any younger. I am not where I thought I would be at this point in my life. I hate the thought of starting over. Everyone else in my peer group is engaged, getting married, and having children. What is wrong with me?

After submitting this post, Jenny is thanked and then briefed on the main features of the site (for a full transcript of this text, see Appendix A.1). Meanwhile, her post is sent out to several sets of crowd workers. As soon as she clicks 'share', a complex sequence of crowdwork is set into motion. The precise algorithm that ensues depends, in part, on her progress within the site and her status as a user. Since she has only just registered on the site, the system first solicits messages of support.

Before any messages are composed for Jenny, workers from MTurk review her post to make sure it is appropriate and to assess whether there is any indication that she might harm herself or others. In the rare case that her post gets flagged, a follow-up email is sent out immediately. The email includes links to mental health resources and reminders that *Panoply* is a self-help tool and is not to be used for crisis situations.

4.3.3 Responding to Others

Once Jenny submits her first post and completes the site orientation, she is encouraged to switch over to the role of respondent. On *Panoply*, responding to others is framed not just as a good deed, but also as a chance to help oneself. Users are told that each time they respond to others they get to practice stress management techniques that can be applied to themselves offline. They are reminded that teaching others can be an exceptionally great way to learn. This concept, sometimes referred to as peer-based learning, has been studied at length in pedagogical research (Cohen, Kulik, & Kulik, 1982). While it is a huge and complex area of research, meta-analyses consistently suggest that teaching

others is a powerful way to deeply absorb new concepts (Cohen et al., 1982; Rohrbeck, Ginsburg-Block, Fantuzzo, & Miller, 2003).

Panoply also highlights the possible mood-boosting effects of simply being kind to others. To reinforce this notion, the interface provides links to positive psychology research showing how small 'acts of kindness' can significantly enhance positive emotions and well-being (Buchanan & Bardi, 2010; Lyubomirsky, Dickerhoof, Boehm, & Sheldon, 2011).

In its current instantiation, *Panoply* solicits responses from two separate populations: (1) MTurk workers who are paid in cash and (2) *Panoply* users like Jenny who are encouraged to help others for free. The system design for MTurk workers differs significantly from what *Panoply* users experience and it will be discussed in detail in section 4.3.6. For now, I limit the discussion to the experience of users like Jenny who are not paid for creating responses. These users may be volunteers or they may be persons experiencing emotional distress who want to help others as part of their own therapeutic process.

Practice Panel

To create responses, users must navigate to the *Practice* panel (see Figure 4.4). The practice panel has three modules, displayed in increasing order of difficulty: *support*, *debug*, and *reframe*.

Previous research on positive psychology interventions has found that, while variety is important (Sin & Lyubomirsky, 2009), too much choice can be overwhelming and may weaken the overall efficacy of the intervention. For instance, Schueller & Parks (2012) compared interventions that offered a choice of two, three, or six positive psychology exercises. Those assigned to do two or three exercises significantly outperformed those assigned to do six. Those in the six exercise condition may have been overwhelmed by the 'paradox of choice' (Schwartz, 2005) and may also have found it difficult to attend to any one particular intervention deeply enough for its effects to take hold. For these reasons, *Panoply* only requires users to master three different modules.

The *debug* and *reframe* modules were chosen because they both involve reconstruing stressful thoughts and situations. Together, they offer two complementary flavors of

reappraisal. The debug module specifically targets distorted negative thoughts, while the reframe module offers users a more open-ended way to reappraise the meaning of a situation (as well as any negative thoughts that might accompany it). The 'support' module functions as an 'acts of kindness' intervention and is a fairly easy concept to grasp.

Each module includes a short, interactive tutorial culminating in the opportunity to create responses for other *Panoply* users. First time users like Jenny are initially granted access to the *support* module (the other modules remain hidden behind a locked panel until additional tutorials are completed – see Figure 4.4).
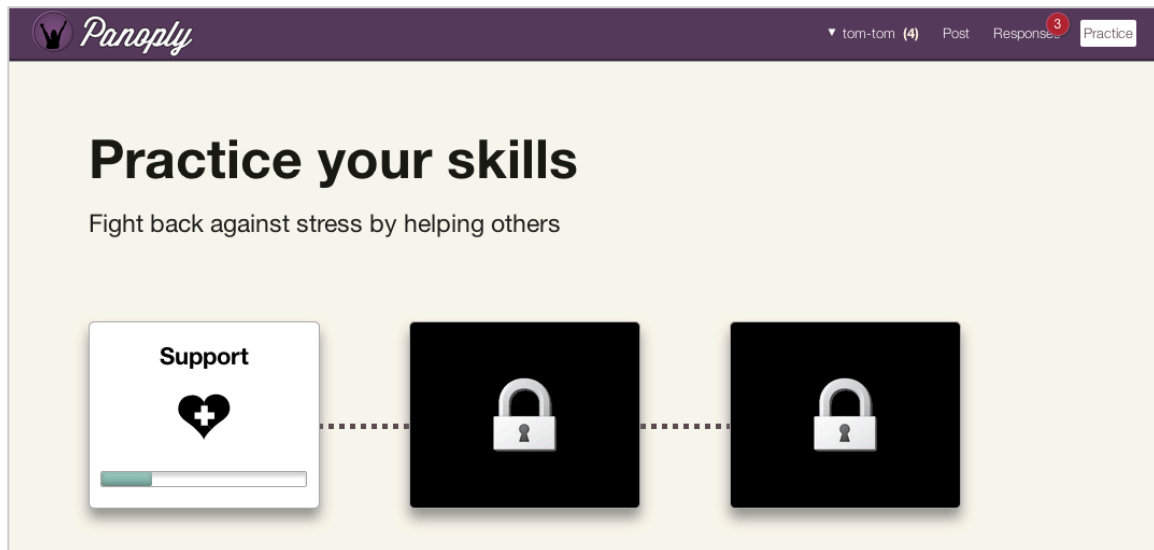
Figure 4.4: A screenshot of the practice panel. This user has completed part of the 'support' training. Additional modules are unlocked as the user progresses through the site.

Thus, the practice modules are introduced in a graded, hierarchical fashion, not unlike the way many video games are structured. To further emphasize this game-like component, users are awarded progress points each time they complete a module (for more about the point system, see section 4.3.5).

One of the most difficult design challenges was finding ways to teach concepts like *support*, *debug*, and *reframe* in an extremely short period of time. No preexisting knowledge of psychology could be assumed for any of the *Panoply* users or MTurk

workers. In fact, most of the design work for this platform was focused on the instructional tutorials. The content of these tutorials was redesigned considerably until users could complete them in a few minutes and yet still demonstrate a thorough understanding of the concepts. Because these training sessions were so important, the next sections describe them in detail.

Support Module

The support module teaches active listening skills and offers a chance to practice 'acts of kindness.' To introduce the concept, users are asked to first imagine a hypothetical situation in which they are about to take a test, but are really nervous. They are then given a set of example responses and are asked to choose which best exemplify the qualities of a good, supportive response. After making their choice, the interface explains why their selection was or was not empathetic (see Figure 4.5).

These explanations help further reinforce the key attributes of an empathetic, supportive response. After assessing four sets of example responses, users are congratulated and told they are ready to compose a real support response to another *Panoply* user. They are shown another user's post and they are given some additional suggestions to help compose their response. Specifically, they are told to (a) address the user directly, (b) share how they might feel if they were in the user's situation, and (c) let the user know his/her emotion makes sense, given the situation.

After composing a response, users are asked to review their own work before submitting it. This particular strategy has been used before in crowdsourcing contexts and can significantly enhance the quality of crowdsourced responses (Dow, Kulkarni,
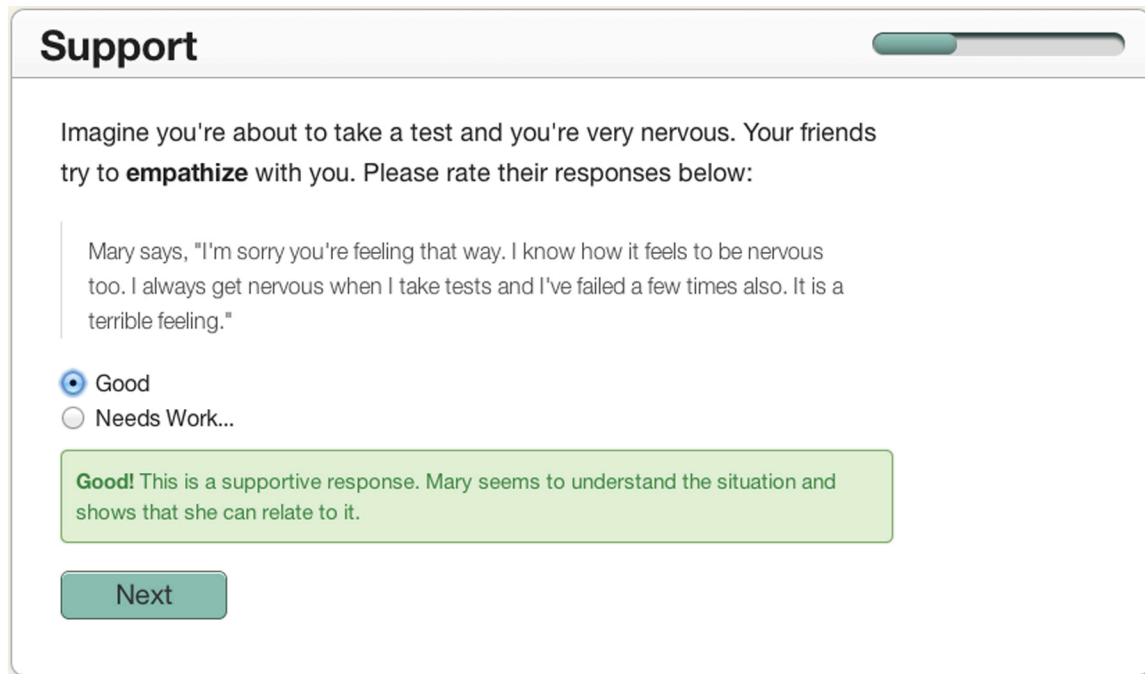
Figure 4.5: A screenshot of the *support* tutorial.

Klemmer, & Hartmann, 2012). In the *support* interface, users must acknowledge that what they've written offers support, not problem solving advice. As was discovered in the large-scale feasibility studies, as well as subsequent pilot studies, the first instinct for many users is to try to solve the problem. This is certainly not a bad approach generally speaking, but it is inappropriate in this context for two reasons: (1) the respondent usually does not have enough information about the person or the person's situation to offer legitimate advice and (2) the point of the *support* exercise is to help a person feel understood, not to redirect the conversation toward problem solving tactics.

Once users complete the *support* tutorial, their progress is logged in the database so they do not have to do the training again. The interface continues to provide helpful hints, however, and users are always able to repeat the tutorial, if they feel like they need to refresh the concepts. This design pattern is used for all the practice modules, and serves as a way to provide support and guidance when needed, without adding excess clutter to the interface.

Debug Module

The *debug* module introduces the concept of cognitive distortions and cognitive restructuring. Conceptually, cognitive restructuring is a subtype of reappraisal, one whose emphasis is on reframing distorted negative thoughts, rather than negative situations. The process involves: (1) identifying logical fallacies within negative thoughts, (2) disputing these fallacies, and (3) replacing them with thoughts that are more rational and adaptive. This process is frequently used in cognitive-behavioral therapy for depression (Beck, 2011) and it can also be helpful for treating conditions like posttraumatic stress disorder (Marks, Lovell, Noshirvani, Livanou, & Thrasher, 1998) and social phobia (Mattick, Peters, & Clarke, 1989).

Of course, as with all language employed on *Panoply*, the *debug* module is mostly bereft of clinical terms like "logical fallacies" and "cognitive restructuring." Considerable effort was made to ensure concepts like cognitive restructuring could be explained in the quickest, simplest way possible. Usability testing in the lab suggested that the most expedient way to introduce the *debug* concept was to use an interactive tutorial, similar to the one employed in the *support* module.

In the *debug* tutorial, users are introduced to the concept of cognitive distortions ('bugs', in *Panoply* parlance). They are then asked to read several negative thoughts and determine whether or not they are distorted. Feedback is given after each selection they make to help reinforce the concepts. After completing the short tutorial, users are then given the chance to evaluate real negative thoughts submitted by other *Panoply* users. For each post that comes into the system, they are asked to first evaluate whether it appears distorted or not. If it appears distorted, they are then asked to determine which cognitive distortion best applies (see Figure 4.6).

To keep the interface simple, users are only offered four cognitive distortions to choose from: *All-Or-Nothing Thinking*, *Overgeneralization*, *Fortune-Telling*, *Mind-Reading*. If none of the four distortions can be applied, users are able to select 'other' as a default, fifth option. The four distortions included were the ones I considered the most common and most orthogonal to each other. Future versions of *Panoply* could certainly include additional distortions beyond the four listed.

Select the bug that best fits ███████ **'s negative thought.**

███████ says, **"Adjusting to moving in with my boyfriend, used to having my own space... I worry that we'll fight or that I'll be too rigid about the way to do things in the house or he won't be able to find work and be at home all the time... "**

○ **All-Or-Nothing:** Seeing things in black-and-white categories.
see examples...

○ **Fortune-Telling:** Predicting the future with absolute certainty. Anticipating that things will turn out badly.
see examples...

○ **Overgeneralization:** Making an overly broad generalization from an isolated event.
see examples...

○ **Mind-Reading** Overestimating your ability to read other people's thoughts.
see examples...

○ **Other:** This appears buggy, but none of the labels above apply. That's OK. There are many types of bugs we haven't introduced yet.

Figure 4.6: A screenshot of the *debug* task. If a post is deemed distorted, users are given the chance to apply a cognitive distortion label. The post shown in this figure is from a study participant and so the username is redacted. Even though users were expressly told to select anonymous usernames, they are redacted as an extra precaution.

Reframe Module

In the third and final module, users are taught to reframe negative situations using cognitive reappraisal techniques. Like the other modules, users must complete an introductory tutorial before they can compose responses for other users.

The *reframe* tutorial begins with a brief vignette. Users are asked to imagine making a small mistake at work, incurring the boss's wrath, and then getting screamed at in front of all their colleagues. They are told they might initially feel mortified and distraught in this situation, but that reappraisal techniques can be used to change the trajectory of their

emotional response. Several reappraisals for this situation are presented and the users are asked to identify the best and worst among them.

The bad examples highlight several common mistakes I observed during my first forays into crowdsourced reappraisal (see section 3.2). Just as I saw in the *support* category, novice users would often first try to solve the problem. For example, in response to the boss vignette, one novice respondent said, "you should march right up to your boss and tell him you quit!" While this might not be altogether bad advice, it offers no way to reinterpret the meaning of the situation.

Early on, I also saw rampant Pollyanna-ism; many responses were unrealistic and over-sweetened with bright-sided optimism. For example, one person wrote, "it sounds like your boss will be fired soon, so thankfully this won't be a problem in the future." Thinking about the inevitability of this boss's termination is certainly a positive perspective to consider, but it may not be very realistic.

By contrast, the good examples in the *reframe* tutorial avoid these pitfalls and illustrate some of the properties I have found to underlie ideal reappraisals[3]. The following (a real response composed by an MTurk worker) was chosen as an exemplary reframe of the boss situation:

> Perhaps this boss has some serious anger issues? I could even feel sorry for this person. Not being able to control your temper can be a serious health problem. It also makes you pretty unlikeable!

There are several qualities that make this a good reappraisal. First, there are no attempts to solve the problem; rather, the focus is on different ways to view the situation. Specifically, this reframe shifts the narrative away from the employee and whatever mistake s/he made and instead focuses on the impropriety of the boss's behavior. Also, instead of casting too many aspersions on the boss, this reframe invokes sympathy. It is

---

[3] It is worth noting that there is very little research to date on what constitutes good and bad reappraisals in interpersonal contexts. While the *Panoply* platform could help answer these questions in the future, for the time being informed intuition guided most of these design choices.

Figure 4.7: A screenshot of the *reframe* task. Users are asked to review their response before submitting them. The interface includes a link to various reappraisal strategies, to help users compose their response.

quite possible that the boss in question does indeed have trouble controlling aggression –
a foible that can have serious repercussions for one's health and social life[4].

After this quick examination of good and bad reappraisals, users are given the chance
to create their own reframes for other *Panoply* members (Figure 4.7). They are not asked
to use any one particular reappraisal strategy, but instead are given a bulleted list of

---

[4] As an aside, I should mention that such an interpretation is not something I would have
thought of by myself. But I now revisit this response anytime I'm confronted by rude or
aggressive behavior in the city (a situation that is regrettably not too uncommon during
the rush hour in Boston). I consider how aggression rarely makes a person feel good and I
try to sympathize with the person, instead of getting heated myself. This speaks to the
value of crowdsourcing reappraisals; when incentivized properly, crowd workers can use
their own unique perspective to generate novel reframes that might not ordinarily occur
to the end-user.

tactics to consider in case they need inspiration. The tactics include considering whether the situation in question is permanent or temporary, whether it contains a silver lining, whether something can be learned from it, whether it might say something positive about the person posting it, etc. These prompts were culled from reappraisal taxonomies cited in the research literature (McRae, Ciesielski, & Gross, 2012) as well as my own design intuitions. In the future, the interface might be able to predict which subset of strategies is most likely to help for a given situation or type of person and could provide a more targeted set of reappraisal suggestions.

As with the *support* module, users are asked to review their reframes before submitting them. They are asked to make sure their response offers ways to rethink the problem, not ways to solve it. They are also reminded that good reframes are creative, but realistic.

4.3.4 Receiving Responses

To best describe how responses are displayed on *Panoply*, we'll revisit the user scenario with Jenny, the woman who recently broke up with her boyfriend. As Jenny explores the *Panoply* site and works her way through the *practice* panel, multiple sets of crowd workers help create and curate responses to her situation.

When a crowdsourced response is ready to be viewed, users are notified in one of two ways. If they are actively using the site, they will see a notification icon appear in the upper right-hand corner, hovering above the link to the response panel. As with many other social sites, this notification is strategically positioned in the periphery, where the eyes are most sensitive to movement and contrast (see Johnson, 2010). Users are also notified by email. A cron task on the Heroku server runs twice daily and sends an automatic email message to any users that have new, unread responses. The message contains a link directing users to the response panel and the new messages they have received.

The response panel itself features a button-based navigation bar that lets users switch between *support*, *debug*, and *reframe* responses (Figure 4.8). The support button is selected by default and support messages are always shown first when a user visits the
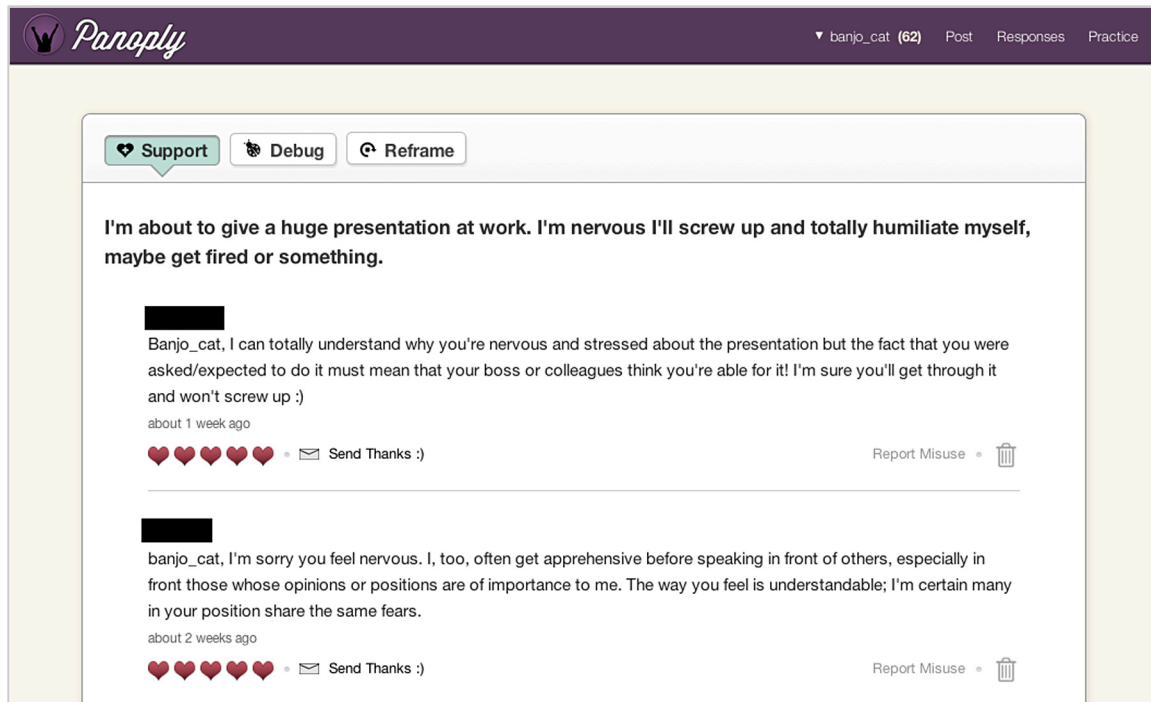
Figure 4.8: A screenshot of the response panel. By default, users are shown support messages first. They can switch between different responses by selecting the tabs at the top of the panel. Respondents in this screenshot were MTurk workers and their user names are redacted. "Banjo_cat" is an administrator account, and is not redacted from the interface.

response panel. Support messages offer a socioaffective cushion of sorts, helping make subsequent cognitive-based responses land a bit softer. Recontextualizing a negative thought takes effort, even if a crowd is doing much of the legwork for you, and it might be abrasive to have an interface deliver these responses first. As John Dewey wrote, "The path of least resistance and least trouble is a mental rut already made. It requires troublesome work to undertake the alternation of old beliefs." To make this process a bit easier, support responses are always shown first. The hope is that 'support' messages help users feel understood and comforted before they are challenged to rethink their interpretations.

As with the practice panel, the debug and reframe responses are locked at first (Figure 4.9). Users are told they can 'unlock' these additional responses as soon as they complete
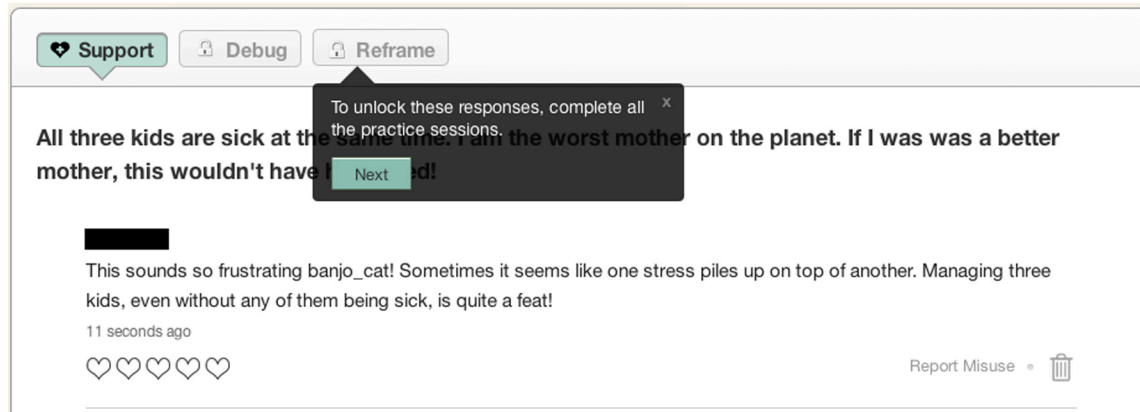
Figure 4.9: When users are first introduced to the system, they are only given 'support' responses. Additional responses remain locked until the user progresses further within the platform.

the relevant practice modules. This design feature nicely illustrates how social exchanges can motivate participation in an application. Users are likely to be curious about the hidden responses they've received, and so they should be willing to work a bit to unlock them. Further, this design strategy also prevents users from seeing responses they do not yet understand. Seeing a list of cognitive distortions, without knowing anything about the concept first, would be a confusing and frustrating user experience. The *debug* panel is therefore locked until users have had the relevant training in that practice module.

The layout of the response page depends on which response category is selected. The sections that follow describe how responses are presented from each of the three categories.

Support Responses

When users first open the response panel, they are given a short, guided tour of its various features. Users are told they can rate each response, by assigning it 1 to 5 hearts. A 5-point Likert scale such as this is not typically used in social question-answer sites, but it was employed on our platform to help facilitate data analysis in the randomized controlled trial. If a response garners three or more hearts, an option appears to send a

This sounds so frustrating banjo_cat! Sometimes it seems like one stress piles up on top of another. Managing three kids, even without any of them being sick, is quite a feat!

11 seconds ago

❤❤❤❤❤ • ✉ Send Thanks :)   Report Misuse • 🗑
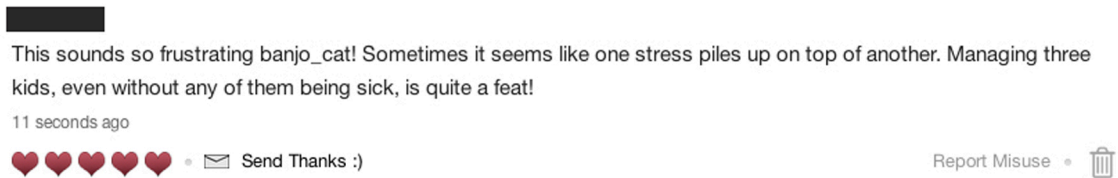
Figure 4.10: A screenshot of a *support* response. Users can rate responses and send thank you notes. If a particular response is abusive or off-topic, it can be reported.

thank you note (Figure 4.10). This affords users a chance to thank respondents personally, if the mood should strike them. Each response also has an option to 'Report Misuse', in case any of the responses are malicious or inappropriate. Responses are presented in a newsfeed format, with the most recent appearing at the top (see Figure 4.8).

