EXPLORABILITY, SATISFICING, AND SATISFACTION IN PARAMETER SPACES

by

Alberto Gonzalez

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Computing and Information Systems

Charlotte

2015

Approved by:

____
Dr. Celine Latulipe

____
Dr. Heather Lipford

____
Dr. Mary Lou Maher

____
Prof. Christopher Beorkrem

____
Prof. Heather Freeman

ABSTRACT

ALBERTO GONZALEZ. Explorability, satisficing, and satisfaction in parameter spaces. (Under the direction of DR. CELINE LATULIPE)

Many modern software applications have a fairly large set of variables that allow for the dynamic change of content. While a large parameter space provides users with fine-grained control over the content they are creating, even the best designed interfaces can't help users visit every possible combination of parameters. When there are more choices than can be feasibly explored, the user will go with the best option they have seen. I call this *exploratory satisficing.* When the complexity of interface options is too high, the user may be forced to give up and settle for a 'good enough' result (i.e., satisfice), when with another interface they might have explored more options or had a different threshold of 'good enough.' For some users, not exploring additional options leaves them feeling like they could potentially be missing out on the discovery of more desirable possibilities. Additionally, more options means these users could spend excessive amounts of time searching through the design space, hoping they haven't missed anything useful or interesting. My work on explorability, satisficing, and satisfaction is aimed at understanding the relationship between supporting the fluid exploration of large parameters spaces, the satisficing that results from interface and interaction design, and the user's satisfaction with both their finished work and creative process. I contribute insights about the relationship between explorability, satisfaction, and interaction design using exploratory satisficing as the basis for that understanding. I demonstrate that measuring explorability lets me

look at how engaged and immersed users are during the exploratory process as well as understanding the tradeoff users are making, borrowing from behavioral sciences and applying notions of satisficing and maximizing behavior.

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Celine Latulipe for her support and guidance throughout my time as a PhD student. I would also like to thank my dissertation committee: Dr. Mary Lou Maher, Dr. Heather Lipford, Professor Chris Beorkrem, and Professor Heather Freeman. Each of them has provided me with invaluable feedback and I could not have asked for a better committee. I would also like to thank Dr. Maher for letting me know my original dissertation topic could be improved.

I would like to thank my colleagues at UNCC: Professor Sybil Huskey for letting me be a part of her choreographic process, Dr. Vikash Singh for being my dissertation writing buddy, Dr. Michael Whitney for the deep (sometimes intellectual) conversations, Alexander Adams for his work that helped push my PhD forward, both current and previous members of the HCILab and InDe Lab, and my collaborators on the Dance.Draw and Tangible Creativity projects.

I would also like to thank my colleagues from Northwestern University, who encouraged me to pursue a PhD: Dr. Justine Cassell for seeing something in me and giving me my start in academia, Dr. Darren Gergle for connecting me to UNCC, and other friends/colleagues from the CTSB, Articulab, and Collabolab.

Finally I would like to thank my family and friends who supported me through this whole process, more specifically, my good friend Dan Anderson, for the emotional support that I often needed.

Some components of this dissertation work were supported by the NSF CreativeIT Grant #IIS-0855882 and the NSF REU Grant #IIS1156822.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Creative content generation tools once thought to be for "Professionals Only" have become widely available and more accessible to the everyday consumer, allowing hobbyists and enthusiasts to generate creative content on their own. One of the hurdles to using such tools is navigating the overwhelming number of possibilities made available by the tool. For example, Adobe® Photoshop® has a highly extensible filter and plugin library. Each filter can have several different options or parameters. Further, the effect of any particular setting may be dependent on the effects of other settings, leading to an colossal array of possibilities. There are so many possibilities that users are forced to satisfice, or choose options that are "good enough," even though a better solution may exist. My goal has been to understand the relationships between the design space users are exploring, where satisficing is occurring, as well as the effect that interface and interaction design have on a user's satisfaction with their final design and design process.

1.1 Exploring Parameter Spaces

The size and heterogeneity of a Design Space varies by domain. Design Spaces consist of all the possible configurations of every choice and option available to users (e.g., devices, materials, layouts, etc.) including the different connections that can be made between each of them [25]. For example, Card et al. researched the design space of input devices and contributed a taxonomy to support future designs [10]. They looked at the input mappings between different designs and quantified the design space along two indices: footprint and bandwidth. One of the design spaces I will be looking at in Chapter 6 of this dissertation is the design space of image creation. Beyond the computer interface, the larger design space includes variables such as print material, print size, display lighting, framing, etc. Therefore, I propose the use of the term, *Parameter Space*, to describe the realm of all possible combinations of variables that can be directly manipulated using the interface. For example, most photo editing applications have sliders for adjusting contrast, brightness, saturation, etc. The *parameter space* focuses specifically on the design space of possibilities that is offered by adjusting the variables available in the software. I make this distinction in order to describe the problem space in a way that is highly quantifiable.

I will discuss different parameter spaces throughout this proposal. One of the simpler examples to use is that of the Hue, Saturation, Brightness (HSB) color parameter space. One way to control HSB is by mapping each parameter to three different GUI sliders which adjust each parameter to integer values between 0 and 255. The total number of colors that can be created using these sliders is over 16 million. Most users are not going to want to attempt all 16 million colors. If the user has a general idea of what color they want, say green, they will move the sliders to create the nearest green color, and slowly refine that green until it is to their liking. Another strategy may be to 'jump' around the color space looking for different colors. Based on Schön's description of experiments, the former strategy of finding the nearest green can be called a move-testing experiment [45]. The latter strategy is more exploratory. The two approaches are not mutually exclusive, and many users may find themselves using both strategies intermittently. This combination of refining and jumping has been linked to creativity by Noy et al., who identify these types of behaviors as *scavenging* and *creative leaps*, respectively. This is the heart of explorability, to support a user's mental model of the parameter space that allows them the flexibility to scavenge and leap fluidly through the parameter space.