Returning to our user scenario, let us consider a support statement that was composed in response to Jenny's dilemma about being single at an age when many of her friends are engaged or married. Keep in mind her *Panoply* username is FortuneCookie and that these responses are from real people using the system.

> FortuneCookie, I'm sorry to hear you're going through a tough time. It is totally normal to feel this way when your friends around you seem to be moving on with their lives and you feel stagnant. It's better to find the right person than rush into something.

This response offers empathy and describes a shared experience. No efforts are made to minimize the situation or attempt to fix it. Jenny gave this response 4 out of 5 hearts and composed the following thank you note:

> Thank you! You are totally right. I keep reminding myself it is better to wait then to rush and regret. Thanks again for your understanding words.

Debug Responses

Responses in the debug panel are presented in graphical form (see Figure 4.11). Receiving this feedback seemed somewhat judgmental for some of our pilot users, so extra effort was taken to assure people that distortions are normal occurrences and happen to everyone. The first time users see the debug panel a large disclaimer to this effect is provided.

In the debug panel, users are shown a depiction of the type and frequencies of distortions identified by the crowd. They can hover over each distortion to learn more about it and to discover ways to dispute it (see Figure 4.11).
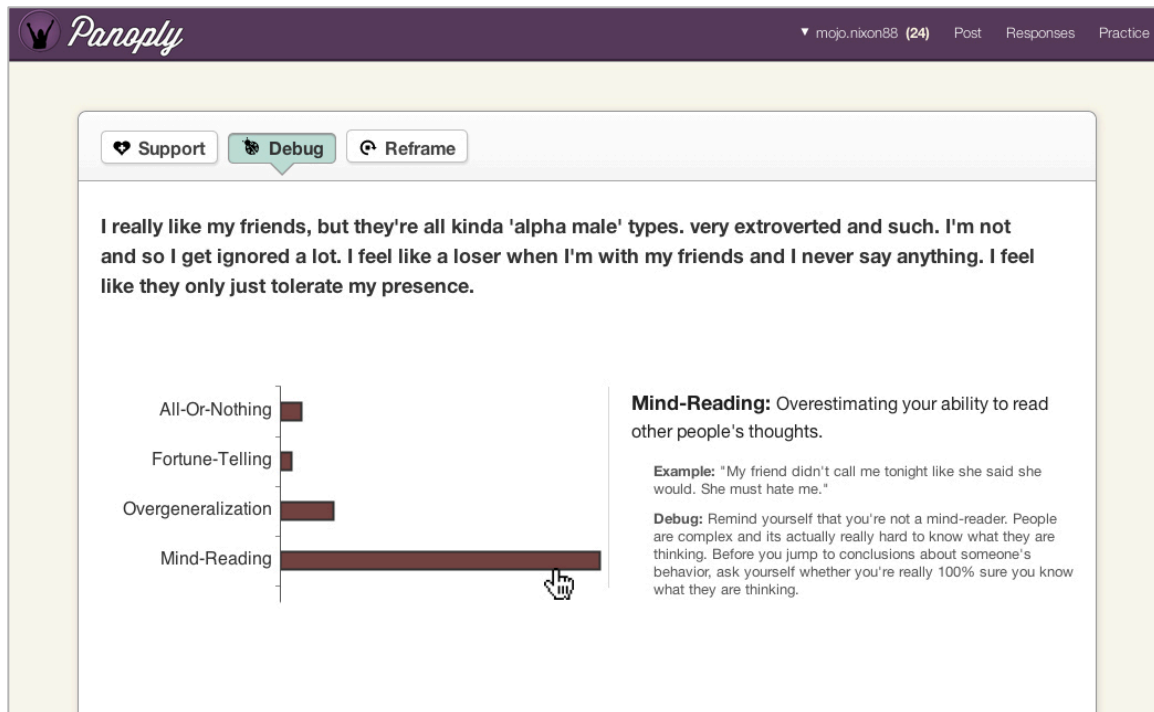
Figure 4.11: A screenshot of the *debug* panel. Distortions identified by crowd workers are presented on the left of the interface. Users can hover over these distortions to discover ways to dispute them.

Reframe Responses

The *reframe* panel is very similar to the support panel, but includes one important difference: the responses aren't revealed outright. Instead, users are asked to compose a

You have reframes pending!
Want to see what people wrote? Unlock responses by submitting your own reframe first.

> *Compose a quick reframe for yourself...*

Unlock Responses

Figure 4.12: Before receiving new reappraisals from the crowd, users must first compose one for themselves.

reframe for themselves before they can access responses from the crowd. In most CBT practices, alternative interpretations of stressful events are not usually fed directly to the patient. Rather, the therapist might engage in 'Socratic questioning' – a practice that involves asking pointed questions to help patients tease out new interpretations on their own. One problem with crowd feedback is that it sidesteps this process and may prevent users from exercising the techniques on their own and coming up with their own conclusions. Crowd feedback provides important emotion regulatory assistance and offers an engaging way to learn new techniques, but it does not explicitly ask users to exercise the techniques on their own.

By asking users to first create reframes for themselves, *Panoply* helps users understand that successful reappraisals can come from within as well as without. There is, of course, a risk that withholding information in this way will frustrate users. However, this was not observed in pilot tests or in the randomized controlled trial.

As in the support panel, users can rate reframes returned by the crowd. One of the top-rated responses for Jenny's dilemma was composed by an MTurk worker. It reads as follows:

> A good thing to remember is that with that chapter of your life ending, a new chapter is beginning. Starting over can be scary but it's also a great opportunity. The fact that you were able to step out of your relationship into an uncertain

future says that you are a strong and brave individual ready to experience new things that life has to offer.

This response was especially potent for Jenny. She was moved to composed the following thank you note:

> Wow! Thank you for your response. It about brought me to tears. As much as I don't feel strong or brave right now I realize now how strong and brave I had to be to make that decision. Thank you for helping me realize this.

4.3.5 Points System

Much like other Q&A sites, *Panoply* has a points system that helps users track the quality of their contributions. The points system is designed to be subtle and the site does not highlight its importance. A large emphasis on getting points, perhaps to earn prizes or to earn reputation on the site, could adversely change the incentives for participation. Adding as many positive incentives as possible will not necessarily produce a monotonic rise in engagement; human motivation is not always additive (see Shirky, 2011). In some cases, extrinsic rewards like points can 'crowd out' intrinsic rewards like the pleasure of helping others. However, it is also important that users get predictable and regular feedback on their performance (Zichermann, 2011) and so a subtle point system was implemented, not unlike what is used on sites like *Quora*.

Points on *Panoply* are awarded anytime a user completes a practice module or composes a well-received response. Because this user base is likely to be sensitive, and because the perceived quality of a given response is highly subjective, feedback is only awarded for the top-rated responses. Low scores are not displayed to the user and are instead cloistered away in the database and used only for analysis purposes.

Thank you notes

When users receive thank you notes, they are sent an email notification with the subject line: "Someone thinks you're awesome!" The email links users to their points panel on *Panoply*. The points panel displays scores for top-rated responses and it shows users how

far they've progressed in the practice panel (Figure 4.13). If a response has a thank you note, users can click on it and see the entire communication.

Figure 4.13: A screenshot of the points panel. Users are shown a list of posts they've responded to, along with any points they have received. Responses in the points panel can be expanded to show the entire communication.

4.3.6 MTurk Workers

Hybrid Crowd Design

So far, I have only examined the system from the perspective of users like Jenny, a registered, volunteer *Panoply* user. But, as specified throughout this document, paid

crowd workers are also recruited to both create and curate site content. Workers from MTurk are asked to review posts, compose responses, and rate responses from other workers. Since MTurk workers need to be paid, it is reasonable to question their use at all. Why can't *Panoply* just be a peer-to-peer system comprised entirely of unpaid volunteers? Incorporating multiple populations certainly costs more, both in terms of money and development time. However, when done carefully, a commingling of crowds can offer great advantages, especially on platforms that do not yet have large user bases.

On *Panoply*, MTurk workers fill in the cracks as needed, performing tasks that are unappealing to unpaid volunteers or stepping in when no other site users are available. MTurk provides an elastic user base that can grow or shrink as needed. While the number of logged in *Panoply* users may be small, the pool of MTurk workers is always large and can always be tapped when necessary. Having a huge on demand workforce in reserve helps create the appearance of a popular, vibrant social system (when in reality only a handful of registered *Panoply* users might be active at any given time). With MTurk, *Panoply* can simulate the quick, interactivity found within large, successful Q&A sites like *StackOverflow*, *Quora*, and *Whisper*, to name a few.

Reviewing Posts

As described previously, MTurk workers examine each post submitted to *Panoply*. They are asked to determine whether it contains any profanity or any suggestions of self-harm. This is a very quick, simple task and the payment is only $.01.

Composing Responses

MTurk workers are also recruited to compose responses for other *Panoply* users. Ideally, responses from MTurk workers should be as good as those coming from registered *Panoply* users. Yet, *Panoply* users and MTurk workers are likely to differ in several key respects, including expertise, English fluency, and demographics (e.g., age and nationality). Most notably, though, they differ in terms of motivation. While MTurk users report varying motivations for performing crowdwork, chief among them is the desire to make money (Ipeirotis, 2010). This cannot be said for users on *Panoply*. By contrast, *Panoply* respondents are principally motivated by the desire to learn new techniques or to

help others. While some MTurk workers may also be driven by similar, non-pecuniary motivations, the fact remains that many of them will only be interested in making money. A reliance on extrinsic motivations such as money can sometimes lead to poor performance (Deci, 1971). Some workers may try to fake their qualifications in order to access the task. Others might not cheat outright, but might put forth as little effort as possible in order to get paid. To manage these issues, MTurk workers must pass several quality assurance requirements that are not required of unpaid *Panoply* users.

Quality Assurance

In pilot tests of this system, many of the MTurk responses were poorly constructed and showed very little mastery of the English language. To address this issue, various qualifications were put in place. First, before accessing any of the *Panoply* tasks, all MTurk workers must successfully answer three SAT verbal questions. Their IP address must also originate from within the United States[5]. Once these fairly strict procedures were put into place, there was a dramatic improvement in the quality of the responses.

A voting stage was also introduced, not unlike the *verify* procedure described by Bernstein et al. (2010). After an MTurk worker composes a response, three additional workers are asked to rate it. At first, workers were asked to assess the overall quality of each response using a 7-point Likert scale. Unfortunately, this task was far too subjective. Workers could not seem to agree on the quality of responses. To make the task easier and more reliable, workers were simply asked to indicate whether a response was (a) confusing, (b) rude or (c) poorly-written. Framed thusly, agreement on the rating task increased considerably. Workers seemed to all agree on which responses needed to be filtered out of the system.

---

[5] U.S. workers are more likely to be native English speakers than other countries represented on MTurk (e.g., India).

4.4 Conclusion

In this chapter, I outlined various design principles that governed the development of the *Panoply* platform. I described the components of the platform and discussed the rationale behind each design decision.

Once the complete *Panoply* system was built and tested for usability, it needed to be evaluated for user experience and therapeutic efficacy. Specifically, I wanted to know whether people would use the system repeatedly, without being paid directly to do so. I also wanted to know whether repeated use of the system might significantly reduce negative mood, increase positive mood, and change self-reported emotion regulatory habits. Finally, I wanted to test elements of the system design itself, such as the speed, cost, and quality of the crowdsourced responses.

Chapter 5

Psychological Outcomes

The evaluation of this thesis spans three lines of inquiry. First, I assess *Panoply*'s effect on psychological outcomes, such as subjective happiness, depression, perseverative thinking, and self-reported use of emotion regulation strategies. Next, I evaluate the user experience of the system, with an emphasis on behavioral measures of engagement, such as activity level. Finally, I examine various aspects of the system design, such as the cost, speed, and quality of the crowdsourced contributions. This chapter addresses psychological outcomes, specifically, and describes the experimental procedures of the study. To best evaluate this dissertation along all three lines of inquiry, I deployed the system for two months and conducted a randomized controlled trial (RCT) with 166 participants.

5.1 Method

5.1.1 Participants

Participants were recruited through various online channels, including social media posts, Craigslist advertisements, psychology bulletin boards, and email mailing lists. Flyers were also placed around college campuses in the Boston area. The recruitment materials advertised the study as a three-week, online stress reduction intervention. To meet the

inclusion criteria, each participant needed to be a native English speaker between 18 and 35 years old. They also had to have access to a tablet or PC running an updated web browser (such as Chrome, Safari, or Firefox). The age requirement was used to reduce variability in the sample. Younger individuals are more likely to have experience using online social networking platforms and should find the *Panoply* system self-explanatory (therefore requiring fewer interactions with the experimenters). All aspects of this study, including participant recruitment, experimental procedures, and data collection, were approved by the Institutional Review Board at the Massachusetts Institute of Technology.

Participants were not paid directly for participation in the study. Instead, all participants who completed the baseline and follow-up assessments were offered a chance to win an iPad Mini (valued at $300). Use of the technology itself was optional and was not a factor for reimbursement. Additionally, all study instructions were disseminated online, through email, with zero face-to-face interaction with the experimenters. These procedures helped enhance the ecological validity of the experiment and helped better assess how *Panoply* might be used as an open access application, where adherence is not influenced by reimbursement or oversight from experimenters.

466 individuals expressed interest in the study and were given an online consent form explaining the study procedures as well as a pretest questionnaire which assessed demographic variables and baseline psychological variables (e.g., baseline depression symptoms). 270 individuals completed the consent form and the assessment. Five were not native English speakers and were therefore excluded from the study. Individuals with scores of 16 and higher on the Center for Epidemiologic Studies Depression (CES-D) scale at baseline were sent a follow-up email with links to mental health resources. They were told they could continue to participate in the study, but they were reminded to seek formal mental health resources at any time, should they need them. After completing the consent form and baseline assessment, participants were randomly assigned login credentials for either the control or treatment application. 217 individuals activated their accounts and, of these, 166 completed the final follow-up assessments. Of the individuals that dropped out of the study, three said they were not interested in reducing their own stress but were instead just curious to try new Media Lab technology. Another individual

reported not having enough time to participate. The rest could not be reached for comment.

```
                    ┌─────────────────┐
                    │   Enrollment    │
                    └─────────────────┘
                            │
            ┌───────────────────────────────────┐
            │ Expressed interest in study (n=466)│
            └───────────────────────────────────┘
                            │
            ┌───────────────────────────────────┐
            │ Randomized                         │
            │  •  Panoply (n=236)                │
            │  •  Expressive Writing (n=230)     │
            └───────────────────────────────────┘
                            │
            ┌───────────────────────────────────┐
            │ Completed pretest questionnaires   │
            │  •  Panoply (n=137)                │
            │  •  Expressive Writing (n=133)     │
            └───────────────────────────────────┘
                            │
                            │    ┌────────────────────────────────────────┐
                            ├────│ Excluded (n=53)                        │
                            │    │  •  Declined to activate account (n=48)│
                            │    │  •  Did not meet inclusion criteria (n=5)│
                            │    └────────────────────────────────────────┘
```

Allocation

Allocated to *Panoply* (n=108)   Allocated to *Expressive Writing* (n=109)

Follow-Up

Lost to follow-up (n=24)   Lost to follow-up (n=27)

Analysis

Analyzed (n=84)   Analyzed (n=82)
• Excluded (n=0)   • Excluded (n=0)

Figure 5.1: Trial diagram illustrating enrollment patterns for the treatment group (*Panoply*) and the control group (online expressive writing).

5.1.2 Control Group

Participants were randomly assigned to either a control condition or the complete *Panoply* intervention. Group assignment was blind and was conducted using a python-generated random allocation schedule. Users assigned to the control condition were introduced to the concept of expressive writing and they were instructed to compose descriptions of negative thoughts and situations. They were told that doing so might help them contextualize their problems and give them new perspectives on challenging emotions. The web and interface design for the control condition was identical to the full *Panoply* intervention and the instructions for describing stressful situations and negative thoughts were exactly the same. These participants were given the same tutorial and the same introduction to the concept of a thought record. They were not, however, given any opportunities to participate in crowdsourced interactions. They did not receive feedback from the crowd nor were they given the chance to provide feedback to others; their primary task was to write deeply about their own negative thoughts and stressful situations.
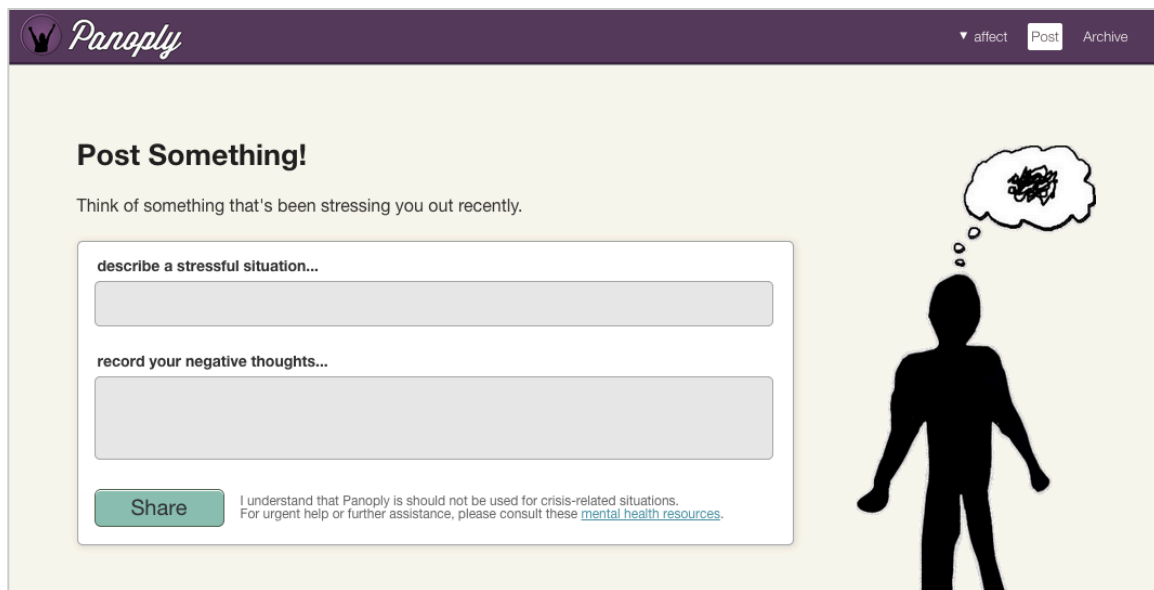
Figure 5.2: A screenshot of the control condition, illustrating its similarity to the *Panoply* platform. All nonspecific factors were matched (including the application name, the graphic design, the login procedures, and the orientation text).

Expressive writing is a well-documented intervention that has been studied for almost thirty years. Meta-analyses suggest it can significantly improve physical and psychological health outcomes (Baikie & Wilhelm, 2005). Many RCTs use inert controls that are hypothesized to confer no therapeutic benefits. But many control conditions that are thought to be inert may in fact be 'nocebos', causing participants to get worse. In a recent meta-analysis of CBT RCTs, Furukawa et al. (2014) found that people put on waitlists performed worse than those who were given no treatment. For these reasons, the present experiment matched *Panoply* with an expressive writing intervention - an active control that has purported therapeutic benefits of its own.

Expressive writing was also chosen because it differs only with respect to the active ingredients that are unique to the *Panoply* intervention. It matches *Panoply* on nonspecific factors (e.g., web design, user registration) but does not contain reappraisal training or crowdsourced interactions as active ingredients.

Finally, the expressive writing condition was also chosen to help assess engagement. It was hypothesized that a crowdsourcing platform, replete with social interactions and interactive content, would be significantly more engaging than a typical online journaling application. Online journaling applications are frequently used in positive psychology interventions, such as cultivating gratitude and reflecting on signature strengths (Seligman, Steen, Park, & Peterson, 2005) and therefore serve as a useful point of comparison.

It is worth noting that the original expressive writing procedure involved writing about traumatic life events for 15 minutes a day for four consecutive days (Pennebaker & Beall, 1986). By contrast, participants in the current experiment were asked to write about *any* negative event for a minimum of 25 minutes a week, *at intervals of their choosing*. To call this an expressive writing task might be disingenuous if it differs substantially from the accepted format. Yet, there doesn't seem to be one singular, agreed upon format for expressive writing interventions. Since the original procedure was published, hundreds of follow-up studies have been conducted, many of which involve considerable variation on the original paradigm. On the whole, there is little evidence to suggest that expressive writing procedures need to perfectly match the original protocol. In a recent review, Smyth and Pennebaker (2008) note that 'positive effects accrue if people write on three

occasions over a single hour or even as briefly as for 5 minutes on different days" (Smyth & Pennebaker, 2008). Further, they argue that expressive writing needn't focus on traumatic life events. Participants can write about all kinds of experiences, positive and negative, traumatic and quotidian, and still experience beneficial health outcomes. Given the loose boundary conditions of this procedure, it seems that writing about negative thoughts and situations for 25 minutes a week can be described as a form of expressive writing. Hereafter, the control condition will be referred to as an expressive writing intervention.

5.1.3 Procedure

The overarching study design was the same for participants in both the expressive writing intervention and *Panoply* groups. After obtaining login information, participants created an anonymous account. They were reminded to use the application for 25-minutes per week, for three weeks. They were told they could use it as frequently as they liked, so long as the total amount of the time they spent matched or exceeded 25-minutes per week. To best approximate real usage with an unmoderated application, participants were not given any further instructions about how to use the system. Instead, participants were told to use the application in ways that best fit their schedules and interests. Participants in both groups received four automated emails throughout the study reminding them to use their assigned application. After three weeks, participants were emailed a link to the follow-up assessments.

5.1.4 Assessments

Psychological assessments were taken at baseline and follow-up to examine any changes in positive affect, negative affect, and risk factors for depression (e.g., perseverative thinking and maladaptive use of emotion regulatory strategies). Additional, follow-up analyses examined *Panoply*'s effect on perceived benefits, from both an intra and interpersonal perspective. The psychological assessments used five empirically-validated questionnaires, as described below.

Subjective Happiness Scale (SHS; Lyubomirsky & Lepper, 1999).

The SHS is a 4-item measure of global subjective happiness. The response format is a 7-point Likert scale and the questions assess the extent to which respondents consider themselves happy individuals, generally speaking. Although designed to be very brief, the SHS nonetheless is an important tool in the study of well-being and it has high internal consistency and test-retest reliability.

The Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988).

The PANAS assesses positive and negative affect using two 10-item scales. Respondents indicate the extent to which they felt various positive states (e.g., enthusiastic, alert) and negative states (e.g., guilt, disgust) over the past week. Scores for positive affect (PA) and negative affect (NA) are computed and used in the analyses. High PA scores indicate a positive affective outcome, whereas high NA scores indicate a negative affective outcome.

Center for Epidemiologic Studies Depression Scale (CES-D; Radloff, 1977).

The CES-D is a 20-item self-report scale that assesses symptoms of depression. Respondents are asked to indicate the extent to which they've felt various depression symptoms over the past week. The questions address symptoms such as loss of appetite, depressed mood, and feelings of loneliness. A score of 16 or higher suggests a high level of depression and is often used as a cut-off to determine clinically relevant symptoms.

The Emotion Regulation Questionnaire (ERQ; Gross & John, 2003)

The ERQ assesses individual differences in the habitual use of two emotion regulation strategies: cognitive reappraisal and expressive suppression. Cognitive reappraisal, as discussed at length in this dissertation, is a strategy that involves reinterpreting the meaning of a thought or situation to change its emotional trajectory. Expressive suppression involves inhibiting an emotional response by stifling behavioral and experiential aspects of emotional experience. Reappraisal is considered an adaptive

regulatory strategy and is associated with positive psychological functioning, including increased positive affect, well-being and interpersonal functioning (Gross & John, 2003). Suppression, by contrast, is considered maladaptive and is associated with negative affect, reduced well-being, and poor interpersonal functioning. The ERQ is a 10-item questionnaire that produces scores for the self-reported use of cognitive reappraisal and expressive suppression.

Perseverative Thinking Questionnaire (PTQ: Ehring et al., 2011)

Depressive rumination is a cognitive style that involves repetitive elaboration of the symptoms and causes of distress. In essence, it is an unproductive form of reappraisal. Instead of recasting a situation in ways that lead to positive recontextualizations and problem solving insights, depressive rumination mires individuals in circular reinterpretations that only serve to magnify distress. Rumination is considered a risk factor for depression and suicide and is thought to play a causal role in the development and maintenance of depressive illness (Nolen-Hoeksema, 2000). The PTQ is a 15-item scale that assesses three components of rumination: its repetitiveness, its unproductiveness, and its tendency to capture mental capacity.

All participants who completed the follow-up questionnaires were included in the analyses, irrespective of whether they used the intervention as suggested (i.e., 25min/week for three weeks). Statistical analyses were conducted to examine differential changes across the various outcome measures: SHS, PA, CES-D, NA and depression risk factors (ERQ-Reappraisal, ERQ-Suppression, PTQ).

Additionally, separate analyses were conducted for the entire cohort and for the subset that scored 16 and higher on the CES-D (i.e., participants with clinically relevant levels of depressive symptoms at baseline). A median split was also performed on the baseline reappraisal scores to compare outcomes between low reappraisers and high reappraisers. Since *Panoply* aims to train reappraisal, first and foremost, it was hypothesized that it would be of most benefit to individuals who reported deficiencies in this skill at baseline (i.e., low reappraisers).

5.2 Results

5.2.1 Baseline Analyses

Analyses confirmed no significant difference in follow-up completion between participants assigned to the treatment vs. control tasks ($\chi^2 = .03$, $p = .87$). Of those included in the final analyses, there were no significant differences in age ($t_{163.07} = .41$, $p = .68$) or gender ($\chi^2 = .07$, $p = .79$) between the two experimental groups. The mean age was 23 years old. 119 females and 47 males completed the study. A full account of the demographic characteristics, including levels of education, are provided in Table 5.1. T-tests were conducted to examine potential differences between the control and treatment groups on each psychological variable at baseline. None of the tests were significant.

Analyses were also conducted to determine whether there were any differences at baseline between participants who dropped out vs. those who completed their posttest assessments. T-tests compared baseline scores for drop-outs and completers across all psychological outcome variables. None of these between-group tests were significant.

5.2.2 Full Cohort

Analyses were first conducted across the entire sample of participants, irrespective of baseline depression or reappraisal scores. It was hypothesized that, compared to the expressive writing condition, *Panoply* would yield better outcomes across all psychological variables.

A 2 (Group: Panoply, Writing) × 2 (Time: Pre, Post) repeated-measures analysis of variance (ANOVA) was conducted for each outcome variable. Follow-up within group t-tests were also conducted to examine changes from baseline to follow-up for both the *Panoply* condition and the writing condition (see Table 5.2).

Baseline Characteristics

	Panoply (n=84)	Writing (n=82)
Females, n (%)	62 (73.8%)	62 (75.6%)
Age, M (SD)	23.51 (5.20)	23.85 (5.48)
Education, n (%)		
Doctoral Degree	3 (3.6%)	2 (2.4%)
Masters	13 (15.5%)	11 (13.1%)
4-Year Degree (BA, BS)	21 (25%)	26 (30.1%)
2-Year Degree (Associates)	1 (1.2%)	2 (2.3%)
Some College	32 (38.1%)	36 (42.9%)
High School / GED	14 (16.7%)	5 (6.1%)
Baseline Positive Affect, M (SD)		
Subjective Happiness Scale (SHS)	16.85 (5.17)	17.45 (5.38)
Positive Affect (PANAS-PA)	29.77 (7.56)	30.45 (7.58)
Baseline Negative Affect, M (SD)		
Center for Epidemiologic Studies Depression Scale (CES-D)	19.38 (10.16)	18.55 (10.61)
Negative Affect (PANAS-NA)	23.45 (6.83)	23.57 (6.63)
Baseline Risk Factors, M (SD)		
Cognitive Reappraisal Frequency (ERQ-R)	25.99 (6.91)	26.74 (6.66)
Expressive Suppression Frequency (ERQ-S)	14.23 (5.16)	14.52 (4.82)
Perseverative Thinking Questionnaire (PTQ)	46.76 (10.70)	48.23 (11.11)

Table 5.1: Baseline characteristics for individuals assigned to the *Panoply* and expressive writing conditions. Differences in baseline scores between the treatment and control groups were not significant.

Findings from the full cohort revealed a significant interaction of group by time for cognitive reappraisal ($F_{1,160} = 5.33$, $p < .03$, $d = .36$). There were no significant

interactions of group by time for other risk factors (expressive suppression and perseverative thinking) or for the negative and positive affect measures (see F values in Appendix, Table 5.1).

All outcome measures improved over time for both groups; however, the gains were consistently stronger for *Panoply*. Follow-up within-group t-tests (Welch) showed significant increases in self-reported reappraisal use for the *Panoply* condition ($p <$ *.005,*), but not for the expressive writing condition ($p > .58$). There were also significant improvements in depression symptoms over time for *Panoply* ($p < .02$), but not for expressive writing ($p > .18$). Both groups showed significant improvements over time for perseverative thinking, however the effect was stronger for *Panoply* ($d = .41$) than for expressive writing ($d = .33$).

Within Group Comparisons (Full Cohort)

Response Variables	Group	PreTest (SD)	PostTest (SD)	*t-value*	*p-value*	*d* [95% CI]
Positive Affect						
SHS	P	16.85 (5.17)	17.65 (5.24)	1.01	.31	.16 [-.15, .46]
	W	17.45 (5.38)	17.49 (5.40)	0.04	.97	.01 [-.3, .32]
PANAS-PA	P	29.77 (7.56)	31.10 (8.21)	1.09	.30	.17 [-.14, .47]
	W	30.45 (7.76)	31.02 (7.93)	0.47	.64	.07 [-.24, .38]
Negative Affect						
CES-D	P	19.38 (10.16)	15.79 (9.53)	2.36	**.02**	.36 [.06, .67]
	W	18.55 (10.61)	16.33 (10.38)	1.35	.18	.21 [-.1, .52]
PANAS-NA	P	23.45 (6.83)	22.11 (7.86)	1.18	.24	.18 [-.12, .49]
	W	23.57 (6.63)	22.12 (7.03)	1.36	.18	.21 [-.1, .52]
Risk Factors						
ERQ-R	P	25.99 (6.91)	28.92 (6.23)	2.89	**.004**	.45 [.14, .75]
	W	26.74 (6.66)	27.32 (6.78)	0.55	.59	.09 [-.22, .4]
ERQ-S	P	14.23 (5.16)	13.73 (5.32)	0.62	.54	.10 [-.21, .4]
	W	14.52 (4.82)	14.12 (5.07)	0.52	.6	.08 [-.23, .39]
PTQ	P	46.76 (10.70)	42.35 (11.04)	2.63	**.009**	.41 [.1, .71]
	W	48.23 (11.11)	44.21 (13.12)	2.12	**.04**	.33 [.02, .64]

Table 5.2: Within group comparisons for *Panoply* and expressive writing, illustrating means and standard deviations across all outcome measures.

SHS = Subjective Happiness Scale; PANAS-PA = Positive and Negative Affect Scale, Positive Affectivity; PANAS-NA= Positive and Negative Affect Scale, Negative Affectivity; CES-D = Center for Epidemiologic Studies Depression Scale; ERQ-R = Emotion Regulation Questionnaire, Reappraisal Frequency; ERQ-S = Emotion Regulation Questionnaire, Suppression Frequency; PTQ = Perseverative Thinking Questionnaire.

Psychological Outcomes (Full Cohort)
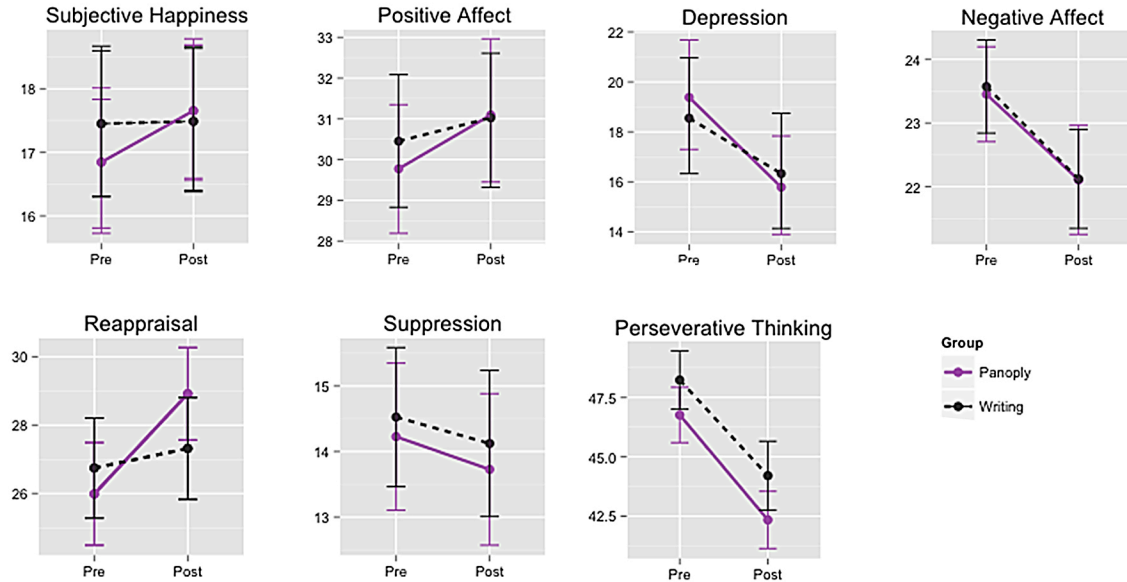
Figure 5.3: Interaction plots comparing the two treatment groups across time. Error bars reflect 95% bootstrapped confidence intervals. The *Panoply* group outperformed the writing group on all measures and there was a significant group x time interaction for self-reported reappraisal use.

Discussion

The difference in performance between the two interventions was only significant with respect to cognitive reappraisal. It is worth noting, however, that (a) all outcome variables trended in the direction of the hypothesis and (b) expressive writing is widely thought to be therapeutic on its own. Had *Panoply* been compared to a less active control condition, such as a wait-list or an inert activity (e.g., making a list of daily activities), larger between-groups effects may have been observed.

The difference in reappraisal outcomes is an important finding. *Panoply* appears to be especially effective at increasing reappraisal use – something the platform was specifically designed to do. Cognitive reappraisal is an adaptive emotion regulatory technique and is thought to be a protective factor against increases in depressive symptomology (Troy, Wilhelm, Shallcross, & Mauss, 2010). A platform that

significantly enhances cognitive reappraisal could help reduce the incidence of depression. Significant changes in depression symptoms and perseverative thinking, as observed in the *Panoply* group, could also help prevent relapses. Future work is needed to investigate these possibilities.

Interestingly, *Panoply* showed no significant benefits over time with respect to subjective happiness or positive affect (see Table 5.2). This is perhaps surprising, as *Panoply* offers many opportunities to help others, and acts of kindness are often thought to increase positive affect and well-being. However, previous research on acts of kindness has typically examined the effects of real-world, offline behavior (Buchanan & Bardi, 2010; Lyubomirsky et al., 2011). Interactions on *Panoply* are anonymous and entirely conducted online. It may also be the case that more immediacy is needed to garner positive effects. Simply replying to an anonymous stranger online may not be sufficient.

5.2.3 Depressed Individuals

The previous analyses examined the entire cohort of participants, irrespective of any depression symptoms they reported at baseline. However, individuals with high CES-D scores may be especially likely to benefit from the *Panoply* platform. To assess this hypothesis, depression classification was added as a categorical variable to the ANOVA models. Participants with CES-D scores of 15 and higher at baseline were classified as depressed, while the rest were classified as nondepressed.

Univariate ANOVAs were conducted for all outcome variables and the effect of interest was defined as the interaction of Group (Panoply, Writing) × Time (Pre, Post) × Depression Classification (Depressed, Nondepressed). Significant interactions were observed for depression symptoms ($F_{1,158} = 4.72$, $p < .04$) and perseverative thinking ($F_{1,158} = 4.98$, $p < .03$). No other interaction effects were significant (see Appendix B.1 for a complete table of F values).

Examining the depressed cohort by itself (n=91), follow-up within group t-tests revealed significant improvements in the *Panoply* group for several outcome measures, including CES-D ($p < .001$), negative affect ($p < .05$), cognitive reappraisal ($p < .008$),

and perseverative thinking ($p < .001$) (see Table 5.3). Effect sizes ranged from moderate to large, with the largest effects observed for pre-post scores on the CES-D ($d = .97$) and PTQ ($d = .73$). Significant improvements in depression scores were also observed for individuals in the writing condition ($p < .02$), though the effect size was smaller ($d = .51$). No other changes from pre to post were significant for the writing condition.

A 50% or greater reduction in symptoms is often considered an important benchmark for treating depression (Rottenberg, 2014). 32% of *Panoply* participants crossed this threshold, achieving more than a 50% reduction in CES-D scores. By contrast, 18% of writing participants experienced this level of improvement. However, this difference between groups was not significant ($\chi^2 = 1.6$, $p = .21$).

Within Group Comparisons (Depressed Individuals)

Response Variables	Group	PreTest (SD)	PostTest (SD)	*t-value*	*p-value*	*d* [95% CI]
Positive Affect						
SHS	P	14.96 (4.83)	16.64 (5.08)	1.64	.10	.35 [-.08, .76]
	W	14.82 (4.72)	15.23 (5.18)	.38	.70	.08 [-.35, .51]
PANAS-PA	P	26.70 (7.23)	30.06 (8.79)	2.02	.05	.42 [.01, .84]
	W	26.77 (6.98)	28.32 (7.60)	.99	.32	.21 [.-.21, .63]
Negative Affect						
CESD-D	P	26.40 (7.76)	17.72 (10.01)	4.7	**.001**	.97 [.53, 1.41]
	W	26.20 (8.44)	21.14 (11.28)	2.39	**.02**	.51 [.07, .94]
PANAS-NA	P	26.34 (6.40)	22.85 (7.77)	2.31	**.02**	.49 [.07, .91]
	W	27.14 (6.49)	24.91 (7.46)	1.49	.14	.31 [-.11, .75]
Risk Factors						
ERQ-R	P	24.00 (6.57)	27.72 (6.65)	2.75	**.008**	.56 [.14, .98]
	W	24.25 (6.34)	26.20 (5.58)	1.54	.13	.32 [-.1, .76]
ERQ-S	P	14.70 (5.43)	13.85 (5.66)	.74	.46	.15 [-.26, .57]
	W	14.80 (5.08)	14.25 (5.19)	.50	.62	.11 [-.32, .54]
PTQ	P	49.83 (9.38)	42.57 (10.73)	3.47	**.001**	.73 [.3, 1.16]
	W	53.02 (9.64)	49.16 (13.01)	1.58	.12	.34 [-.09, .77]

Table 5.3: Within group comparisons for depressed individuals in the *Panoply* and
expressive writing interventions, illustrating means and standard deviations
across all outcome measures.

Psychological Outcomes (Depressed Individuals)
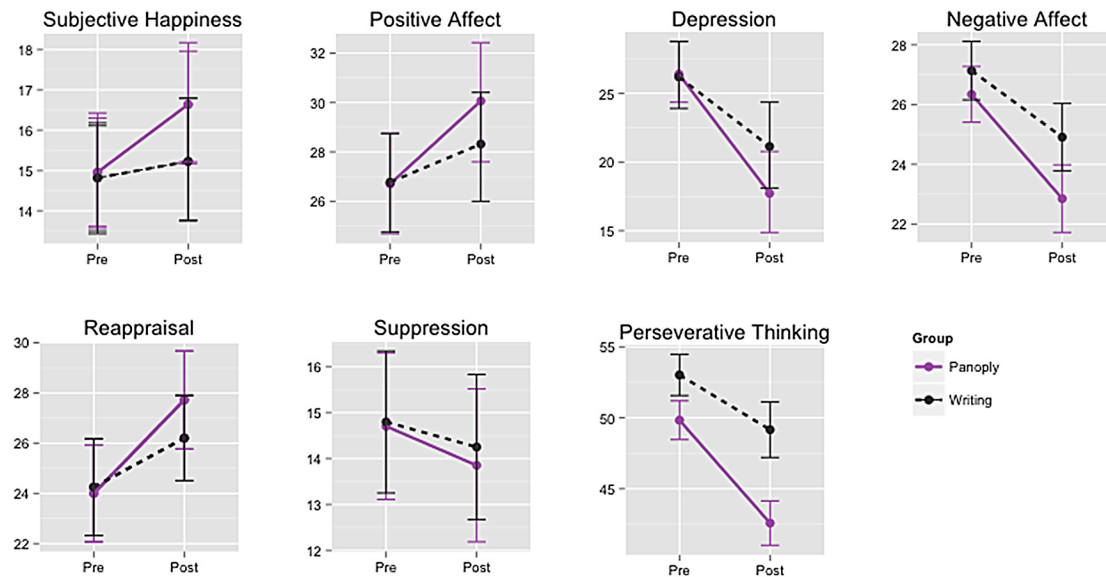
Figure 5.4: Interaction plots for the subset of individuals scoring above the clinical cut-off on the CES-D. Error bars reflect 95% bootstrapped confidence intervals.
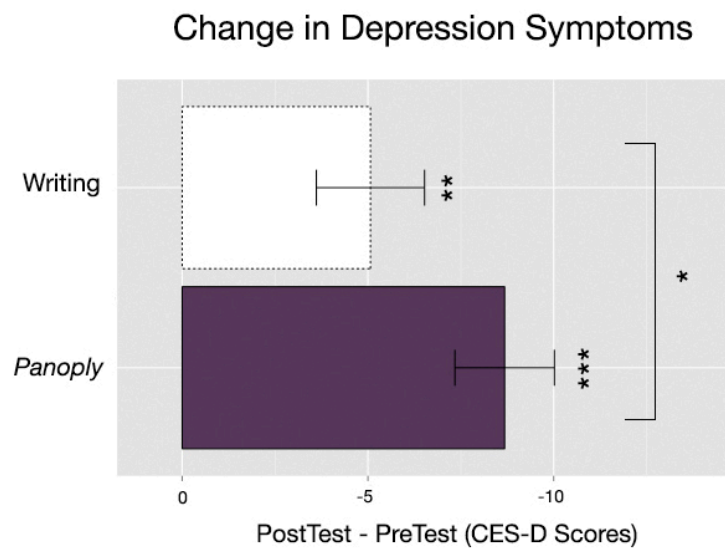
Figure 5.5: Change scores in depression symptoms for the subset of individuals scoring above the clinical cut-off on the CES-D. Error bars reflect 95% bootstrapped confidence intervals. See also Figure B.1 in the Appendix.
p < .05; ** *p* < .01; *** *p* < .001

95

Discussion

On the whole, these results suggest that individuals with elevated depression symptoms stand to benefit more from a platform like *Panoply* than from expressive writing. Compared to the writing task, *Panoply* conferred significantly greater reductions in depression symptoms and perseverative thinking. On average, individuals using *Panoply* experienced a 32% reduction in CES-D scores. Expressive writing, by contrast, yielded a 19% reduction in symptoms.

While these results are encouraging, additional research is needed before *Panoply* could be recommended as a stand-alone treatment for depression. Nonetheless, in its current state, *Panoply* might be a great ancillary treatment, something that could be used alongside traditional therapy to help further reinforce concepts like cognitive reappraisal. *Panoply* might also be a great candidate for 'booster sessions' – brief therapeutic sessions that are sometimes needed when individuals are in danger of relapse.

5.2.4 Low Reappraisers

Since *Panoply* was designed to target cognitive reappraisal first and foremost, it might be most effective for individuals with deficits in this skill at baseline. Those who already report using reappraisal frequently throughout their lives might not benefit as much from further training. To examine this, participants were classified as high or low reappraisers based on their emotion regulation scores at baseline. Since there is no commonly accepted cut-off score for reappraisal, a median split on the ERQ data was used to classify participants as high or low reappraisers. This classification was then used as a categorical variable in the analyses. Univariate ANOVAs examined the interaction between Group (Panoply, Writing), Time (Pre, Post) and Reappraisal Classification (High Reappraiser, Low Reappraiser). Significant three-way interactions were found for subjective happiness ($F_{1,158} = 3.89$, $p = .05$), depression symptoms ($F_{1,158} = 5.78$, $p < .02$), negative affect ($F_{1,158} = 6.56$, $p < .02$), and perseverative thinking ($F_{1,158} = 9.54$, $p < .003$).

Follow-up within group comparisons were conducted for the subset of participants classified as low reappraisers (n=69). In the *Panoply* group, significant improvements across time were observed for low reappraisers with respect to depression symptoms (*p* <

.001), negative affect ($p < .02$), reappraisal ($p < .001$), and perseverative thinking ($p < .001$). Low reappraisers in the writing group, by contrast, only showed significant improvements in reappraisal ($p < .005$).

Within Group Comparisons (Low Reappraisers)

Response Variables	Group	PreTest (SD)	PostTest (SD)	*t-value*	*p-value*	*d* [95% CI]
Positive Affect						
SHS	P	14.71 (4.82)	15.95 (4.78)	1.17	0.24	.26 [-.18, .71]
	W	14.29 (5.08)	14.11 (4.90)	0.13	0.89	.04 [-.51, 58]
PANAS-PA	P	26.51 (7.01)	28.85 (7.62)	1.45	0.15	.32 [-.13, 77]
	W	25.39 (6.73)	27.25 (7.82)	0.95	0.35	.25 [-.29, 80]
Negative Affect						
CES-D	P	22.93 (10.26)	15.12 (8.61)	3.73	**0.001**	.82 [.36, 1.29]
	W	25.14 (12.50)	22.57 (13.05)	0.75	0.45	.20 [-.35, .75]
PANAS-NA	P	25.02 (7.14)	21.27 (7.05)	2.4	**0.02**	.53 [.08, .98]
	W	27.04 (7.95)	26.82 (7.78)	0.1	0.35	.03 [-.51, .57]
Risk Factors						
ERQ-R	P	20.51 (4.52)	26.73 (6.82)	4.87	**0.001**	1.07 [.60, 1.55]
	W	19.39 (4.62)	23.43 (5.71)	2.91	**0.005**	.78 [.21, 1.34]
ERQ-S	P	14.24 (5.36)	13.59 (5.63)	0.54	0.59	.12 [-.32, .57]
	W	14.89 (5.20)	14.39 (4.68)	0.38	0.71	.10 [-.44, .65]
PTQ	P	50.73 (9.74)	42.66 (9.26)	3.85	**0.001**	.85 [.38, 1.31]
	W	53.32 (9.81)	52.07 (12.42)	0.418	0.68	.11 [-.43, .66]

Table 5.4: Within group comparisons for low reappraisers in the *Panoply* and expressive writing interventions.

Psychological Outcomes (Low Reappraisers)
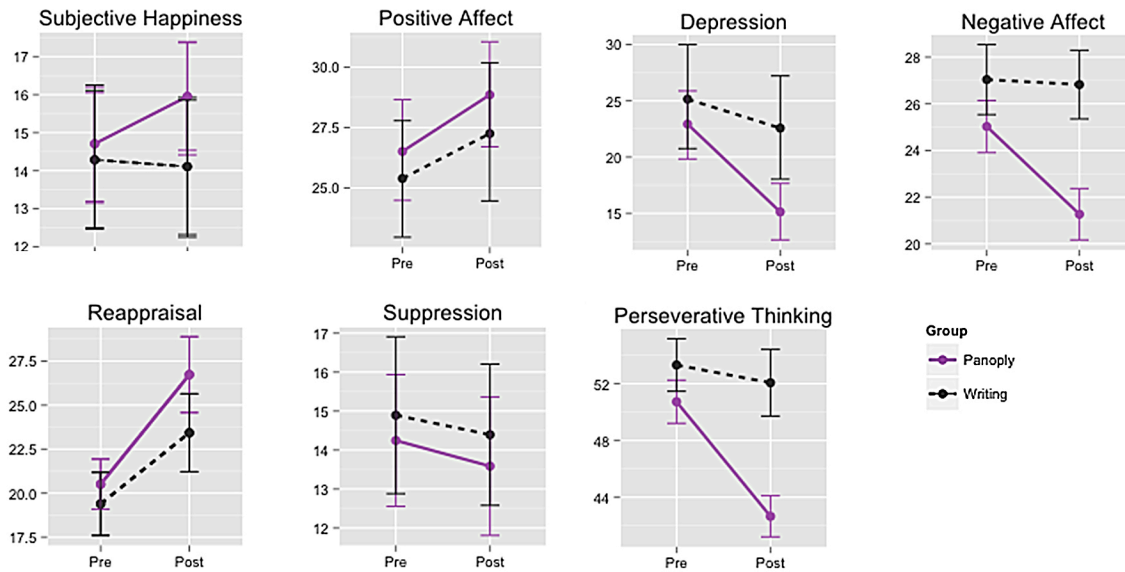
Figure 5.6: Interaction plots for the subset of individuals showing deficits in self-reported reappraisal use at baseline. Error bars reflect 95% bootstrapped confidence intervals.

Discussion

These findings suggest that *Panoply* is especially powerful for those who don't ordinarily regulate emotions by way of reappraisal. In some ways, these findings are not surprising. Because *Panoply* targets reappraisal, and since reappraisal is thought to underlie positive changes in various psychological outcomes, it makes sense that *Panoply* would be most successful for those who typically make little use of it in their daily lives.

Previous research has shown that reappraisal mediates the positive effects of various therapeutic practices. For example, Goldin et al. (2012) found that it mediates the effects of CBT on social anxiety and Garland, Gaylord, and Fredrickson (2011) found it to mediate the stress-reductive effects of mindfulness. It is possible that reappraisal also mediates the positive effects of *Panoply* on depression symptoms. Future work should examine this possibility more thoroughly.

5.2.5 Perceived Improvements

At the end of the study, participants in both experimental groups were asked to reflect on any perceived improvements they might have experienced. In the follow-up questionnaires, participants indicated the extent to which they felt improvements in their ability to manage their own stress as well as the stress of other people. They were also asked to reflect on any perceived changes in empathy towards other people. Using a 7-point Likert scale, participants were asked to indicate the extent to which they agreed or disagreed with the following statements:

- I feel more capable of managing my own stress
- I feel more capable of helping others who feel stressed
- I feel more empathetic towards others

Independent t-tests (Welch) performed across the entire cohort of participants revealed significant differences between the control and treatment groups for all measures of perceived improvements. Individuals in the *Panoply* group reported a greater ability to manage their own stress (*M*=4.71, *SD*=1.19) than individuals in the writing group (*M*=4.06, *SD*=1.33), $t_{159.38} = 3.29$, *p* < .001. Those using *Panoply* (*M*=5.35, *SD*=1.11) also felt more confident in their ability to manage the stress of others than those in the control group (*M*=3.96, *SD*=1.33), $t_{155.37} = 7.25$, *p* < .001. Lastly, self-reported empathy was significantly higher for those in the *Panoply* group (*M*=5.23, *SD*=1.18) than those in the writing group (*M*=4.21, *SD*=1.23), $t_{161.31} = 5.40$, *p* < .001. Similar findings were observed for the cohort of depressed individuals and for those with reappraisal deficits at baseline (see Appendix B.1).

Discussion

The difference in perceived improvements between the two interventions was particularly striking. In part, this may simply reflect the fact that *Panoply* yielded greater psychological benefits overall. Participants may have rightly regarded *Panoply* as the more powerful intervention and perceived it as such. It may also be that the positive effects of expressive writing are delayed and therefore harder to notice. Indeed,
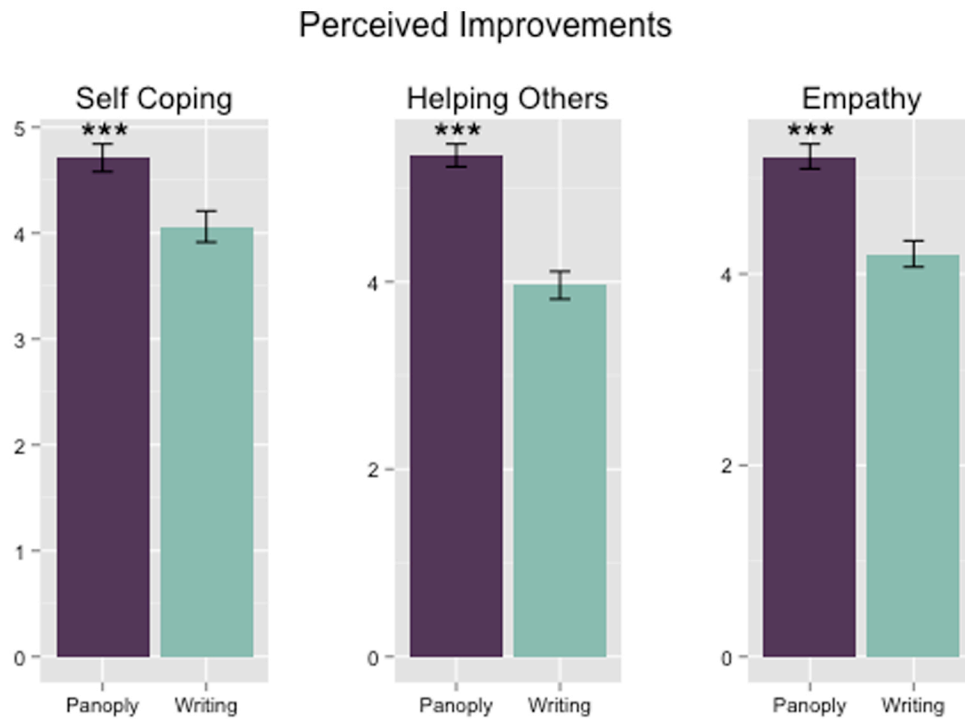
Figure 5.7: Perceived improvements reported by individuals in the *Panoply* and expressive writing groups. Error bars reflect 95% bootstrapped confidence intervals. Note: *** *p* < .001

researchers have found that expressive writing may actually increase negative emotions at first (Smyth & Pennebaker, 2008). By contrast, *Panoply* may offer a more immediate and potent flavor of emotional relief. The temporal dynamics of the two interventions may help explain the differences in their perceived effectiveness.

It is also interesting to note the perceived *interpersonal* benefits of using *Panoply*. Unlike many other computer-based interventions, *Panoply* offers its users the chance to practice therapeutic techniques on other people. Throughout, the site encourages users to compose tactful responses, using skills like active listening and shared understanding. Thus, *Panoply* users not only develop skills that they can apply to themselves, they also learn new ways to help others. This helps create an expanding virtuous circle, spreading therapeutic effects outward beyond those directly involved in the intervention. An interesting line of future research would be to examine whether *Panoply* users behave

differently when their peers are in crisis and whether, in turn, these peers report better functioning as a result.

Chapter 6

Engagement

In the field of computer-based mental health, there exists a confusing array of metrics to assess engagement. One might consider adherence (the extent to which users complete all the recommended tasks), attrition (the number of participants that drop out early), usage (the level of activity within a program), or some combination of all three. Unfortunately, these metrics do not apply equally well to all interventions. *Panoply*, for instance, was designed to be quick and easy to learn and the didactic content is limited to just three distinct modules. All three modules can be completed in about 15 minutes. Adherence rates for module completion, while a common metric for many online mental health interventions, is unlikely to be particularly informative for *Panoply*.

There is also inconsistent data on how engagement metrics affect the outcome of online interventions. The dose-response relationship may not be linear and may not always depend on common metrics of engagement (such as the number of modules completed). Donkin et al. (2013) recently examined the relationship between various engagement metrics and outcome in an online intervention for depression and found that levels of activity were more predictive of outcome than long-term engagement with the program. The total number of modules completed was less important than the level of activity observed per login. It seemed that focused attention was more important than overall adherence. This might reflect the fact that analyses of module completion or time spent using an app poorly reflect the amount of engagement with the content. Individuals who are more engaged with the actual exercises might outperform those who merely go through the assigned steps without absorbing the work.

Based on the findings from Donkin et al. (2013), and considering the unique characteristics of the *Panoply* platform, I primarily examined usage data to analyze engagement. I examined the amount of time users were actively engaged with the platform. Specifically, I compared the average number of words written by individuals in the treatment vs. control group. This is a useful metric because both the *Panoply* and the expressive writing interventions involve a considerable amount of writing. Writing is the only task activity one can perform on the expressive writing task. Similarly, with the exception of the 'debug' exercise, all activities on *Panoply* require writing. In addition to word count, I also explored other usage metrics, such as the time spent per login and the overall number of logins between groups. Finally, I examined user behaviors that suggested intent to access the system in the future.

In addition to these behavioral engagement metrics, I collected self-report data on usability and user experience. I also asked individuals to describe their subjective experiences using the platforms and to indicate what they liked and disliked about their assigned intervention.

6.1 Behavioral Data

6.1.1 Word Count

A python script was used to compute the number of words submitted by individuals in the treatment and control conditions. An independent-samples t-test (Welch) revealed a significant difference between conditions $t_{4.02}$=121.97, $p < .001$, with individuals in the *Panoply* condition writing significantly more words (*M*=1013.28, *SD*=1145.14) than those in the expressive writing condition (*M*=433.85, *SD*=609.59). This is striking, given that writing wasn't the only activity available to *Panoply* users. They could also identify distortions and review crowd-generated feedback – two additional activities that were used frequently, but were not included in the word count metric.

6.1.2 Session Data

To assess the frequency and duration of logins, 'sessions' data from Google Analytics were used. A 'session' is defined as the period of time a user interacts with a site. As of
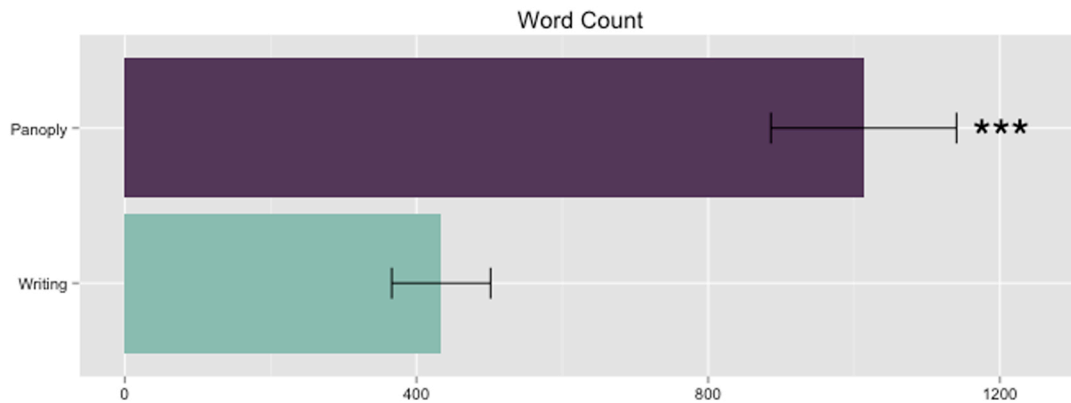
Figure 6.1: Average number of words composed by *Panoply* users and participants in the online writing condition. Error bars reflect 95% bootstrapped confidence intervals. Note: *** $p < .001$

this date, Google sessions expire as soon as a user is inactive for 30 minutes (https://support.google.com/analytics/, 2014). On average, users in the *Panoply* condition logged 21 sessions over the three-week deployment. Their average time per session was 9 minutes and 18 seconds per session. By comparison, users in the expressive writing condition logged an average of 10 sessions, spending an average of 3 minutes and 10 seconds per session.

Since Google Analytics does not reveal data at the level of the user, and since it does not report variance, inferential statistics could not be applied to these reports. Nonetheless, these descriptive statistics show that *Panoply* attracted longer and more frequent visits than the expressive writing task.

6.1.3 Usage Intent

To see whether participants might continue to access the intervention, even after the experiment was over, all accounts were kept open for an additional three weeks. Participants were also told that they were welcome to continue using the system, if they felt compelled to do so. While only 1% continued to use the writing intervention, 12% continued using the *Panoply* system well after the study was over.

Taken together, the behavioral measures described in this section suggest that *Panoply* was significantly more engaging than the expressive writing task. While not everyone was completely smitten with the *Panoply* platform, it is worth noting that many individuals were highly active users. 12.5% submitted over 50 responses and 3% submitted over 120 responses. The system clearly has the potential to attract considerable use from certain individuals.
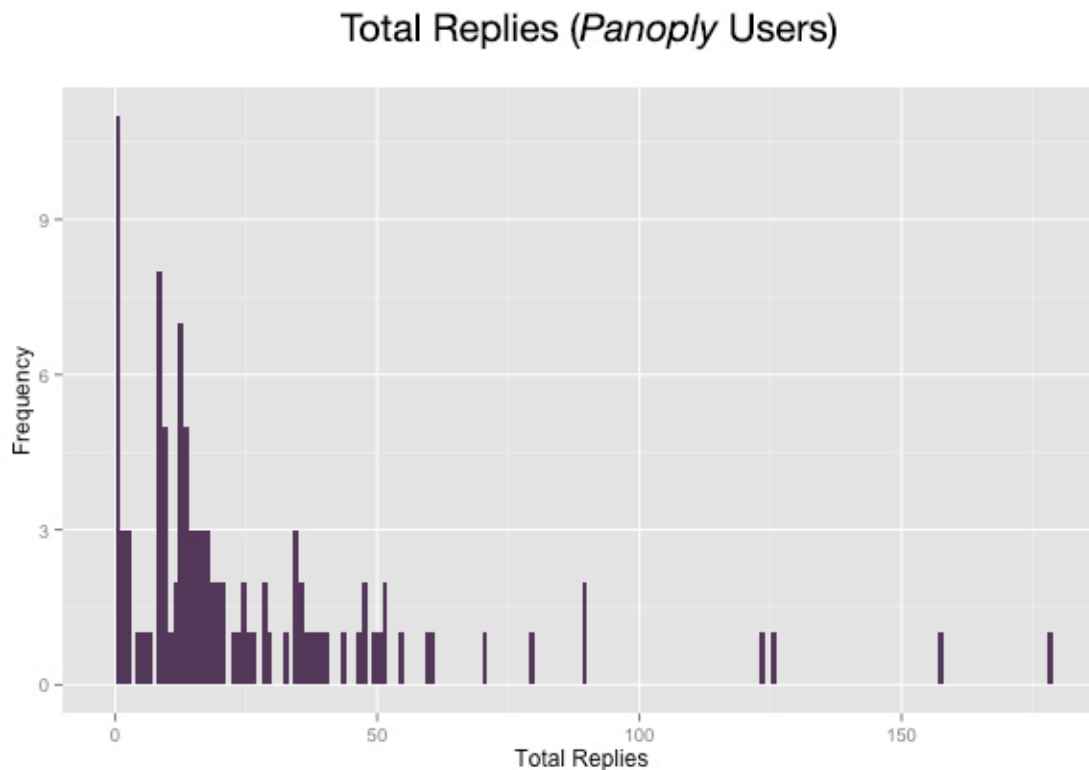
Figure 6.2: Histogram of the total replies submitted by *Panoply* users. Many individuals composed over 50 responses to the system. One individual composed 178 responses. Note: included in this graph are individuals who activated an account but did not participate further or complete the posttest questionnaires.

6.2 Self-Report Data

To examine usability and user experience, I used the System Usability Scale (SUS) and User Experience Questionnaire (UEQ), respectively. I also used a net promoter score to assess participants' likelihood of recommending the system to others.

6.2.1 Assessments

The System Usability Scale (SUS; Brooke, 1996)
The SUS uses a 5-point Likert scale and asks participants to indicate the extent to which they agree or disagree with various usability assessments of the system (e.g., 'I found the system unnecessarily complex', 'I would imagine that most people would learn to use this system very quickly'). This scale is designed to evaluate the ease with which a system is used.

The User Experience Questionnaire (UEQ; Laugwitz, Held, & Schrepp, 2008)
Usability is a necessary, but not sufficient condition for a high quality user experience. To provide excellent user experience, a system must also be stimulating, novel, and aesthetically pleasing. Thus to complement the findings from the SUS, the UEQ was also administered. The UEQ includes 26 pairs of contrasting attributes (e.g., "pleasant vs. unpleasant"; "motivating vs. demotivating") that are ordered along a 7-point bipolar Likert scale. For each test item, the Likert points represent gradations between the two labels. Selections indicate which of the two labels applies best to the technology.

Net Promoter Score
The Net-Promoter Score is based on a single question: "How likely are you to recommend this product to a friend or colleague?" Responses are arranged on a 10-point Likert scale, ranging from "not likely at all" to "extremely likely." The scale is typically used to assess customer satisfaction and is usually administered to evaluate the potential

growth of a company. Its accuracy as a predictive tool has been debated and it is not perfectly suited for the evaluation of a product on its own. Nonetheless, it is a very quick scale and was administered to provide another point of comparison between the control and treatment systems.

6.2.2 Results

Usability

The *Panoply* system is considerably more complex than the writing intervention. It contains several modules, multiple tutorials, and many different interactions between crowd workers and other registered users. At the outset, it was unclear whether participants would consider it as usable as the comparatively simple expressive writing task. Nonetheless, an independent-samples t-test revealed no significant difference in SUS scores between the writing condition (*M*=76.00, *SD*=15.17) and the *Panoply* condition (*M*=75.95, *SD*=14.38), $t_{163.01}$ = 0.02, *p* > .98. The means for both were considerably higher than the average SUS score (68) that is commonly reported in the literature (Sauro, 2011).

User Experience

Panoply was engineered to be an engaging mental health intervention. The final system incorporated many features that were specifically designed to enhance user experience. Indeed, it was hoped that many users would find the crowdsourced interactions particularly novel, motivating, and exciting. Therefore, it was hypothesized that *Panoply* would score higher on user experience than the expressive writing task. This hypothesis was confirmed. UEQ scores were significantly higher for the *Panoply* platform (*M*=137.10, *SD*=20.93) than the expressive writing platform (*M*=122.29, *SD*=20.81), $t_{163.94}$ = 4.57, *p* < .001. These results are in line with the behavioral data for engagement. It is not surprising that the platform that was used most frequently was also the one that was rated highest for user experience.

Net Promoter

It was hypothesized that individuals would prefer the *Panoply* intervention and would therefore be more likely to recommend it to a friend. A t-test revealed a significant difference between the two groups, with participants recommending *Panoply* more highly (*M*=6.35, *SD*=2.63) than the expressive writing intervention (*M*=4.10, *SD*=2.45), $t_{161.62}$ = 5.67, *p* < .001.
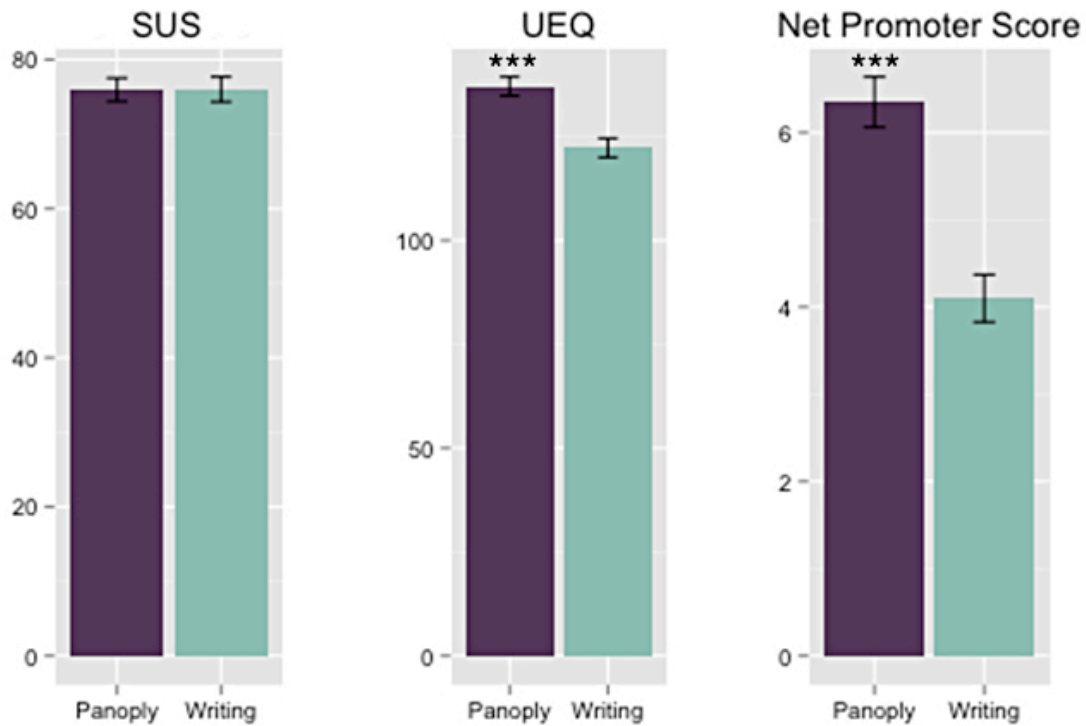
Figure 6.3: Usability and user experience ratings for the *Panoply* platform and the online expressive writing task. Error bars reflect 95% bootstrapped confidence intervals. Note: *** *p* < .001. SUS = Subjective Usability Scale; UEQ = User Experience Questionnaire.

6.3 Qualitative Feedback

The data examined thus far suggest that *Panoply* is an engaging intervention, certainly more alluring than an open-ended journaling task – something that is often used in self-

guided interventions. To follow-up further, participants were asked to describe in their own words what they liked and disliked about the two platforms.

6.3.1 Panoply

Positive Feedback

In general, the *Panoply* platform garnered the most positive comments. In fact, I received many unsolicited positive emails about the *Panoply* system, even while the study was ongoing (no such emails were received from people in the writing condition). For instance, one person wrote:

> So, I just have to say how much I like this system. With every post I am just amazed at the quality of the answers! every single time I get answers that I just didn't think of, and every time I think I am aware of most perspectives of reframing my own thoughts when I post.

Another exclaimed, *"This website really helps and would be great if everybody could use it."* Finally, another wrote, *"I would like to see this in real life (longer than a 3 week study). Maybe if all of MIT could do this it would destress the campus."*

Participants' reasons for liking the system were myriad. However, several consistent themes emerged in the comments that were received in the follow-up questionnaires. In the next sections, I highlight some of the more popular aspects of the platform.

Reappraisals

Many participants found the reappraisal support to be particularly effective. When commenting on what she liked about reappraisal, one individual remarked, *"It forced me to rethink a situation in a new way, and realize that the way I was thinking about it wasn't the only option (or the best option)."* Echoing these sentiments, another participant wrote, "*It helped me see the positive sides of my situation and helped provide hope where it originally seemed that none existed*."

Some individuals expressed enthusiasm for the way reappraisals seemed to break a cycle of negative, ruminative thinking. One individual wrote:

I'm the type of person who gets stuck in a singular channel of thinking about something, and once I think of something I get fixated on it and can't think of other possible reasons. The Reframe responses often helped me look at the situation differently, allowing me to break out of my cycle.

Others described how it often *"takes someone else's perspective and viewpoint to change your own"* and that *"re-framing shed a whole new perspective on the situation that I hadn't considered, which made me feel a lot better."* Finally, some individuals identified the ways repeated exposure to this technique was helpful. One participant wrote, *"After doing more and more reframes that mindset started to permeate into my thinking process and when I faced a difficult scenario I started thinking 'how would I reframe that if I read it on the website'".* Another person wrote that, over time:

I was able to see inconsistencies in the way that I saw the problems of others and the way that I saw my own. I realized that I wasn't really taking my own advice in terms of outlook when looking at my problems versus a similar one of someone else. I realized that it was much easier to find bugs and see the positive in the concerns of others than in my own concerns, which prompted me to try to turn that around. This was a benefit that I did not originally anticipate.

Cognitive reappraisal, while adaptive, is not the easiest emotion regulation strategy to pursue on one's own. Studies have shown that it can be hard to apply for intense emotions and that, when given the freedom to use any emotion regulatory technique, it is largely underutilized (Heiy & Cheavens, 2014; Suri, Whittaker, & Gross, 2014). In his memoir about depression, Andrew Solomon describes this problem quite elegantly. He writes, "It's hard to wrestle with your own consciousness, because you have no tool in this battle except your consciousness itself" (Solomon, 2000). But what if you have access to a *collective* consciousness? With *Panoply*, a crowd can be summoned at will to help reframe situations in ways you might never have considered. Many of the user comments about the platform support this notion.

Support

Reappraisal wasn't the only type of response participants enjoyed. Many people also enjoyed receiving support messages. Some respondents appreciated the pure simplicity of

the support statements. One person said, *"It was just nice to get honest to goodness support and understanding and validation from other people, rather than having them try to fix things."* Another wrote, *"Sometimes, I just need to complain to someone, have them listen, and I'll feel better."* Others felt that support responses were more consistently positive and less critical than the debug and reappraisal responses. Still others enjoyed the way social support exercises were structured. One person wrote, *"I found it was like learning a new useful and important life skill to become a better person."*

Social Camaraderie

A great number of *Panoply* users reported a strong sense of social camaraderie with others on the site. Many found solace in the social aspects of the system and the knowledge that they weren't alone in their struggles. For instance, one individual wrote:

> It helped me remember I am not the only one and provided a sense of camaraderie with the other users, a feeling of "we are all in this together so lets support each other and we will all get through it."

Likewise, another user noted how the platform *"enabled users to feel closer to each other. Unlike other interactive sites, this one seemed to offer a lot of camaraderie at times."* Another user enjoyed the fact that it was a *"safe place where I could go where everyone would be bought into the collective project of helping each other through stress."*

Anonymity and Accessibility

Many users also appreciated the anonymity and accessibility of the system. One user summed up these sentiments as follows: *"It's 24/7. It's anonymous and you have no limits whatsoever to what you can share. I liked the convenience of being able to access this website anywhere at any time."* Another person considered *Panoply* an inexhaustive resource, something that can be turned to at anytime. She wrote, *"I didn't feel like I had to burden loved ones, I could just get it off my chest and then interested users could respond if they wanted."*

Reception from Therapists

Two respondents assigned to the *Panoply* intervention identified themselves as mental health practitioners. Both individuals were very enthusiastic about the system and how it could help support therapeutic practice. One individual wrote:

> Overall, I found the website brilliant - a great idea and while I consider myself to be quite an insightful reflective person (I've actually done some CBT training myself and work in mental health), I was surprised at how helpful it was and how nice it felt to get the feedback from other users, and also to give feedback to them. Seeing other users' issues also helped to put mine into perspective as well.... overall, I think it is a great programme and with my background in mental health could certainly see potential for it to work with clients and just the general public as well, thank you very much!!!

Another participant, a licensed therapist, wrote the following:

> Clearly you have taken the CBT model, broken it down into digestible steps, and presented in in such an easy way. I think it was really cool to also allow for people to learn how to empathize, recognize other's cognitive distortions, and to game-ify it all with pts was all very well thought out, and probably would be effective if they continued to participate.

> As a therapist and educator, I just wanted to say, great job, and I hope in time there's a space to further develop it into a platform for anyone to use.

These comments are encouraging and illustrate how *Panoply* might be used in conjunction with mental health practitioners. Therapists could point clients toward the *Panoply* application, giving them a fun and interesting way to practice some of the CBT techniques they learn in the clinic.

Negative Feedback

Individuals in the *Panoply* system were by no means unanimously enthusiastic about all aspects of the platform. By far, the biggest problem was the fact that users were initially unable to skip posts they found too challenging. One user wrote, *"Sometimes I couldn't think of anything to say for a particular problem, but there was no way to "dismiss" that problem so I could move on to someone else's problem (which I might've been better*

equipped to help with)." As soon as this issue was discovered, a 'skip' button was implemented. Unfortunately, many participants finished the study before this fix was introduced.

Also, some individuals felt as though the platform did not provide enough training and that, with lengthier tutorials, the overall quality of responses would have been higher. One individual wrote:

> I worry that the method used here did not include enough training or feedback for user's feedback. Basically, I think their needs to be more training, more feedback, and more interaction between the original poster and the responses they are getting.

Many participants also disliked the fact that they were asked to create support, debug, and reframe responses for the same post. Several individuals found this demotivating and felt that it forced them to be redundant in their responses.

In general, participants did not find the debug task to be very interesting or informative. Users suggested it could be improved by having the crowd do more than just label the bugs. For instance, one user argued that, *"Debug was a little too simplistic maybe, since we could title the bug, but not explain 'what' the bug was."* Someone else suggested that the debug tasks should have *"... asked for you to highlight the section of the statement that needed debugging."*

Others found various aspects of the platform too challenging. One person wrote, *"I felt powerless when I read other users' issues and I had no valuable feedback, or was at a loss for what to say."* Some wished the word-counts could have been extended. Still others found it hard to reframe problems when very little context was included in the original post.

Finally, several individuals found the support messages to be trite and superficial. One individual dismissively referred to them as 'fortune-cookie' responses, referring perhaps to the clichés and Pollyannaisms that are sometimes submitted on the platform. In the future, machine learning classifiers could be trained to detect these types of responses and prevent them from being returned to users.

6.3.2 Expressive Writing

Positive Feedback

On the whole, the data collected suggests that people overwhelmingly preferred *Panoply* to the writing intervention. This does not mean, however, that the writing condition was poorly received by everyone. Indeed, many individuals found it very therapeutic to record their negative thoughts and feelings. Some felt that it offered catharsis, others a quiet time for reflection. One person argued it *"allows you to see an archive of your past thoughts/stressful situations, which could help identify patterns."* Interestingly, while the expressive writing condition did not specifically prompt users to reframe their thoughts, the activity seemed to lead many people in this direction. One individual remarked:

> I liked that it brought my negative thoughts into my conscious awareness, which helped me realize that they are sometimes unrealistic or inaccurate, and helped me change them when I could. I also found myself analyzing my negative thoughts even when I was away from the application. It was easy to understand and very easy to use.

Another user remarked *"It had never occurred to me to express my negative feelings with the goal of changing how I thought of them, and trying to be more positive. Simply realizing that that could be a goal and activity was actually a useful tool."*

Several users also applauded the simplicity of the system. They seemed to value its minimalist design and enjoyed the fact that they only had to complete one, simple task.

Negative Feedback

While some people valued the system's simplicity, others found it lacking in features and wished it offered more activities. One individual wrote, *"I found it boring to just write my stresses down."* Another participant said, *"It seemed too simple. I thought it would offer a better way to destress rather than to just post about it."* Many users reported forgetting to use the system. Others just didn't feel compelled to use it. One participant wrote, *"I couldn't get myself to use it as much as I should have. It didn't motivate me enough."*

6.3.3 MTurk Experience

Workers on MTurk were always invited to post comments about the tasks they did on *Panoply*. Initially, workers commented on areas of the tasks that needed improving. Once these issues were addressed, the comments were mostly positive. For example, one worker wrote, *"I think this is a great idea with a great purpose. I used to work for a suicide prevention organization and this is exactly what we taught."* Another wrote, *"Interesting hit.*[6] *I like the mental exercise of reframing."*

Future research should be done to more thoroughly examine the user experience from the perspective of the MTurk workers. It will also be important to consider whether these tasks influence affective outcomes for paid crowd workers. It would be interesting to see whether repeated exposure to these tasks produces any psychological benefits for this population. Lately, there have been efforts to design crowdsourcing systems that provide substantial benefits to both the requesters and the workers (Dontcheva, Morris, Brandt, & Gerber, 2014; Kittur et al., 2013). This dissertation presents a system that exemplifies this ethos. In addition to getting paid, workers on MTurk are learning powerful self-regulatory strategies that should have benefits that extend offline.

---

[6] "hit" refers to 'Human Intelligence Task (HIT)', which is how jobs are described on MTurk.

Chapter 7

System Evaluation

In this chapter, I shift the analysis away from considerations of user experience and therapeutic efficacy. Here, I look specifically at the *Panoply* system and I examine how its various features were used throughout the experimental trial. I also examine the cost of the system, as well as the speed and quality of the responses. Together, these analyses offer insights into how the system might perform were it to scale beyond the experimental cohort. Data from this section can also help inform future research and development, providing suggestions for ways to improve the system in the future.

7.1   Usage Patterns

First, I provide an overview of all the content that was submitted to both the *Panoply* and the Writing applications (see Table 7.1). These analyses are not restricted to individuals who completed the follow-up tests. Instead, user behaviors are analyzed from everyone who activated an account and accessed the site (N=214).

Of note is the fact that individuals assigned to the writing condition submitted considerably more posts than those in the *Panoply* condition. There are several possible explanations for this. First, those in the writing condition had only one task to do. Their attentions were never diverted elsewhere; the number of posts they wrote reflects their entire contribution to the site. By comparison, those in the *Panoply* condition could

Usage Patterns (Full Cohort)

Activity	Group	Mean Word Count (SD)	Mean Frequency (SD)
Posts	P	183.30 (218.34)	2.72 (2.76)
	W	337.69 (547.9)	6.62 (8.76)
Responses			
Support	P	371.52 (515.90)	8.94 (11.95)
Debug	P	N/A	10.82 (13.72)
Reframe	P	301.55 (469.64)	5.99 (8.71)

Table 7.1: Average usage patterns for the *Panoply* platform and the online writing application. Note that individuals assigned to the writing task posted more frequently and composed lengthier entries. *Panoply* participants used the site more often overall, however, and spent the most time responding to other users. See text for further discussion.

divide their time between submitting posts, responding to others, and reviewing responses from the crowd. Second, those in the *Panoply* condition may have been more tentative about submitting posts, simply because they had an audience. To the extent that submitting posts was therapeutic, participants in the Panoply condition, on average, received half the dose of those in the writing condition. Future designs of *Panoply* should offer additional incentives for users to post more frequently. For example, users might be given the option to record negative thoughts privately, should they want to reframe their thoughts on their own, without any input from the crowd.

7.2 Dose-Response Relationship

Many Internet interventions observe relationships between engagement metrics and psychological outcomes. In the simplest case, those who participate most frequently experience the greatest gains. This relationship has been observed for several Internet-

based interventions, including programs targeting depression and anxiety (Christensen, Griffiths, & Korten, 2002), smoking cessation (Cobb, Graham, Bock, Papandonatos, & Abrams, 2005), and eating disorders (Troop et al., 1996).

To examine whether similar relationships might exist for *Panoply*, I conducted several linear regressions comparing various log-transformed engagement metrics (e.g., total words composed, number of posts submitted, number of responses submitted) with treatment outcomes (e.g., changes in depression symptoms and use of reappraisal). I also performed a post-hoc median split on the engagement metrics, classifying participants as high- and low-intensity users, and then comparing outcomes between these two groups with independent t-tests. None of these analyses were significant, indicating a poor linear relationship between engagement metrics and outcome.

Similar null dose-response findings can be found throughout the literature (see Donkin et al., 2013). Sometimes the relationship between usage and outcome is not linear and is best approximated with more complex models. For instance, as noted by Donkin et al, treatment gains may saturate after a certain point, creating curvilinear dose-response relationships. Individuals who reap benefits early on may elect to reduce their engagement with the platform. Baseline psychological traits may also add complexity to the dose-response relationship (Blankers, Koeter, & Schippers, 2013) Given these issues, it may not be surprising that simple, linear analyses failed to detect a clear dose-response relationship on *Panoply*. More advanced statistical models may be needed to more thoroughly examine how behaviors on the *Panoply* platform influence psychological outcomes. This is an area for future research.

7.3 Cost

In its current design, *Panoply* incurs fees from MTurk, Heroku (application hosting), Amazon Web Services (data storage), and SendGrid (an email client). Aside from the MTurk costs, the system does not require more resources than a typical data-driven web application. For the purposes of this dissertation, we therefore restrict our analyses to costs incurred from MTurk.

The crowdsourcing costs for the entire study were $326.04. The cost per post varied, depending on the user's progress within the system. A post from a first-time user was the cheapest because the system only solicited 'support' messages. Once the user completed all the modules, however, the system posted MTurk tasks for all three response types. On average, the cost per post was $1.14. The average cost per user for the entire three-week study was $3.11.

7.4 Latency

The *Panoply* platform stored timestamps every time a post or a response was submitted to the database. Response speed was examined by calculating the time elapsed between a user's post and the first response that was received. As expected, MTurk workers were generally the first to return responses. The median response time for MTurk workers was 9 minutes. By contrast, the median latency for registered users was 2 hours. The difference between these groups is striking, but to be expected. The real question is whether the speed of MTurk is worth its additional costs. To better address this question, the quality of the responses needs to be examined.

7.5 Response Quality

In this section, I examine the quality of responses among different groups of individuals. I consider whether MTurk workers outperformed study participants and I also consider whether depressed individuals fared any better or worse than nondepressed individuals. I also examine whether participants preferred one type of response to another. Throughout, I discuss ways these analyses can help inform future designs of this system.

7.5.1 Quality Metrics

Response Scores

During the onboarding process, *Panoply* users were told to rate each 'support' and 'reframe' response on a 5-point Likert scale. Unfortunately, contrary to the instructions, participants did not always rate their responses. Even after excluding responses that were never read to begin with, a large percentage of responses remained unrated. Of the responses that were scored, the distribution was negatively skewed (see Figure 7.1). Only 52% of responses were rated.
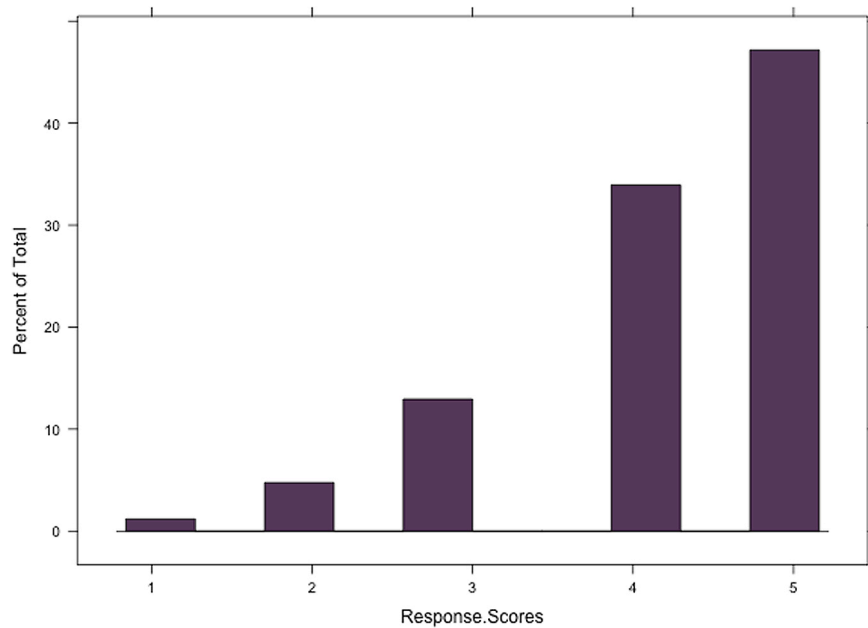
Figure 7.1: A histogram of response scores. Not shown are the remaining 48% of responses that received no ratings at all.

Users may have refrained from rating for several reasons. During the onboarding process, all participants were told that low scores would be kept private, but some users may have forgotten this or simply not believed it. They may have refrained from giving low ratings for fear they might upset someone. Other users may have defaulted to the way online content is typically rated on social sites. On popular platforms like Twitter and

Facebook, a user's content is either applauded or ignored; it is not typically given a Likert rating. Finally, some users might not have seen the value in rating responses and may not have wanted to put forth the effort. Given all these possibilities, the analyses of response scores should be interpreted with caution.

Another complication is due to the fact that respondents did not contribute at equal rates. Some individuals composed responses once or twice and then stopped contributing, while others made dozens of responses a day. This adds some complexity to the interpretation of the data at the level of the individual. For example, an individual that submits hundreds of poor responses is a much greater liability than someone who misses the mark once and then never returns, even though their average scores might be the same. Simply computing average scores per user obscures important information about rates of contribution. When evaluating an individual's performance on the system, therefore, one needs to consider both the average response score and the frequency of responses that were made.

Lacking a satisfactory way to combine these attributes in a single metric, the analyses instead focused on the group level and compared the pooled responses from one group to another. While the average capabilities of different individuals is useful to know, the most important consideration for the end-user of our system is whether, on the whole, responses from one group were better than another. The analyses of response scores use this approach.

Thank You Notes

In addition to response scores, *Panoply* users were given the option to send a thank you note whenever they received a particularly inspiring response. Before conducting the experiment, I was not sure this feature would be used at all and I assumed I would have to rely solely on the Likert response scores. And yet by the end of the study, 297 thank you notes were exchanged between users. These thank you notes provide yet another way to examine the quality of responses, both across user populations (MTurk workers vs. study participants) and across response categories (empathy vs. reappraisal).

7.5.2 MTurk Workers vs. Study Participants

In the first analysis, responses from MTurk were compared with those from the study participants. I hypothesized that, on average, responses from study participants would be rated significantly more favorably than responses from MTurk workers. I also expected more thank you notes to be sent to study participants on average.

Study participants were not paid for composing responses and were instead mainly driven to help others and themselves. I suspected these loftier, more intrinsic motivations would inspire better responses overall.

A Wilcoxon Rank-Sum test was used to compare responses between study participants and MTurk workers. As predicted, the distribution in scores between the two groups differed significantly (W=552,738, p<.01). Responses from the *Panoply* participants received higher scores (*M*=4.21, *SD* = .92) than the MTurk group (*M*=3.73, *SD* = 1.14).
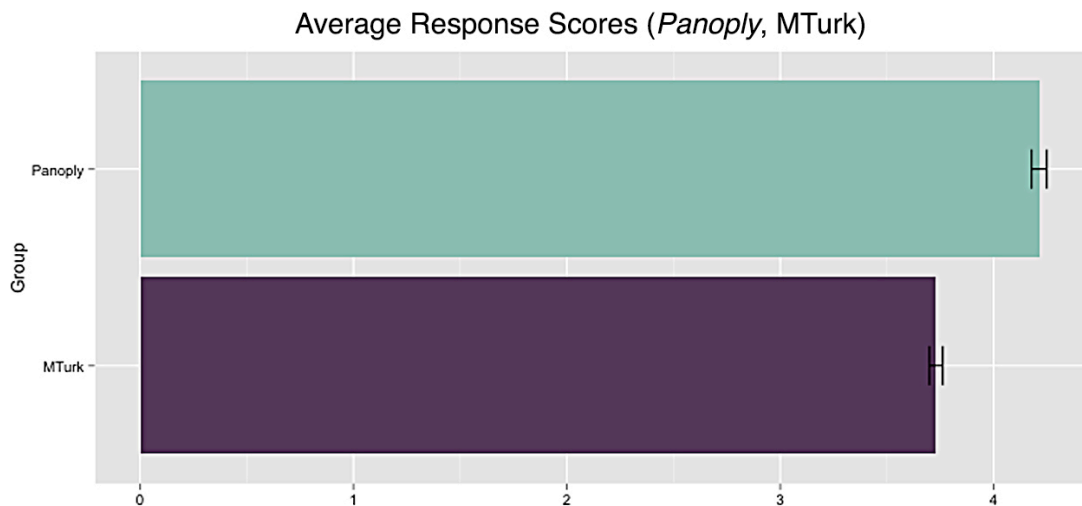
Figure 7.2: Average scores for the *Panoply* and MTurk groups. Error bars reflect bootstrapped 95% confidence intervals.
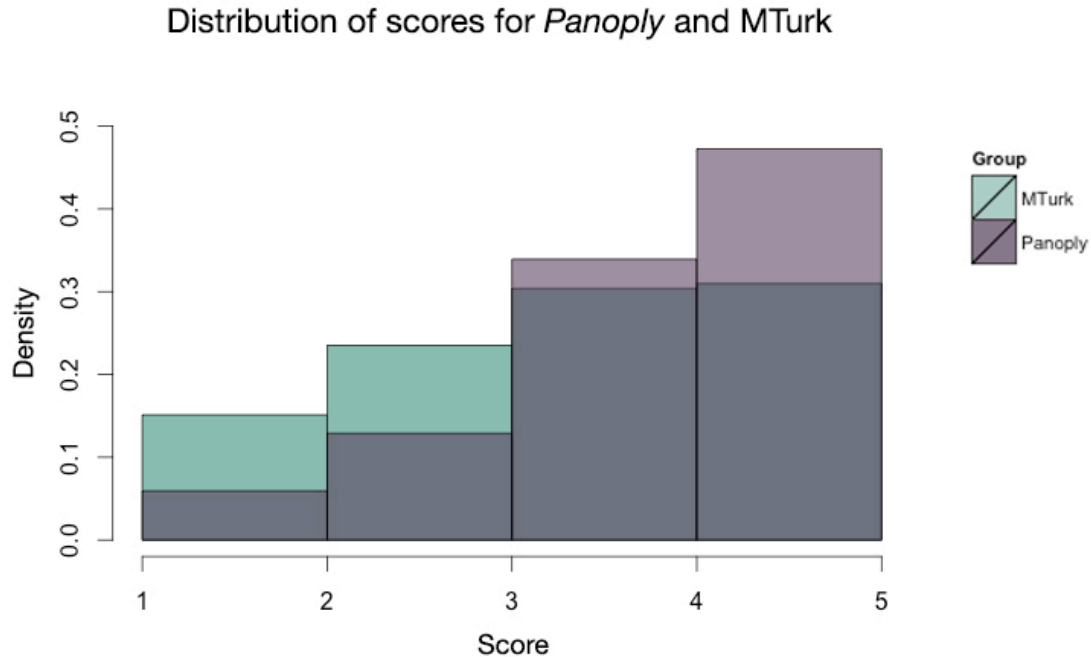
Figure 7.3: Density plot illustrating the distribution of scores from the *Panoply* and MTurk groups. *Panoply* responses garnered higher scores than MTurk responses.

A Chi-Square analysis examined the proportion of thank-you-notes that were earned by study participants compared to MTurk workers. I expected that responses composed by study participants would be more likely to get thank-you-notes. This hypothesis was not confirmed (χ^2=.06, p = .81). 10.01% of study participant responses got thank-you-notes, while 10.41% of MTurk responses got thank-you-notes.

Discussion

These results, while not conclusive, should make us question whether it is worth hiring MTurk workers to compose responses for *Panoply*. Responses from MTurk were rated significantly less favorably than responses from the study participants (who contributed for free). Perhaps the only clear advantage of MTurk is its speed. If users really appreciate quick responses, then it may be worth using MTurk, even if it the costs are higher and the responses aren't quite as good. The question is then whether users really appreciate fast responses.

To address this question, the follow-up questionnaire asked participants to indicate whether they preferred getting quick responses or high quality responses. Using a 7-point Likert scale with the endpoints "quick responses" and "high quality responses", participants were asked to indicate their preference. The average score was 5.7, indicating that participants overwhelmingly preferred high quality responses to quick responses. While it may be hard to trust results from one self-report measure, it is worth seriously reconsidering the value of speed for a system like *Panoply*.

If speed is not that important after all, it may not be worth hiring MTurk workers to compose responses. The system might be perfectly sustainable with a crowd composed of unpaid volunteers. Or, as a compromise, end users could indicate the urgency of their problem and MTurk workers could be recruited only as needed, whenever speed is of the essence. If a user has a public presentation in a few hours and needs immediate reappraisal assistance, MTurk could be leveraged to generate quick responses. For more systemic issues, such as general concerns about one's appearance, users might prefer better crafted responses, even if they take up to a day to arrive.

7.5.3 Depressed vs. Nondepressed Individuals

Depression is often characterized by impaired cognitive functioning and reduced motivation (Radloff, 1977). These deficits could make it difficult for depressed individuals to compose creative, well-written responses on the *Panoply* system. On the other hand, depressed individuals may be better positioned to relate to the struggles of others, having experienced similar adversity themselves. They may be more compassionate and sensitive and may have unique insights that aren't available to other respondents on the system. Further, while most cognitive theories of depression highlight deficits in thinking, the real story appears to be more nuanced. Under certain conditions, low mood and negative affect can confer cognitive advantages, such as reduced inferential biases (e.g., less stereotyping, fewer fundamental attribution errors), improved memory, and enhanced politeness, among others (for a review, see Forgas, 2013). Interestingly, Forgas (2002) found that negative mood led people to be more cautious and

polite in interpersonal situations – a tendency that would certainly be useful when composing responses on *Panoply*.

To examine the performance of depressed individuals, participants were split based on the accepted clinical cut-off for the CES-D; individuals scoring above 15 were categorized as depressed, while those scoring below were categorized as nondepressed. Response data from MTurk was not included because depression assessments were not collected from that population.

Response scores awarded to the depressed cohort were compared with scores awarded to the nondepressed cohort. There was no significant difference in response scores between the depressed group (*M*=4.18, *SD* =.93) and the nondepressed group (*M*=4.28, *SD*=.91), (W=45238, *p*=.14). Interestingly, depressed participants contributed significantly more text responses on average (*M*=9.49, *SD*=12.76) than nondepressed participants (*M*=5.38, *SD*=5.73), (W=650, *p* < .01).

There was no difference in the proportion of thank you notes awarded to depressed (10.69%) vs. nondepressed individuals (9.76%), (χ^2=.08, *p*=.78).
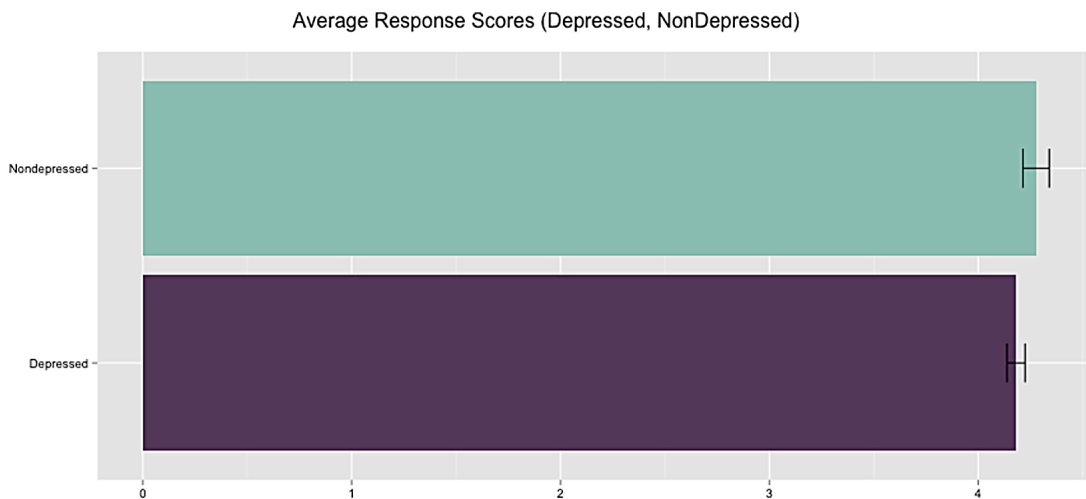
Figure 7.4: Average scores awarded to depressed and nondepressed *Panoply* participants. There was no significant difference in the quality of responses composed by depressed vs. nondepressed individuals.
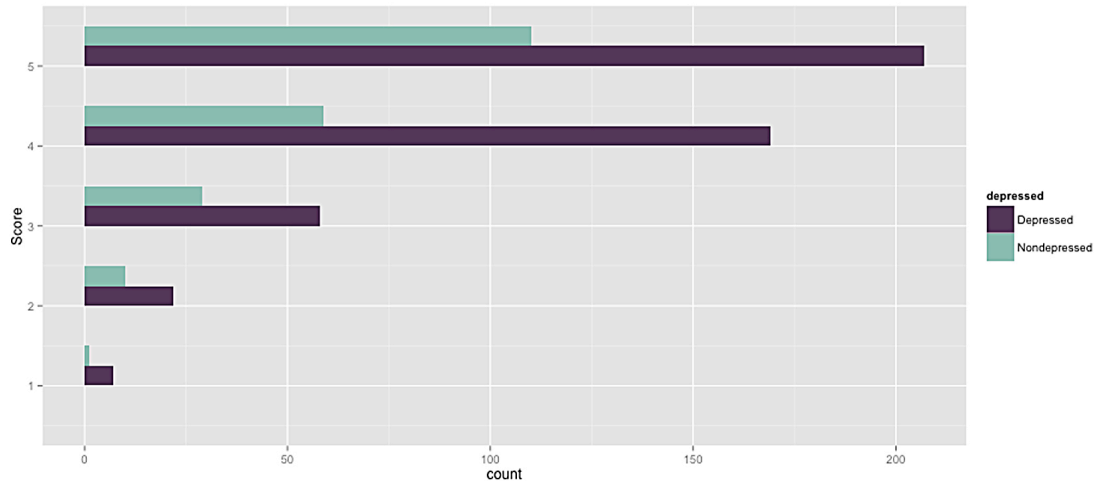
Figure 7.5: Histogram of response scores awarded to messages composed by depressed and nondepressed participants. The shape of the distributions is similar, but depressed individuals contributed significantly more responses overall.

Discussion

One of the risks of developing a system like *Panoply* is the possibility that those needing help will be unable or unmotivated to help others. Another concern is that individuals with depression might inadvertently spread negative attitudes and mindsets to others in the system. However, based on the analyses from this dataset, this does not appear to be the case. Response scores from the depressed cohort were only slightly lower, and not significantly so. Further, depressed individuals were significantly more motivated, contributing well over twice as many responses as their nondepressed counterparts (see Figure 7.5).

7.5.4 Empathy vs. Reappraisal

In the final analysis, ratings for different response types were compared. Since users could not rate 'debug' responses in the interface, I only compared empathy responses to reappraisal responses. A comparison between empathy and reappraisal responses revealed no significant difference in ratings received (W =477,269, *p* =.88), or in the proportion of thank you notes that were sent (χ^2 =2.47, *p* =.11).

These results match the self-report data obtained in the follow-up questionnaires. In the follow-up questionnaires, users had to indicate which response they enjoyed receiving most: support, reframe, or debug. 31% of users preferred support responses, 30% preferred reappraisals, and 13% preferred debug responses.

Discussion

These data suggest that, on the whole, users found empathy and reappraisal responses similarly appealing. Of course, these data reflect the average preference across all users. More refined analyses could help determine whether certain responses are better suited for certain individuals. Future work is needed to determine whether factors such as gender, age, or depression symptoms might influence one's preference for different response types. Updated versions of the *Panoply* system might apply different ratios of responses for different types of users. Different types of posts might also necessitate different combinations of responses. Configuring different response algorithms and testing them with actual users is important future work.

Chapter 8

Vignettes

This chapter examines some of the more noteworthy interactions that took place among *Panoply* users during the experimental trial. These interactions serve to highlight the benefits of the platform as well as its various challenges.

8.1 Positive Behaviors

8.1.1 Self-Generated Reappraisals

Before accessing crowdsourced reappraisals, users are asked to compose a reframe for themselves (see section 4.3.4). This exercise is optional, however, and users can easily side-step it by typing whatever they like into the textbox. At the outset, it was not clear how users might react to this feature. It was added late in the development cycle and was one of the only system components not subjected to extensive user testing. It was quite possible that this exercise would be considered irksome and not taken seriously.

Yet, I could not identify a single self-generated reframe that appeared off-topic or nonsensical. Of the 212 that were submitted, almost all were extremely insightful and well written. Here is a particularly poignant example:

> **Post:** My mom's cancer may be back. My mom had throat cancer a few years ago, and although we were told that it's unlikely to return, they have detected something on the latest PET scan. Waiting for the results is very stressful. I don't

know what insurance they are with now and they are far, so I am not sure how I'll be able to help, if it comes to the worst.

Self Reframe: As with the first time, this could be an important reminder that family comes first and a way to experience love in a deeper way. It may bring our family close and remind us how dear we are to each other.

Another example, addressing a comparatively milder situation, is also worth citing:

Post: My boss is coming back today after three weeks of vacation. I'm so anxious for our meeting! I don't know why I feel this way, because he's not an angry person, but I just am always afraid he'll want things to be a certain way or demand that something happen that I will have to pull off somehow and it'll make me work extra (unpaid overtime) or just make my life harder. So I feel scared!

Self Reframe: Well, look, he's not actually a bad guy, he just has high standards, but he's reasonable and you've never actually tried telling him you don't have time to do something. he'd probably take it better than you think. he probably has no clue how much overtime you work! i bet it will go better than you think because you are PREPARED you little whippersnapper :) Go get 'em!

Both of these examples exemplify excellent reappraisal responses. Both users offer realistic and creative ways to reframe their respective situations.

These examples also illustrate ways in which multiple reappraisals can be incorporated into a single response. The latter example, in particular, includes at least five distinct reappraisal ideas. This author enthusiastically lists one reframe after another, almost as though she is throwing punches in a boxing ring.

It would be interesting to know whether self-generated reappraisals, such as the ones cited above, are more or less powerful than reframes that are generated by the crowd. An argument can be made for both sides. On the one hand, crowd workers are distanced from the situation and may be better poised to think creatively and realistically. On the other hand, crowd workers are not privy to all the nuances of the situation; they may be more likely to misinterpret important details or provide unrealistic suggestions.

The amount of cognitive effort required to create a self-generated reappraisal is also worth considering. It may be more difficult to compose a reframe for yourself than to

integrate reappraisal suggestions from the crowd. Future research should be conducted to examine these possibilities.

Another self-generated reframe worth citing is as follows:

> **Post:** A friend of mine posted a survey to FB and then commented on it, using it as a talking point for her pro-life beliefs, but the survey's irrelevant to her point. It makes me angry. I feel like she's being ignorant and illogical. I've always tried to respect her religion and beliefs, but as a student of science she should know better than to do what she's doing. You should know how to think critically after four years at MIT. There are plenty of valid ways to debate this point, and this is not one.
>
> **User Reframe:** Everyone has their own biases and blind spots, and I'm sure I do as well. It's actually good to understand how other people think about things. This could help me examine the ways I think about things myself and perhaps identify my own biases and blind spots.

This is an interesting example, especially in light of recent research examining the power of reappraisal to successfully mitigate interpersonal (Finkel, Slotter, Luchies, Walton, & Gross, 2013) and even political (Halperin & Gross, 2011) disagreements between people. The user in this example does not change her stance dramatically, but the reframe siphons out the venom considerably. It is helpful to contrast this exchange with those on other online platforms. Online commenting systems can be extremely uncivil and can lead to greater polarization of beliefs (Anderson, Brossard, Scheufele, Xenos, & Ladwig, 2014). A system that encourages people to carefully reframe the way they view their adversaries could be beneficial.

8.1.2 Thank You Notes

Hundreds of thank you notes were submitted on the site. Users were not given any specific hints about what to say, but many offered reasons for why they particularly liked a given response. For example, consider the exchange below:

> **Original Post:** i hate most of my classes and i don't want to do work for them. i don't understand why i'm here or doing this. it seems really pointless and i feel like everyone else is way happier with what they're doing. i'm afraid i'm going to

do badly this semester because i don't like doing work and all i want to do is sleep.

Response: Dear ----, we cannot be sure how other people feel; some show their feelings more than others, but still.. I am sure that others have their moments of frustration with their chosen field and classes they are in.

Thank You Note: hi ----, thanks for this! i often think other people are coping better than i am, and it's always great to be reminded that they may just be hiding it better.

This thank you note illustrates a pattern that came up again and again on the site. Users would often latch on to a particular reframe and then restate it using their own words, almost as though they were rehearsing it. It is possible that doing so helps the person encode the reappraisal more deeply, helping them adjust farther away from their original interpretation of the situation. A few additional examples of this are presented below:

Thank You Note (1): very encouraging thanks, funny how I forget what I - and my family - have already accomplished just by moving here, adapting to the new context, etc. I may just not be patient enough.

Thank You Note (2): thanks for your thoughtful response! It's sometimes tough for me to consider why someone is doing something in the midst of a conflict, and your response helped me consider my Dad's side. I appreciate it.

Thank You Note (3): Thanks ----, for re-framing this situation as an opportunity to learn and practice empathy and patience. :) I like that concept, turning my problems into opportunities for growth.

8.1.3 Improvements

In the follow-up questionnaires, several individuals reported finding the reframe task somewhat challenging at first. It is likely that many users grew better at this task over time, as they acquired more experience on the site. Consider, for example, the following post that was submitted to the system:

Post: Things that stress out other people on Panoply stress me out too so I can't reply to them to rephrase it. I don't know how to respond to some people here on panoply because I have the same stresses. How can I help other people when I am struggling helping myself deal with stuff?

At first glance, this post seems to point to a fundamental problem of peer support systems like *Panoply*. That is, individuals experiencing stress may not be prepared to manage the stress of others. Interestingly, after writing this post, rather than giving up on the site, this user went on to compose 89 additional responses, 17 of which were reframes. Moreover, by all accounts, this person was actually quite good at composing reframes. In fact, she was among the top 10 highest scoring users on the site. By the end, this person gave the site high marks on her follow-up questionnaires, suggesting she felt more comfortable with the platform over time.

While this user may have overcome her initial feelings of self-doubt on her own, additional assistance may be needed for other users. Future versions of the system should do more to reassure users and help them understand that reappraisals can be hard to generate at first and that the process often gets easier with practice. This message could be made explicit in the actual copy presented in the application. Better yet, novice users could be given more structured help when first composing reappraisals. With enough data, the interface might be able to predict the kinds of responses that are likely to help for different kinds of situations. Based on this data, novice users could be given additional hints and prompts to help get them started.

8.2 Troubling Behaviors

Throughout the study, no malicious behaviors were observed from site users or MTurk workers. Even when some of the site content was challenging (and indeed, at least one participant noted that she found some of the posts 'hard to agree with'), respondents remained remarkably sensitive and understanding, as per their task instructions. This is perhaps not surprising, given the user population. The online workers were financially incentivized to follow the task instructions, or else they risked losing wages or, worse

132

still, getting blocked from MTurk. Study participants were not risking financial outcomes, but many probably assumed some oversight by the experimenter.

I did, however, observe one participant who used the site inappropriately, though not maliciously. Should this site be made publicly available, I suspect others like him would participate and it is therefore important to describe his case in some detail. This individual (hereafter referred to as BG, for anonymity) was a 20-year old male with a CES-D score of 35 at intake, suggesting severe levels of baseline depression.

BG's second post was flagged by online crowd workers, suggesting possible harm to himself or others. Once the post was flagged, an email was automatically sent to BG with links to mental health resources. He was also reminded that *Panoply* is a self-help tool and is not intended to be a substitute for formal mental health care. It is worth noting that BG's troubling post was made at 1:47am and follow-up resources were mailed within the hour. This speaks to the benefits of systems like Mechanical Turk, where online workers are available anytime, day or night. Had our system been composed of a select group of specialized workers, such as oDesk workers[7], the post would probably not have been caught until many hours later.

As per the site protocol, BG's troubling post was immediately hidden from other *Panoply* users. All subsequent posts made by this individual were sent only to MTurk. Also, while he was still able to compose responses for other *Panoply* users, they were never delivered. He was never made aware of these site restrictions, however. These safety precautions were made in consultation with the MIT IRB.

After this incident, BG continued to write challenging posts on the site. Although it is impossible to make any diagnostic assessments from a handful of anonymous text entries, many of BG's posts were suggestive of persecutory delusions (e.g., people following him or hiding near his house or otherwise attempting to smear his reputation). While we have no right to question the veracity of these troubling statements, especially with so little information to draw upon otherwise, it is worth considering how *Panoply* might respond to truly delusional thinking. One possible solution to this problem is to have the users

---

[7] www.odesk.com. For an example of a crowd powered system that uses oDesk, see (Kokkalis et al., 2013)

help moderate the posts more aggressively. Posts that seem delusional or fall outside of the accepted norms of the platform could be blocked.

Interestingly, all the MTurk respondents handled BG's posts with considerable aplomb. They responded with caring statements of empathy and they did not pass judgment (though one worker wrote to us directly in the 'comments' section of the task and noted that this was "a strange HIT"). On the one hand, these responses could be very therapeutic for BG. In real-life, his persecutory statements may not be met with such calm consideration, making it hard for him to relate to others socially. However, some of the reappraisals he received suggest that *Panoply* could potentially inflame or exacerbate paranoid delusions, insofar as they do not debunk them outright. In many cases, workers played along with the story and circumstances of BG's posts and accepted most everything he wrote at face value.

Online Q&A sites are frequently beset by troubling behavior, either because the user is trolling and trying to provoke controversy or, sadly, because the user is troubled himself. In some cases, these individuals are ignored or bullied, perhaps making their troubles even worse. *Panoply*, by contrast, has mechanisms to review each post that is entered into the system. Users who seem troubled are quickly sent resources directing them to formal mental health hotlines and services. Responses composed on the platform are also reviewed, significantly reducing the likelihood that anyone would be bullied. However, until more edge cases like BG are spotted, it is hard to know how well these safety mechanisms will work in practice. This is an important area of future work.

Chapter 9

Conclusion

In the past decade, there has been an explosion of technologies designed to augment human cognition. Google's Maps application helps us navigate new environments. Google Translate helps us understand foreign languages. Smartphones of all kinds augment our memories by storing information about upcoming appointments, contact info, and the like. But, in addition to managing the information around us, to be truly smart, to be emotionally intelligent, we also need to manage our internal processes, our inner thoughts and emotions. *Panoply* offers a safe, anonymous platform to rehearse these skills alongside a group of supportive crowd helpers.

The experiments described in this dissertation suggest that such a platform, when used repeatedly, confers psychological benefits to both healthy and depressed individuals. In particular, it appears to inspire greater use of cognitive reappraisal among healthy individuals and, for depressed individuals, it reduces depression symptoms and perseverative thinking. Reappraisal may mediate the effects of this intervention and *Panoply* may be most useful for those who typically underutilize reframing techniques in their daily lives.

I also find that *Panoply* offers multiple pathways for engagement. A variety of social feedback mechanisms prompt users to return to the site regularly. Various measures of engagement show significantly greater usage of *Panoply* than a matched control task.

9.1 Limitations and Future Directions

9.1.1 Further Experimentation

The findings from the randomized controlled trial were extremely encouraging, but by no means definitive. For one, roughly three quarters of the participants enrolled in the study were female. While no gender effects were observed in any of the analyses, it is unclear whether the effects observed in the study would generalize to a predominantly male population. Also, a larger sample size in general might have been better able to detect more significant differences between the control and treatment conditions, particularly with respect to measures of positive affect and subjective happiness.

Additional RCTs should also be conducted to further examine questions of efficacy and engagement. It would be interesting, for example, to compare *Panoply* with existing computer-based CBT interventions, instead of expressive writing. *Panoply* offers a considerably shorter course of treatment than typical online CBT interventions. However, *Panoply* might be comparable to longer, top-tier interventions when examined as an 'open access' intervention. If *Panoply* attracts more use, it might be more beneficial overall, even if it is a comparatively shorter intervention. Additional experiments are needed to examine this possibility.

9.1.2 Cognitive Strategies

In its current instantiation, *Panoply* offers three modules: support, debug, and reframe. In the future, additional modules could be added as the user progresses through the platform. For example, to expand on the cognitive techniques already introduced, the platform could challenge users to identify and dispute some of their deeper core beliefs (e.g., the need to be perfect at everything). Core beliefs are sometimes harder to identify than cognitive distortions, because they may not manifest themselves directly in negative thoughts and they may lie at deeper, less consciously accessible levels of cognitions (Beck, 1979). But carefully constructed crowd interactions could help individuals ferret out and confront these core beliefs.

9.1.3 Behavioral Strategies

Panoply could also be extended to address some of the behavioral components of CBT. In its current state, *Panoply*'s primary emphasis is on cognitive techniques like reappraisal and cognitive restructuring. However, a full course of traditional CBT also incorporates a variety of behavioral methods, such as exposure therapy, deep breathing and behavioral activation. Additional modules to the *Panoply* platform could involve crowdsourced suggestions for activities that are likely to boost positive emotion. Activities from the field of positive psychology could be especially relevant and useful for crowd-powered systems like *Panoply*. To illustrate, I will briefly examine how crowds might be leveraged to help individuals set goals and utilize signature strengths.

Goal Setting and Planning

Goal setting and planning interventions can significantly enhance well-being. Crowd-powered technologies like *Panoply* could intervene in at least two cognitive components believed to be important for healthy goal pursuit – "pathways thinking" and "agency thinking."

Pathways thinking, as described by Snyder (2002), is the process of identifying the specific routes and steps needed to attain a goal. Crowds could be recruited to help users define clear blueprints for goal attainment. They could also be recruited to intervene whenever an individual reaches an impasse and needs to brainstorm alternate pathways to achieve a goal. Research by Zhang et al. (2012) shows how crowds can be recruited to help plan vacation itineraries, making sure each activity is a reasonable precursor to the next (both in terms of the time required to complete the task and the distance needed to travel to the next stop). Similar design approaches could be used to help individuals identify specific, manageable goals and the various routes for achieving those goals. In practice, such an approach would be most appropriate for goals that do not require considerable domain expertise (e.g., finding ways to read more vs. perfecting a machine-learning algorithm). However, crowdsourcing platforms that incorporate social network information can match users with respondents who have similar interests and skills.

Systems such as these might be useful for individuals that need help planning goals within specialized domains.

Agency thinking is defined as the perceived ability to achieve one's goals (Synder, 2002). This is another component thought to be important in successful goal pursuit. To bolster this faculty, crowds could be hired to congratulate users when goals are achieved or boost their confidence when goals seem unduly challenging. Simple messages of encouragement, crowdsourced at opportune times, might help individuals feel more confident in their ability to attain their goals.

Utilizing Personal Strengths

Utilizing personal strengths is associated with a host of positive outcomes, such as enhanced well-being, less stress, and greater work satisfaction (Peterson, Stephens, Park, Lee, & Seligman, 2009). Crowd-powered technologies could be used to augment strengths interventions in various ways. For instance, crowds could be trained to help people identify optimal ways to utilize signature strengths. As crowds help people identify goals and planning strategies, as discussed earlier, they could tailor their recommendations to align with the user's signature strengths. Indeed, evidence suggests that strengths use mediates the relationship between goal pursuit and well-being (Linley, Nielsen, Wood, Gillett, & Biswas-Diener, 2010).

Crowds could also help a user brainstorm new ways to use certain types of signature strengths throughout daily life, outside the context of goal interventions. For those whose signature strengths include 'Love of Learning,' for example, crowds could help users find ways to pursue new pathways to knowledge (such as providing links to interesting online courses, blogs, or adult education classes). Previous work in human-computer interaction has outlined ways to successfully crowdsource idea generation in domains as diverse as product ideation (Poetz & Schreier, 2012), poetry translation (Kittur, 2010), and furniture design (Yu & Nickerson, 2011). Crowds could similarly produce useful ideas for strengths use that may not have occurred to the user. While automatic recommendation systems for movies and music are now commonplace and may eventually suggest interesting and apt ways to use strengths, these algorithms require massive amounts of user data to work effectively (Resnick & Varian, 1997). To predict what movies you

might like, for instance, recommendation algorithms need to know a lot about your own movie tastes and how they compare with similar others. Until individuals are commonly reporting ways in which their strengths align with the activities they pursue, subtle and personalized recommendations for strengths use will likely remain the purview of human intelligence.

Crowds could also be tapped to provide encouragement and positive feedback as individuals pursue strengths use interventions. Buckingham (2010) suggests that only 17% of people report using their strengths 'most of the time' each day. Offering individuals motivation, as well as more interesting and varied opportunities to exercise strengths, might help encourage greater use of this technique.

9.1.4 Different Crowd Populations

In its current form, *Panoply* primarily draws from two populations: MTurk workers and registered site users. The site users can be individuals seeking help with negative thoughts or they can be unpaid volunteers who simply want to log on and help others. For now, everyone in the system is anonymous, including the respondents. While many of our study participants appreciated the anonymity of the system, there may be value in revealing the identity of respondents, especially if they are friends or family members. It would be interesting to know whether support from known peers is perceived as more or less valuable than support from a cadre of anonymous crowd workers. Future research could also examine whether participants prefer responses from crowd workers, trained therapists, or autonomous agents. An interesting future experiment might involve randomly labeling MTurk responses as either from the crowd, a therapist, or an artificially intelligent agent. It is quite possible that the supposed source of the response has important implications for how it is received.

MTurk has its shortcomings and future work could examine whether other crowdsourcing labor markets are preferable. Companies like oDesk facilitate closer relationships between the requester and the crowd worker; on oDesk, workers can be trained at length and supervised closely. If speed is not an imperative for systems like *Panoply*, crowd workers from platforms like oDesk could be ideal. Their work can be

scrutinized over time and groomed for quality. And while further analyses are needed, it is likely that oDesk workers would be cheaper and easier to hire than trained therapists.

Finally, future experiments should examine whether *Panoply* could be sustained solely from the contributions of unpaid site users. The analyses suggest that these users contribute better responses than MTurk workers overall. And, while the median response time from this group was on the order of several hours, this degree of latency may be tolerable for many users. Many of the posts submitted during the RCT involved systemic issues that did not appear to require immediate attention. Further, the majority of users on the system reported preferring quality over speed.

9.1.5 Artificial Intelligence

Various techniques from the field of artificial intelligence could be used to enhance the *Panoply* platform. Advances in natural language processing, in particular, could help groom responses from the crowd, ensuring that malicious or inappropriate responses are not returned to the user. Similar approaches have already been applied to automatically detect cyberbullying on online platforms (Dinakar, Jones, Havasi, Lieberman, & Picard, 2012). Algorithms can also be trained to automatically detect text features that correlate with good responses (e.g., sentiment, grammar), obviating the need for crowdsourced voting. Reappraisals that are excessively bright-sided could also be filtered out automatically with a properly trained text classifier. This could improve the content of the site overall and eliminate Pollyannaisms that might annoy and offend users (e.g., 'you will certainly get better', 'I just know things will work out').

9.2 *Panoply* as a Research Platform

Beyond its immediate practical use for individuals needing emotion regulatory support, *Panoply* can also be used as a research platform. The corpus of data collected from the RCT can be used for many research applications beyond those discussed in this dissertation. For instance, the debug data can be used to build machine-learning algorithms that automatically detect cognitive distortions within natural language.

Research in this area is currently ongoing. Such a tool could be used in many health informatics applications and could help classify millions of lines of text without the need for human raters.

The platform itself could also be used to examine important questions within the field of emotion regulation. To date, there has been little work on the interpersonal aspects of emotion regulation. Which strategies do peers typically use to support their friends? Which of these strategies are most successful? Could social support be improved if peers are trained to use adaptive strategies, such as cognitive reappraisal? Studies conducted on *Panoply* could help answer these questions.

Panoply could also be used to examine various aspects of intrapersonal emotion regulation. Cognitive reappraisal comes in many forms (e.g., seeing the silver lining, making a counterfactual comparison), and some strategies may be more powerful than others. Additional research could use *Panoply* to examine which reappraisal strategies work best for which situations and which individuals. Individual differences with respect to depression status or various personality types may necessitate different forms of emotion regulatory assistance. Eventually, as our understanding of strategies like reappraisal become more nuanced, platforms like *Panoply* could become more flexible and more powerful. Indeed, in recent years, psychological flexibility has been touted as an important skill for maintaining mental health and well-being (Kashdan & Rottenberg, 2010). In the real world, no singular strategy is best suited to meet every challenge, be it major life trauma or quotidian abrasion; rather, stress is perhaps best managed by applying different strategies to different situations. Future versions of *Panoply* could be similarly flexible and could cater its support to better meet the needs of particular situations and the various personalities of its users.

9.3 Insights for Crowdsourcing Practitioners

In this section, I consider various design lessons that can be drawn from the *Panoply* platform. I focus the discussion primarily for those designing crowd-powered systems, but many of the insights will likely be useful for anyone working in crowdsourcing.

9.3.1 Manage Quality with SAT Verbal Tasks

As discussed at length in section 4.3.6, maintaining quality can be a particular challenge when using labor markets like MTurk. *Panoply* employed many quality assurance techniques discussed previously in the crowdsourcing literature, such as creating qualification tests (Kittur, Chi, & Suh, 2008), designing short, circumscribed tasks with clear instructions (Morris & McDuff, 2014), appealing to intrinsic motivations (Rogstadius et al., 2011), and screening for IPs that do not originate from English speaking countries (Chandler, 2013).

However, even after implementing the 'gold standard' approaches described above, improvements were still sorely needed. Perhaps the most valuable technique was the inclusion of a short, SAT verbal qualification test (see Figure 9.1). MTurk workers could not access any of the *Panoply* tasks until they first successfully completed two fill-in-the-blank SAT questions. A dozen questions were prepared and two were randomly presented for workers on each page load. The quality of the responses from MTurk improved substantially after this procedure was put into place. This technique is highly advisable for any crowdsourcing tasks that involve writing. I found the SAT verbal task to be the quickest and easiest way to find workers with a nuanced command of the English language.

9.3.2 Reassess the Need for Speed

Many crowd-computing systems and crowdsourcing platforms boast extremely fast response times (e.g., Bernstein et al., 2011; Mamykina et al., 2011). Intuitively, it seems as though speed would almost always be of paramount importance. When requesting work from a crowdsourcing system, one typically wants responses as quickly as possible. At first glance, it might seem that speed would be especially important for emotional support systems like *Panoply*. Individuals feeling distressed might want help as soon as they can get it. However, when asked, the *Panoply* participants overwhelmingly preferred high quality responses to quick responses (see section 7.5.2). In part, this might reflect the fact that many of the posts submitted to *Panoply* did not involve urgent, time-sensitive stressors. Rather, many of the posts involved longstanding concerns (i.e.,

This HIT requires strong English skills. To work on this HIT, please complete the following two sentences.

Note: if you pass this exam, you won't need to take it again for any of our other HITs.

Because theaters refused to show it when it was first released, Citizen Kane was a ____ failure, though now it is considered one of the greatest American films ever made.

○ perennial ○ commercial ○ revolutionary ○ aesthetic ○ personal

Lavender has a ____ effect; its aroma alleviates tension and anxiety.

○ iridescent ○ painful ○ joyous ○ fragrant ○ soothing

Done

Figure 9.1: A screenshot of the SAT verbal qualification task that was deployed on MTurk.

chronic family issues, body image problems, generalized anxiety, etc). For issues such as these, carefully crafted responses may be far more valuable than ones delivered as quickly as possible. Also, data from empirical studies on emotion regulation suggests that techniques like cognitive reappraisal may be less effective immediately after intense emotional experiences (Sheppes & Gross, 2011). Reappraisal takes cognitive effort and, even with crowdsourced support, it may be hard to accept and integrate new reinterpretations during or immediately after an intense emotional experience. It is possible that these types of responses may actually be more effective if they are delayed somewhat.

Considerable engineering effort went into ensuring that *Panoply* responses were delivered promptly. However, after having built and tested the system, it is not clear that such efforts were worthwhile. For platforms like *Panoply*, more effort should be spent

ensuring that the responses are of high quality. Crowdsourcing designers should carefully assess the trade-offs between speed and quality early on in the design process.

9.3.3 Utilize Hybrid Crowd Designs

Panoply recruited two crowd populations: MTurk workers and unpaid *Panoply* users. The virtues of combining these populations are discussed in detail in section 4.3.6, but are also summarized here briefly.

The merits of using MTurk were myriad. MTurk was a highly useful prototyping population. Workers from the site were also available and always ready to test new user interface designs or interaction patterns. They also acted as a stand-in for a large, active user community, making the platform seem more vibrant than it would have been had it been limited solely to other *Panoply* users. This helped enhance the ecological validity of the RCT and provided a compelling glimpse of what such a system could be like with thousands of active users. Finally, because they were being paid, MTurk workers willingly performed maintenance tasks such as carefully moderating posts and responses. Asking *Panoply* users to do this may have degraded their user experience.

Panoply users, on the other hand, were valuable because they were intrinsically motivated to compose excellent responses. Doing so was also part of their therapeutic program and many may have been genuinely excited to help their peers on the platform. Overall, their contributions were stronger than those from MTurk.

Combining these crowd populations enhanced the platform as a whole and made it easier to answer many crucial research questions. Other crowdsourcing systems could benefit from following this model.

9.3.4 Offer Reciprocal Benefits for Workers and Requesters

Crowdsourcing practices that utilize labor markets like MTurk are sometimes seen as exploitative. Indeed, Jonathan Zittrain – a fellow at the Harvard Berkman Center – has widely derided MTurk as a 'digital sweatshop' (Zittrain, 2009). Sadly, there is some element of truth to this claim. Many of the tasks posted on platforms like MTurk offer

low wages and involve nothing more than repetitive and menial information processing tasks (e.g., labeling images or transcribing text). Recently scholars from human-computer interaction have begun to propose new ways to benefit these crowdworkers, so that the practice as a whole can be more sustainable (Kittur et al., 2013). As argued by Dontcheva and colleagues (this author included), one approach is to design tasks that offer new pathways for learning and new opportunities that extend beyond the crowdsourcing environment (Dontcheva et al., 2014). *Panoply* exemplifies this approach; the workers on the platform are not only getting paid, they are also learning valuable new skills to help them manage stress in their daily lives. For example, after working on a *Panoply* task, an MTurk worker sent the following message:

> Hello! I'm just contacting your group to say that this HIT was very interesting. Combined with the ease of completion, **I was really pleased that I was able to identify some own bugs in my thinking with this method.** So for that, thank you!

Providing MTurk workers with new skills and opportunities for learning is something that can benefit requesters as well. Workers that are driven by non-monetary incentives (such as the opportunity to learn) may be more engaged and may produce better work overall.

9.4   Insights for Designers of Online Psychotherapies

Throughout this dissertation, I've documented some the shortcomings that plague many existing computer-based psychotherapies (see section 2.3 in particular). As noted, many existing applications suffer from a poor user experience. This, in turn, leads to problems with engagement and adherence. To address these issues, I argue that designers consider adopting the following design principles (see also section 4.1 for a longer discussion):

- Consider how people use technology today and build interventions that fit these usage patterns. For example, applications should be 'snackable' and should not require long, extended periods of sustained attention. Ideally, the content should

be packaged so that it can be consumed in short bursts (such as while waiting in line for coffee).

- Offer pathways for autonomy and competence. New interventions should provide choice and continuous feedback for the user. The interactive capabilities of new technologies offer interesting new ways to make users feel competent and in control.

- Incorporate persuasive design strategies to support engagement and habit formation. As stated in section 4.1, it can be very useful to borrow persuasive strategies that work well in other contexts. *Panoply*, for instance, uses many of the same design features that are found on popular Q&A sites.

- Consider using organic social interactions to help foster engagement. Social exchanges offer a nice, intuitive hook to keep users coming back to an application. Other researchers have suggested using email prompts or quick telephone calls to remind and encourage users to continue using online interventions (e.g., Mohr et al., 2013). Another, perhaps more organic approach, is to package reminders within naturally occurring social interactions. In *Panoply*, for example, users are sent notifications anytime they receive new messages or thank you notes. These messages are not framed as reminders, per se, but are instead natural ways to lead users back onto the site (see sections 4.1.4 and 4.3.4 for more details).

The above design patterns worked extremely well for *Panoply*. However, this is not to say that these approaches are appropriate for every computer-based psychotherapy platform. As always, designers should use good judgment when incorporating techniques such as these.

9.5 Concluding Remarks

Much of crowdsourcing research is framed as a sort of stopgap measure, a technique to be used temporarily until machine intelligence becomes sophisticated enough to replace it. However, there are many applications of crowdsourcing that may never be supplanted by machines. While studies have shown that individuals can form attachments to machines (Kidd, 2008; Turkle, Taggart, Kidd, & Dasté, 2006), it is quite possible that machines will never be able to comfort people quite as well as other humans. To truly empathize with someone, one needs a theory of mind and some type of similar lived experience. And, some may argue, the ability to feel real pain. Even if machines are able to generate truly wonderful reappraisals and comforting support statements, their efforts may ultimately be rejected as inauthentic simply because the machines haven't 'been there' themselves. More research is needed to test this supposition. For now, however, there is great comfort in the knowledge that crowds of anonymous strangers can be brought together to help one another in ways that were never before possible. This dissertation described new technology that makes this possible.

To my knowledge, *Panoply* is the first crowd-powered system that has demonstrated improvements in mental health in a randomized controlled trial. Hopefully, it is the first of many. Additional research and technological innovation could improve this type of platform considerably. Ideally, this work should inspire researchers to build computer-based therapies that place a similar emphasis on engagement and interactivity. While we will always need therapists and trained mental health professionals, we also need tools that can expand their reach. Highly engaging and accessible behavior intervention technologies are sorely needed if we want to resolve the mental health crisis in this country.

Appendix A

Panoply Text

A.1 Onboarding Text

This section includes the complete transcript of the *Panoply* onboarding text. The text is presented in a 'wizard' interface and is segmented into small, readable chunks (see Figure A-1).
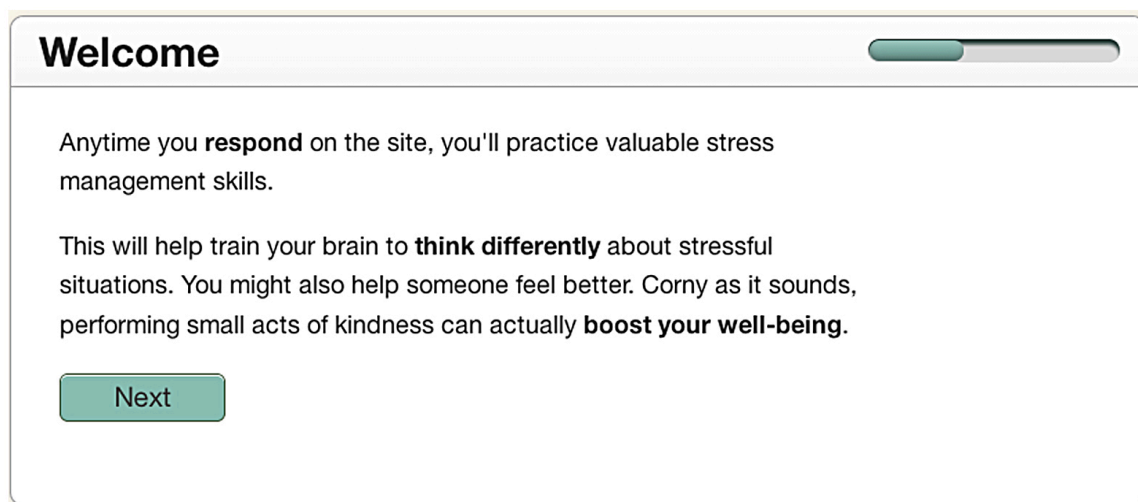
Figure A.1: A screenshot of the site orientation wizard.

A.1.1 Welcome Text

Welcome!

Thanks for joining us in this bold new experiment!

Panoply is a new **crowdsourcing** site, made up of people just like you. It's an **anonymous** place to vent your problems, manage stress, and help others. Let's take a tour!

A.1.2 Thought Record Tutorial
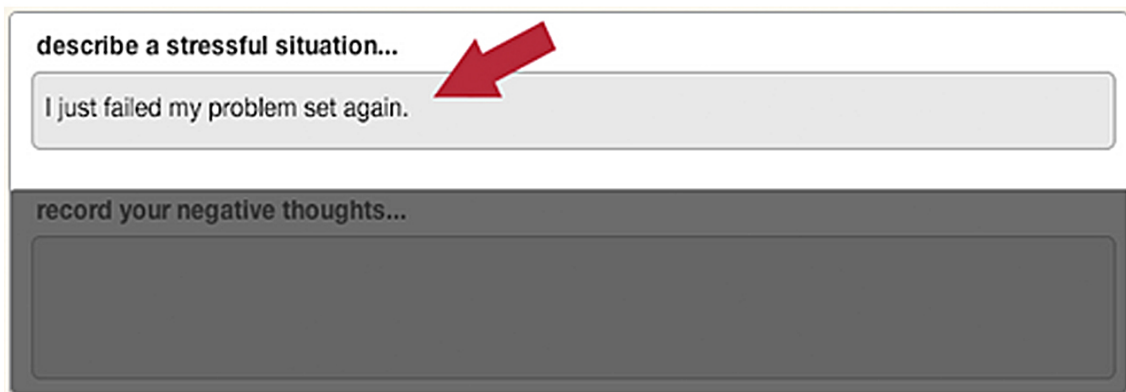
Post Something!

To start, we'll have you describe something that's **stressing you out**. What's making you anxious? What's bumming you out? It could be anything, even something really small (ideally from the past 24 hrs).

Your text will be sent **anonymously** to other Panoply members, each of whom will send you something back in return.

Panoply uses something called a **thought record** to help you get to the root of what's stressing you out. You'll want to use it frequently, as stressors crop up throughout the week. Let's check out a quick example…

First, just describe the **situation.**

Next, describe any **negative thoughts** you might be having about the situation.

Negative thoughts are what go through your mind when you think about the situation.

Ask yourself: Why does this situation stress me out? What's the worst thing about this? How does it make me feel about myself? The world? Other people? Dig deep! The more honest and raw you are, the better!

It may take some practice to capture negative thoughts at first. But it's worth it.

Learning to record your negative thoughts and separate them from the situation is an extremely powerful **stress reduction tool**. Over time, and with help from your fellow Panoply members, you'll learn that your thoughts play a huge role in how you manage stressful situations.

You'll also learn that your thoughts are not facts. They are something you can change. And by changing the way you think about a situation, you can dramatically change how you feel.

Ready to try it?
[At this point, users are shown the post interface described in section 4.3.2]

A.1.2 Site Tour

What is Panoply?

It's a peer-to-peer **crowdsourcing** site, kinda like any other question and answer site on the Internet. Here's the twist:

Each time you submit a post or respond to someone, you're building powerful new skills to **reduce stress**. These are skills you can use throughout your daily life to boost your mood, manage your anxiety, and possibly even lower your risk for depression.

Here's how it works: Anytime you **submit a post**, you'll get responses back from other *Panoply* members like you. We'll try to help you wrap your head around what's stressing you out.

Anytime you **respond** on the site, you'll practice valuable stress management skills.

This will help train your brain to **think differently** about stressful situations. You might also help someone feel better. Corny as it sounds, performing small acts of kindness can actually **boost your well-being**.

We'll track your progress as you explore the site. You'll unlock **new skills** and you'll **earn points** each time someone likes one of your responses.

We'll even give you **points** just for doing this orientation! [*At this point in the tutorial, users are awarded 10 points and are shown a large celebratory image applauding their progress thus far. The points mechanism will be discussed in more depth later on.*]

Before you get started, it's important to remember that *Panoply* is still **an experiment**. The feedback you get will come from peers, many of whom are still learning, just like you.

We'll do our best to get you good responses *as quickly as possible*. Your job will be to **rate them**. See something lame? Rate it! See something awesome? Rate it! **Your feedback will help improve the site over time.**

Great Job! You've finished the orientation! We'll let you know when people respond to your post. In the meantime, check out the
practice session to interact with **real users!**

Appendix B

Supplementary Tables & Figures

Change Scores (Full Cohort)

Response Variables	Panoply Condition (n=84)		Writing Condition (n=82)		F-value	p-value	d [95% CI]
	M	SD	M	SD			
Positive Affect							
SHS	0.81	3.24	0.04	2.80	2.64	0.11	.25 [-.06, .56]
PA	1.32	6.96	0.57	6.12	0.53	0.47	.11 [-.19, .42]
Negative Affect							
CES-D	-3.6	9.86	-2.22	8.24	0.93	0.34	.15 [-.16, .46]
NA	-1.35	6.68	-1.45	6.33	0.01	0.92	.02 [-.29, .32]
Risk Factors							
ERQ-R	0.49	1.02	0.09	1.41	5.34*	**0.02**	.36 [.05, .67]
ERQ-S	-0.08	0.67	-0.07	0.56	0.28	0.87	.03 [-.28, .33]
PTQ	-4.42	10.13	-4.02	9.54	0.06	0.8	.04 [-.27, .35]

Table B.1: Changes from pre to post for all participants assigned to the *Panoply* and the expressive writing groups. F scores represent the interaction of Group (*Panoply*, Writing) × Time (Pre, Post).

SHS = Subjective Happiness Scale; PANAS-PA = Positive and Negative Affect Scale, Positive Affectivity; PANAS-NA= Positive and Negative Affect Scale, Negative Affectivity; CES-D = Center for Epidemiologic Studies Depression Scale; ERQ-R = Emotion Regulation Questionnaire, Reappraisal Frequency; ERQ-S = Emotion Regulation Questionnaire, Suppression Frequency; PTQ = Perseverative Thinking Questionnaire.

Change Scores (Depressed, Nondepressed)

Response Variables	Panoply Condition		Writing Condition		F-value	p-value	d [95% CI]
	Depressed (n=47)	Nondepressed (n=37)	Depressed (n=44)	Nondepressed (n=38)			
	M (SD)	M (SD)	M (SD)	M (SD)			
Positive Affect							
SHS	1.68 (3.45)	-.30 (2.60)	.41 (2.72)	.39 (2.86)	1.59	.21	.23 [-.08, .54]
PA	3.36 (7.45)	-1.27 (5.33)	1.55 (6.48)	-.55 (5.56)	1.60	.21	.19 [-.12, .5]
Negative Affect							
CES-D	-8.68 (9.18)	2.86 (6.32)	-5.07 (9.66)	1.07 (4.41)	4.72	**.03**	.33 [.02, .64]
NA	-3.49 (6.75)	1.37 (5.57)	-2.23 (7.34)	-.55 (4.81)	2.56	.11	.24 [-.07, .55]
Risk Factors							
ERQ-R	.62 (0.98)	.32 (1.05)	.33 (1.09)	-.17 (1.15)	0.34	.56	.08 [-.23, .39]
ERQ-S	-.14 (0.73)	-.01 (.57)	-.09 (0.56)	-.04 (.58)	0.17	.68	.09 [-.21, .4]
PTQ	-7.26 (10.94)	-.81 (7.73)	-3.86 (10.67)	-4.21 (8.18)	4.98	**.03**	.35 [.04, .66]

Table B.2: Changes from pre to post for depressed and nondepressed individuals assigned to the *Panoply* and the expressive writing groups. F scores represent the interaction of Group (*Panoply*, Writing) × Depression (Depressed, NonDepressed) × Time (Pre, Post).

154

Change Scores (Low, High Reappraisers)

Response Variables	Panoply Condition		Writing Condition		F-value	p-value	d [95% CI]
	Low (n=41)	High (n=43)	Low (n=28)	High (n=54)			
	M (SD)	M (SD)	M (SD)	M (SD)			
Positive Affect							
SHS	1.24 (3.28)	.40 (3.19)	-0.18 (3.09)	0.15 (2.66)	3.88	.05	.34 [.03, .65]
PA	2.34 (6.77)	0.35 (7.08)	1.86 (7.11)	-0.09 (5.50)	0.08	.78	.07 [-.24, .38]
Negative Affect							
CESD	-7.81 (8.86)	0.42 (9.14)	-2.57 (10.73)	-2.04 (6.71)	5.78	**.02**	.39 [.08, .70]
NA	-3.76 (5.99)	0.95 (6.55)	-0.21 (8.16)	-2.09 (5.11)	6.55	**.01**	.40 [.09, .71]
Risk Factors							
ERQ-R	1.04 (1.08)	-0.03 (.59)	0.67 (1.19)	-0.20 (1.00)	1.38	.24	.2 [-.11, .51]
ERQ-S	-0.11 (0.58)	-0.06 (0.74)	-0.08 (0.65)	-0.06 (0.52)	0.18	.68	.08 [-.23, .39]
PTQ	-8.07 (11.36)	-93 (7.37)	-1.25 (10.72)	-5.46 (8.62)	9.54	**.002**	.48 [.17, .79]

Table B.3: Changes from pre to post for low and high reappraisers assigned to the *Panoply* and the expressive writing groups. F scores represent the interaction of Group (*Panoply*, Writing) × Baseline Reappraisal Status (High, Low) × Time (Pre, Post).
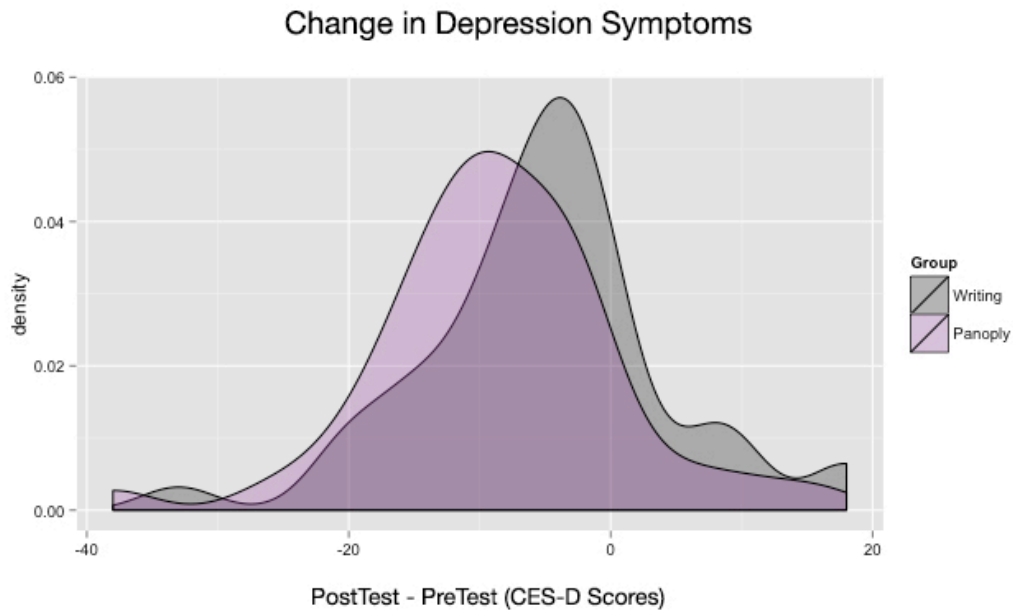
Figure B.1: A density plot illustrating changes in depression symptoms for individuals scoring above the clinical cut-off on the CES-D.
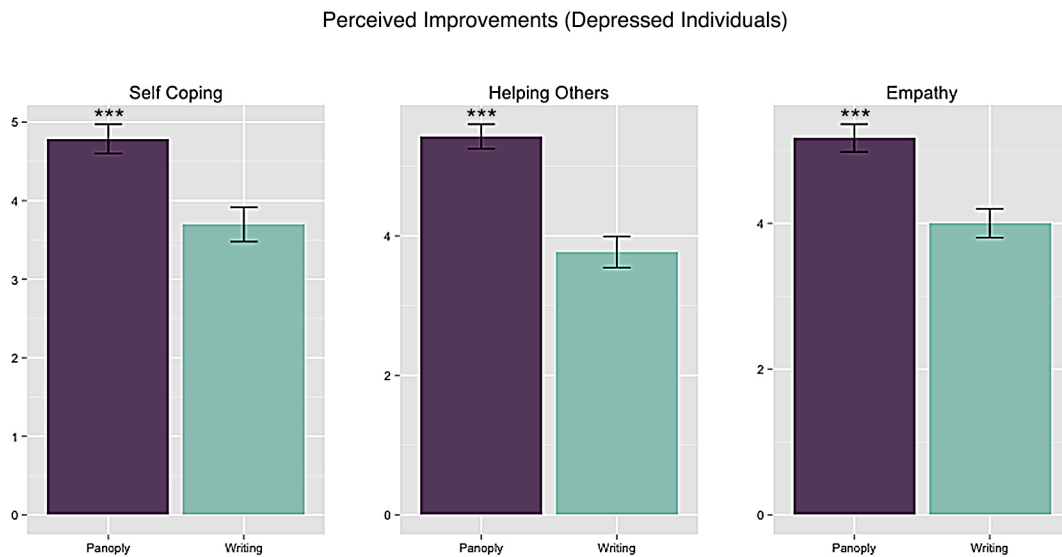
Figure B.2: Perceived improvements reported by depressed individuals in the *Panoply* and expressive writing groups. Error bars reflect 95% bootstrapped confidence intervals. Note: *** *p* < .001
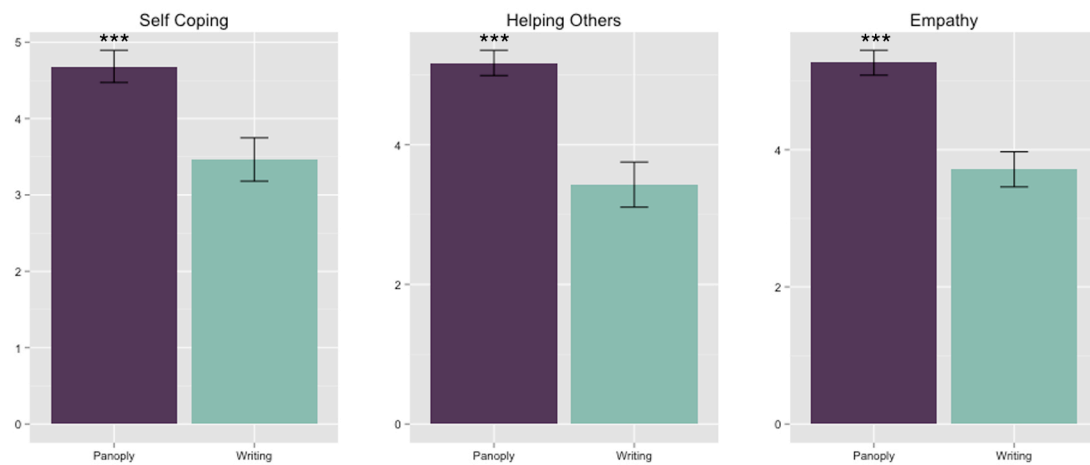
Figure B.3: Perceived improvements reported by low reappraisers in the *Panoply* and expressive writing groups. Error bars reflect 95% bootstrapped confidence intervals. Note: *** *p* < .001

References

Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A., & Ladwig, P. (2014).
The "Nasty Effect:" Online Incivility and Risk Perceptions of Emerging
Technologies. *Journal of Computer-Mediated Communication*, *19*(3), 373–387.
doi:10.1111/jcc4.12009

Baikie, K. A., & Wilhelm, K. (2005). Emotional and physical health benefits of
expressive writing. *Advances in Psychiatric Treatment*, *11*(5), 338–346.
doi:10.1192/apt.11.5.338

Beck, A. T. (1979). *Cognitive therapy of depression*. Guilford Press.

Beck, J. (2011). *Cognitive Behavior Therapy, Second Edition: Basics and Beyond*
(Second Edition, Second Edition.). The Guilford Press.

Benkler, Y. (2007). *The Wealth of Networks: How Social Production Transforms
Markets and Freedom*. Yale University Press.

Bernstein, M., Little, G., Miller, R., Hartmann, B., Ackerman, M., Karger, D., …
Panovich, K. (2010). Soylent: A Word Processor with a Crowd Inside. In
*Proceedings of the 23th annual ACM symposium on User interface software and
technology* (pp. 313–322). New York, New York: ACM Press.

Bernstein, M. S., Brandt, J., Miller, R. C., & Karger, D. R. (2011). Crowds in two
seconds: enabling realtime crowd-powered interfaces. In *Proceedings of the 24th
annual ACM symposium on User interface software and technology* (pp. 33–42).
New York, NY, USA: ACM. doi:10.1145/2047196.2047201

Bigham, J. P., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R. C., … Yeh, T. (2010).
VizWiz: nearly real-time answers to visual questions. In *Proceedings of the 23nd*

annual ACM symposium on User interface software and technology (pp. 333–342). New York, NY, USA: ACM. doi:10.1145/1866029.1866080

Blankers, M., Koeter, M. W., & Schippers, G. M. (2013). Baseline predictors of treatment outcome in Internet-based alcohol interventions: a recursive partitioning analysis alongside a randomized trial. *BMC Public Health*, *13*(1), 455. doi:10.1186/1471-2458-13-455

Brooke, J. (1996). *SUS: A quick and dirty usability scale*.

Bryant, J., & Jones, M. (2012). Responsive Web Design. In *Pro HTML5 Performance* (pp. 37–49). Apress. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4302-4525-4_4

Buchanan, K. E., & Bardi, A. (2010). Acts of Kindness and Acts of Novelty Affect Life Satisfaction. *The Journal of Social Psychology*, *150*(3), 235–237. doi:10.1080/00224540903365554

Burns, D. D. (1999). *Feeling Good: The New Mood Therapy* (Reprint.). Harper.

Butler, A. C., Chapman, J. E., Forman, E. M., & Beck, A. T. (2006). The empirical status of cognitive-behavioral therapy: a review of meta-analyses. *Clinical Psychology Review*, *26*(1), 17–31. doi:10.1016/j.cpr.2005.07.003

Cartreine, J. A., Ahern, D. K., & Locke, S. E. (2010, March 17). A Roadmap to Computer-Based Psychotherapy in the United States [research-article]. Retrieved April 30, 2013, from http://informahealthcare.com/doi/abs/10.3109/10673221003707702

Christensen, H., Griffiths, K. M., & Farrer, L. (2009). Adherence in internet interventions for anxiety and depression. *Journal of Medical Internet Research*, *11*(2), e13. doi:10.2196/jmir.1194

Christensen, H., Griffiths, K. M., & Korten, A. (2002). Web-based Cognitive Behavior Therapy: Analysis of Site Usage and Changes in Depression and Anxiety Scores. *Journal of Medical Internet Research*, *4*(1). doi:10.2196/jmir.4.1.e3

Christensen, H., Griffiths, K. M., Korten, A. E., Brittliffe, K., & Groves, C. (2004). A Comparison of Changes in Anxiety and Depression Symptoms of Spontaneous Users and Trial Participants of a Cognitive Behavior Therapy Website. *Journal of Medical Internet Research*, *6*(4). doi:10.2196/jmir.6.4.e46

Cobb, N. K., Graham, A. L., Bock, B. C., Papandonatos, G., & Abrams, D. B. (2005). Initial evaluation of a real-world Internet smoking cessation system. *Nicotine & Tobacco Research: Official Journal of the Society for Research on Nicotine and Tobacco*, *7*(2), 207–216. doi:10.1080/14622200500055319

Cohen, P. A., Kulik, J. A., & Kulik, C.-L. C. (1982). Educational Outcomes of Tutoring: A Meta-analysis of Findings. *American Educational Research Journal*, *19*(2), 237–248. doi:10.3102/00028312019002237

Cohen, S., & Wills, T. A. (1985). Stress, social support, and the buffering hypothesis. *Psychological Bulletin*, *98*(2), 310–357. doi:10.1037/0033-2909.98.2.310

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, *18*(1), 105–115. doi:10.1037/h0030644

Dinakar, K., Jones, B., Havasi, C., Lieberman, H., & Picard, R. (2012). Common Sense Reasoning for Detection, Prevention, and Mitigation of Cyberbullying. *ACM Trans. Interact. Intell. Syst.*, *2*(3), 18:1–18:30. doi:10.1145/2362394.2362400

Donkin, L., Hickie, I. B., Christensen, H., Naismith, S. L., Neal, B., Cockayne, N. L., & Glozier, N. (2013). Rethinking the Dose-Response Relationship Between Usage and Outcome in an Online Intervention for Depression: Randomized Controlled Trial. *Journal of Medical Internet Research*, *15*(10), e231. doi:10.2196/jmir.2771

Dontcheva, M., Morris, R. R., Brandt, J. R., & Gerber, E. M. (2014). Combining Crowdsourcing and Learning to Improve Engagement and Performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3379–3388). New York, NY, USA: ACM. doi:10.1145/2556288.2557217

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). New York, NY, USA: ACM. doi:10.1145/2145204.2145355

Ehring, T., Zetsche, U., Weidacker, K., Wahl, K., Schönfeld, S., & Ehlers, A. (2011). The Perseverative Thinking Questionnaire (PTQ): Validation of a content-independent measure of repetitive negative thinking. *Journal of Behavior Therapy and Experimental Psychiatry*, *42*(2), 225–232. doi:10.1016/j.jbtep.2010.12.003

Farvolden, P., Denisoff, E., Selby, P., Bagby, R. M., & Rudy, L. (2005). Usage and Longitudinal Effectiveness of a Web-Based Self-Help Cognitive Behavioral Therapy Program for Panic Disorder. *Journal of Medical Internet Research*, *7*(1). doi:10.2196/jmir.7.1.e7

Finkel, E. J., Slotter, E. B., Luchies, L. B., Walton, G. M., & Gross, J. J. (2013). A Brief Intervention to Promote Conflict Reappraisal Preserves Marital Quality Over Time. *Psychological Science*, 0956797612474938. doi:10.1177/0956797612474938

Fogg, B. (2009a). A Behavior Model for Persuasive Design. In *Proceedings of the 4th International Conference on Persuasive Technology* (pp. 40:1–40:7). New York, NY, USA: ACM. doi:10.1145/1541948.1541999

Fogg, B. (2009b). Creating Persuasive Technologies: An Eight-step Design Process. In *Proceedings of the 4th International Conference on Persuasive Technology* (pp. 44:1–44:6). New York, NY, USA: ACM. doi:10.1145/1541948.1542005

Forgas, J. P. (2002). Feeling and Doing: Affective Influences on Interpersonal Behavior. *Psychological Inquiry*, *13*(1), 1–28. doi:10.1207/S15327965PLI1301_01

Forgas, J. P. (2013). Don't Worry, Be Sad! On the Cognitive, Motivational, and Interpersonal Benefits of Negative Mood. *Current Directions in Psychological Science*, *22*(3), 225–232. doi:10.1177/0963721412474458

Frankl, V. E., Winslade, W. J., & Kushner, H. S. (2006). *Man's Search for Meaning* (1 edition.). Boston: Beacon Press.

Furukawa, T. A., Noma, H., Caldwell, D. M., Honyashiki, M., Shinohara, K., Imai, H., … Churchill, R. (2014). Waiting list may be a nocebo condition in psychotherapy trials: a contribution from network meta-analysis. *Acta Psychiatrica Scandinavica*, n/a–n/a. doi:10.1111/acps.12275

Garland, E. L., Gaylord, S. A., & Fredrickson, B. L. (2011). Positive Reappraisal Mediates the Stress-Reductive Effects of Mindfulness: An Upward Spiral Process. *Mindfulness*, *2*(1), 59–67. doi:10.1007/s12671-011-0043-8

Garnefski, N., & Kraaij, V. (2006). Relationships between cognitive emotion regulation strategies and depressive symptoms: A comparative study of five specific

samples. *Personality and Individual Differences*, *40*(8), 1659–1669. doi:10.1016/j.paid.2005.12.009

Goldin, P. R., Ziv, M., Jazaieri, H., Werner, K., Kraemer, H., Heimberg, R. G., & Gross, J. J. (2012). Cognitive reappraisal self-efficacy mediates the effects of individual cognitive-behavioral therapy for social anxiety disorder. *Journal of Consulting and Clinical Psychology*, *80*(6), 1034–1040. doi:10.1037/a0028555

Griffiths, K. M., Calear, A. L., & Banfield, M. (2009). Systematic Review on Internet Support Groups (ISGs) and Depression (1): Do ISGs Reduce Depressive Symptoms? *Journal of Medical Internet Research*, *11*(3). doi:10.2196/jmir.1270

Gross, J. J. (1998). Antecedent- and Response-Focused Emotion Regulation: Divergent Consequences for Experience, Expression, and Physiology. *Journal of Personality and Social Psychology*, *74*(1), 224–237.

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, *85*(2), 348–362.

Halperin, E., & Gross, J. J. (2011). Emotion regulation in violent conflict: Reappraisal, hope, and support for humanitarian aid to the opponent in wartime. *Cognition & Emotion*, *25*(7), 1228–1236. doi:10.1080/02699931.2010.536081

Heiy, J. E., & Cheavens, J. S. (2014). Back to Basics: A Naturalistic Assessment of the Experience and Regulation of Emotion. *Emotion (Washington, D.C.)*. doi:10.1037/a0037231

Hofmann, S. G., Heering, S., Sawyer, A. T., & Asnaani, A. (2009). How to handle anxiety: The effects of reappraisal, acceptance, and suppression strategies on anxious arousal. *Behaviour Research and Therapy*, *47*(5), 389–394. doi:10.1016/j.brat.2009.02.010

Horrigan, J. B. (2001). Online Communities: Networks that nurture long-distance relationships and local ties. *Pew Internet and American Life Project*.

Ipeirotis, P. G. (2010). Demographics of Mechanical Turk. *SSRN eLibrary*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1585030

Johnson, J. (2010). *Designing with the Mind in Mind: Simple Guide to Understanding User Interface Design Rules* (1 edition.). Amsterdam ; Boston: Morgan Kaufmann.

Kaltenthaler, E., Brazier, J., De Nigris, E., Tumur, I., Ferriter, M., Beverley, C., … Sutcliffe, P. (2006). Computerised cognitive behaviour therapy for depression and anxiety update: a systematic review and economic evaluation. *Health Technology Assessment (Winchester, England)*, *10*(33), iii, xi–xiv, 1–168.

Kaplan, K., Salzer, M. S., Solomon, P., Brusilovskiy, E., & Cousounis, P. (2011). Internet peer support for individuals with psychiatric disabilities: A randomized controlled trial. *Social Science & Medicine (1982)*, *72*(1), 54–62. doi:10.1016/j.socscimed.2010.09.037

Kashdan, T. B., & Rottenberg, J. (2010). Psychological flexibility as a fundamental aspect of health. *Clinical Psychology Review*, *30*(7), 865–878. doi:10.1016/j.cpr.2010.03.001

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., … Wang, P. S. (2003). The epidemiology of major depressive disorder: Results from the National Comorbidity Survey Replication (NCS-R). *JAMA: Journal of the American Medical Association*, *289*(23), 3095–3105. doi:10.1001/jama.289.23.3095

Kessler RC, Chiu W, Demler O, & Walters EE. (2005). PRevalence, severity, and comorbidity of 12-month dsm-iv disorders in the national comorbidity survey replication. *Archives of General Psychiatry*, *62*(6), 617–627. doi:10.1001/archpsyc.62.6.617

Kidd, C. D. (2008). *Designing for Long-term Human-robot Interaction and Application to Weight Loss*. Massachusetts Institute of Technology, Cambridge, MA, USA.

Kiluk, B. D., Sugarman, D. E., Nich, C., Gibbons, C. J., Martino, S., Rounsaville, B. J., & Carroll, K. M. (2011). A methodological analysis of randomized clinical trials of computer-assisted therapies for psychiatric disorders: toward improved standards for an emerging field. *The American Journal of Psychiatry*, *168*(8), 790–799. doi:10.1176/appi.ajp.2011.10101443

Kittur, A. (2010). Crowdsourcing, collaboration and creativity. *XRDS*, *17*(2), 22–26. doi:10.1145/1869086.1869096

Kittur, A., Chi, E. H., & Suh, B. (2008). Crowdsourcing user studies with Mechanical Turk. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (pp. 453–456). New York, NY, USA: ACM. doi:10.1145/1357054.1357127

Kittur, A., Nickerson, J. V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., … Horton, J. (2013). The future of crowd work. In *Proceedings of the 2013 conference on Computer supported cooperative work* (pp. 1301–1318). New York, NY, USA: ACM. doi:10.1145/2441776.2441923

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). New York, NY, USA: ACM. doi:10.1145/2047196.2047202

Kobak, K. A., Greist, J. H., Jefferson, J. W., & Katzelnick, D. J. (1996). Computer-administered clinical rating scales A review. *Psychopharmacology*, *127*(4), 291–301. doi:10.1007/s002130050089

Kokkalis, N., Köhn, T., Pfeiffer, C., Chornyi, D., Bernstein, M. S., & Klemmer, S. R. (2013). EmailValet: Managing Email Overload Through Private, Accountable Crowdsourcing. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (pp. 1291–1300). New York, NY, USA: ACM. doi:10.1145/2441776.2441922

Kummervold, P. E., Gammon, D., Bergvik, S., Johnsen, J.-A. K., Hasvold, T., & Rosenvinge, J. H. (2002). Social support in a wired world: use of online mental health forums in Norway. *Nordic Journal of Psychiatry*, *56*(1), 59–65. doi:10.1080/08039480252803945

Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and Evaluation of a User Experience Questionnaire. In *Proceedings of the 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society on HCI and Usability for Education and Work* (pp. 63–76). Berlin, Heidelberg: Springer-Verlag. doi:10.1007/978-3-540-89350-9_6

Linde, K., Fässler, M., & Meissner, K. (2011). Placebo interventions, placebo effects and clinical practice. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1572), 1905–1912. doi:10.1098/rstb.2010.0383

Linley, P. A., Nielsen, K. M., Wood, A. M., Gillett, R., & Biswas-Diener, R. (2010). Using signature strengths in pursuit of goals: Effects on goal progress, need satisfaction, and well-being, and implications for coaching psychologists. *International Coaching Psychology Review*, (5), 6–15.

Little, G., Chilton, L. B., Goldman, M., & Miller, R. C. (2010). TurKit. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology - UIST '10* (p. 57). New York, New York, USA. doi:10.1145/1866029.1866040

Locke, S. E., Kowaloff, H., Safran, C., Slack, W. V., Cotton, D., Hoff, R., … Page, P. (1990). Computer-Based Interview for Screening Blood Donors for Risk of HIV Transmission. *Proceedings of the Annual Symposium on Computer Application in Medical Care*, 835–839.

Lyubomirsky, S., Dickerhoof, R., Boehm, J. K., & Sheldon, K. M. (2011). Becoming happier takes both a will and a proper way: An experimental longitudinal intervention to boost well-being. *Emotion*, *11*(2), 391–402. doi:10.1037/a0022575

Lyubomirsky, S., & Lepper, H. S. (1999). A Measure of Subjective Happiness: Preliminary Reliability and Construct Validation. *Social Indicators Research*, *46*(2), 137–155. doi:10.1023/A:1006824100041

Mamykina, L., Manoim, B., Mittal, M., Hripcsak, G., & Hartmann, B. (2011). Design Lessons from the Fastest Q&a Site in the West. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2857–2866). New York, NY, USA: ACM. doi:10.1145/1978942.1979366

Marks I, Lovell K, Noshirvani H, Livanou M, & Thrasher S. (1998). Treatment of posttraumatic stress disorder by exposure and/or cognitive restructuring: A controlled study. *Archives of General Psychiatry*, *55*(4), 317–325. doi:10.1001/archpsyc.55.4.317

Marroquín, B. (2011). Interpersonal emotion regulation as a mechanism of social support in depression. *Clinical Psychology Review*, *31*(8), 1276–1290. doi:10.1016/j.cpr.2011.09.005

Mattick, R. P., Peters, L., & Clarke, J. C. (1989). Exposure and cognitive restructuring for social phobia: A controlled study. *Behavior Therapy*, *20*(1), 3–23. doi:10.1016/S0005-7894(89)80115-7

Mauss, I. B., Cook, C. L., Cheng, J. Y. J., & Gross, J. J. (2007). Individual differences in cognitive reappraisal: experiential and physiological responses to an anger provocation. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *66*(2), 116–124. doi:10.1016/j.ijpsycho.2007.03.017

McRae, K., Ciesielski, B., & Gross, J. J. (2012). Unpacking cognitive reappraisal: Goals, tactics, and outcomes. *Emotion*, *12*(2), 250–255. doi:10.1037/a0026351

Mohr, D. C., Duffecy, J., Ho, J., Kwasny, M., Cai, X., Burns, M. N., & Begale, M. (2013). A Randomized Controlled Trial Evaluating a Manualized TeleCoaching Protocol for Improving Adherence to a Web-Based Intervention for the Treatment of Depression. *PLoS ONE*, *8*(8), e70086. doi:10.1371/journal.pone.0070086

Morris, R.R., & McDuff, D.J (2014). Crowdsourcing techniques for affective computing. In Calvo, R., D'Mello, S.K., Gratch, J., & Kappas, A. (Eds.), *The Oxford Handbook of Affective Computing*.

Nolen-Hoeksema, S. (2000). The role of rumination in depressive disorders and mixed anxiety/depressive symptoms. *Journal of Abnormal Psychology*, *109*(3), 504–511. doi:10.1037/0021-843X.109.3.504

Ochsner, K. N., & Gross, J. J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences*, *9*(5), 242–249. doi:10.1016/j.tics.2005.03.010

Oulasvirta, A., Tamminen, S., Roto, V., & Kuorelahti, J. (2005). Interaction in 4-second Bursts: The Fragmented Nature of Attentional Resources in Mobile HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 919–928). New York, NY, USA: ACM. doi:10.1145/1054972.1055101

Panzarella, C., Alloy, L. B., & Whitehouse, W. G. (2006). Expanded Hopelessness Theory of Depression: On the Mechanisms by which Social Support Protects Against Depression. *Cognitive Therapy and Research*, *30*(3), 307–333. doi:10.1007/s10608-006-9048-3

Pennebaker, J. W., & Beall, S. K. (1986). Confronting a traumatic event: Toward an understanding of inhibition and disease. *Journal of Abnormal Psychology*, *95*(3), 274–281. doi:10.1037/0021-843X.95.3.274

Peterson, C., Stephens, J. P., Park, N., Lee, F., & Seligman, M. E. P. (2009). Strengths of character and work. In *Oxford Handbook of Positive Psychology and Work*. Oxford, UK: Oxford University Press.

Pinker, S. (2012). *The Better Angels of Our Nature: Why Violence Has Declined* (Reprint edition.). Penguin Books.

Poetz, M. K., & Schreier, M. (2012). The Value of Crowdsourcing: Can Users Really Compete with Professionals in Generating New Product Ideas? *Journal of Product Innovation Management*, *29*(2), 245–256. doi:10.1111/j.1540-5885.2011.00893.x

Radloff, L. S. (1977). The CES-D Scale A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement*, *1*(3), 385–401. doi:10.1177/014662167700100306

Reger, M. A., & Gahm, G. A. (2009). A meta-analysis of the effects of internet- and computer-based cognitive-behavioral treatments for anxiety. *Journal of Clinical Psychology*, *65*(1), 53–75. doi:10.1002/jclp.20536

Rimé, B. (2009). Emotion Elicits the Social Sharing of Emotion: Theory and Empirical Review. *Emotion Review*, *1*(1), 60–85. doi:10.1177/1754073908097189

Rogstadius, J., Kostakos, V., Aniket Kittur, Smus, B., Laredo, J., & Vukovic, M. (2011). An Assessment of Intrinsic and Extrinsic Motivation in Crowdsourcing Markets. In *ICWSM11*. Barcelona, Spain.

Rohrbeck, C. A., Ginsburg-Block, M. D., Fantuzzo, J. W., & Miller, T. R. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology*, *95*(2), 240–257. doi:10.1037/0022-0663.95.2.240

Rottenberg, J. (2014). *The Depths: The Evolutionary Origins of the Depression Epidemic*. New York: Basic Books.

Rubin, J., & Chisnell, D. (2008). *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. John Wiley & Sons.

Ryan, R., & Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*, *55*, 68–78.

Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The Motivational Pull of Video Games: A Self-Determination Theory Approach. *Motivation and Emotion*, *30*(4), 344–360. doi:10.1007/s11031-006-9051-8

Sapolsky, R. M. (2007). Stress, Stress-Related Disease, and Emotional Regulation. In *Handbook of emotion regulation* (pp. 606–615). New York, NY, US: Guilford Press.

Schueller, S. M., Muñoz, R. F., & Mohr, D. C. (2013). Realizing the Potential of Behavioral Intervention Technologies. *Current Directions in Psychological Science*, *22*(6), 478–483. doi:10.1177/0963721413495872

Schueller, S. M., & Parks, A. C. (2012). Disseminating Self-Help: Positive Psychology Exercises in an Online Trial. *Journal of Medical Internet Research*, *14*(3), e63. doi:10.2196/jmir.1850

Schwartz, B. (2005). *The Paradox of Choice: Why More Is Less*. Harper Perennial.

Seligman, M., Steen, T. A., Park, N., & Peterson, C. (2005). Positive Psychology Progress: Empirical Validation of Interventions. *American Psychologist*, *60*(5), 410–421. doi:10.1037/0003-066X.60.5.410

Sheppes, G., & Gross, J. J. (2011). Is Timing Everything? Temporal Considerations in Emotion Regulation. *Personality and Social Psychology Review*, *15*(4), 319–331. doi:10.1177/1088868310395778

Shirky, C. (2011). *Cognitive Surplus: How Technology Makes Consumers Into Collaborators*. Penguin Group (USA) Incorporated.

Sin, N. L., & Lyubomirsky, S. (2009). Enhancing well-being and alleviating depressive symptoms with positive psychology interventions: a practice-friendly meta-analysis. *Journal of Clinical Psychology*, *65*(5), 467–487. doi:10.1002/jclp.20593

Smyth, J. M., & Pennebaker, J. W. (2008). Exploring the boundary conditions of expressive writing: In search of the right recipe. *British Journal of Health Psychology*, *13*(1), 1–7. doi:10.1348/135910707X260117

Spek, V., Cuijpers, P., Nyklicek, I., Riper, H., Keyzer, J., & Pop, V. (2007). Internet-Based Cognitive Behaviour Therapy for Symptoms of Depression and Anxiety: a Meta-Analysis. *Psychological Medicine*, *37*(3), 319–328. doi:http://dx.doi.org.libproxy.mit.edu/10.1017/S0033291706008944

Suri, G., Whittaker, K., & Gross, J. J. (2014). Launching Reappraisal: It's Less Common Than You Might Think. *Emotion (Washington, D.C.)*. doi:10.1037/emo0000011

Troop, N., Schmidt, U., Tiller, J., Todd, G., Keilen, M., & Treasure, J. (1996). Compliance with a self-care manual for bulimia nervosa: predictors and outcome. *The British Journal of Clinical Psychology / the British Psychological Society*, *35 (Pt 3)*, 435–438.

Troy, A. S., Wilhelm, F. H., Shallcross, A. J., & Mauss, I. B. (2010). Seeing the silver lining: Cognitive reappraisal ability moderates the relationship between stress and depressive symptoms. *Emotion (Washington, D.C.)*. doi:10.1037/a0020262

Turkle, S., Taggart, W., Kidd, C. D., & Dasté, O. (2006). Relational artifacts with children and elders: the complexities of cybercompanionship. *Connection Science*, *18*(4), 347–361. doi:10.1080/09540090600868912

Vaish, R., Wyngarden, K., Chen, J., Cheung, B., & Bernstein, M. S. (2014). Twitch Crowdsourcing: Crowd Contributions in Short Bursts of Time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing System* (pp. 3645–3654). New York, NY, USA: ACM. doi:10.1145/2556288.2556996

Watson, D., Clark, L., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scale. *Journal of Personality and Social Psychology*, *54*, 1063–1070.

Wegner, D. M., Erber, R., & Zanakos, S. (1993). Ironic processes in the mental control of mood and mood-related thought. *Journal of Personality and Social Psychology*, *65*(6), 1093–1104. doi:10.1037/0022-3514.65.6.1093

Yu, L., & Nickerson, J. (2011). Cooks or cobblers? crowd creativity through combination. In *CHI 2011*. Vancouver, B.C.: ACM Press.

Zhang, H., Law, E., Miller, R., Gajos, K., Parkes, D., & Horvitz, E. (2012). Human computation tasks with global constraints. In *Proceedings of the SIGCHI*

Conference on Human Factors in Computing Systems (pp. 217–226). New York, NY, USA: ACM. doi:10.1145/2207676.2207708

Zichermann, G. (2011). *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps* (1 edition.). Sebastopol, Calif: O'Reilly Media.

Zittrain, J. (2009, December 7). The Internet Creates a New Kind of Sweatshop. Retrieved September 19, 2014, from http://www.newsweek.com/internet-creates-new-kind-sweatshop-75751