There is a good deal of tedious mode switching that occurs when you have to adjust the different sliders for HSB, which could deter from the experience of exploring the color space. In Chapters 3 and 5, I discuss a color exploration tool I created that enables fluid exploration of the HSB color space, without tedious mode switching. I also discuss findings from experimental studies of the color exploration tool.

1.2 Tyranny of Choice

Earlier I mentioned that while the HSB sliders enable users to view 16 million colors, most of them will prefer to visit a tiny subset of those colors. Herein lies the issue. Introducing users to the full spectrum of parameter space possibilities bares the risk of overwhelming the user with the sheer magnitude of possibilities to choose from. Barry Schwartz refers to this phenomenon as the 'tyranny of choice' [46]. Schwartz presents several lessons for handling situations where the number of possibilities is monstrous, one of which is *learning to accept "good enough."* This can also be described as 'satisficing' [49]. According to Schwartz, users fall into a spectrum between satisficers and maximizers [46]. Satisficers will go with a good enough option that is readily available and be satisfied. Maximizers will seek out new and better alternatives, but are often still thinking about what they've missed (i.e., they tend to be less satisfied).

In my work with Adams et al. [1], I have described the user's struggle to deal with an immense space of interacting parameter values as *Exploratory Satisficing*, where the user cannot possibly consider every alternative and *must be satisfied with a less than full exploration of the design space*. Later in this proposal, I'll elaborate further on the implications of large parameter spaces on satisficing and satisfaction. For now, I want to make it clear that users will at some point stop exploring these large parameter spaces. Numerous factors contribute to satisficing (e.g., time, fatigue), with personality playing a major role as well [47]. I am interested in understanding how an interface's design or an interaction technique contributes to when the user stops exploring. This is important to knowing how to better support users creating content in the face of enormous parameter spaces, as well as better understand what helps make that exploration less monotonous and more fluid and satisfying.

1.3 Motivation

Many modern software applications have large sets of variables that allow for the dynamic change of content. While a large parameter space provides users with fine-grained control of their content, naive design approaches can result in interaction techniques and interfaces that are excessively complex. Applications should support the fluid exploration of large parameter spaces. Additionally, the design of the interface should account for exploratory satisficing, supporting the user in reaching their threshold for what is acceptable in the quality of their creative content. Most of the current metrics focus on helping the user finish a task more quickly. In the case of creativity support tools, immersion can be a much more relevant factor than time [11]. There is evidence to suggest that user satisfaction is impacted by when a user chooses to satisfice [47, 23, 14] and the amount of options (i.e., explorability) an interface presents [42]. In addition, spending more time on a creative task is not a negative if the results are worth the effort [11]. I am looking at the combination of explorability and satisficing in parameter spaces as a way to evaluate interfaces. I believe the goal of interface designers should be to understand the role of the interface in satisficing. Users should only be satisficing because of external constraints and preferences, not because the parameter space exploration is difficult, slow, or tedious.

1.4 Thesis Statement

> *Interaction techniques and size of search spaces (defined by the granularity of control) for the exploration of large parameter spaces can positively and negatively effect the amount of options users explore as well as the users' satisfaction with their creative process and final designs. Some of these effects impact Maximizers differently than Satisficers.*

1.5 Contributions

1. Evidence to support my claim that explorability is a fundamental aspect of interaction and usability that cannot be captured by standard usability metrics (e.g., time and error rates), as well as a high level approach to measuring explorability, illustrated with domain-specific examples.

2. Groundwork for understanding the relationship between the level of explorability an interface supports, the satisficing that results, and the user's satisfaction with their work. This includes measuring a user's inclination towards satisficing and indirectly measuring the satisficing that occurs within interfaces.

3. Design implications for how interfaces may need to adapt in order to support parameter space exploration for all users who are satisficers, maximizers, or somewhere in between.

1.6 Dissertation Overview

Chapter 2 contains relevant background to the work I did: creativity support tools, interaction techniques, cognitive aspects of exploration, measuring satisfaction, experimentation, and satisficing. Chapter 3 provides a description of the Bimanual Color Exploration Plugin (BiCEP) [19], a tool that supports fluid exploration of the HSB color space. This chapter also includes a study I conducted for evaluating the usability and creativity support of BiCEP. I discuss the results of the study and its implications for design. While this study focused on the efficiency of movement in a parameter space, the analysis and results were a stepping stone to the more in-depth studies of exploration done in Chapters 5 and 6. Furthermore, the study in Chapter 5 also includes the use of BiCEP tool. Chapter 4 serves as an introduction to my approach to understanding the topics of explorability, satisficing, and satisfaction in parameter spaces. It includes the measurements I use, the rational behind them, and the research questions that drive this work. This chapter serves as a summary of my methodology and approach, and can be used by other researchers to understand the role of explorability, satisficing, and satisfaction in their study of interfaces that support the exploration of large parameter spaces. Chapter 5 presents a follow up study of BiCEP designed to show how different interaction styles impact the core topics of this dissertation. This study compares two different interaction techniques for exploring the same size search space. In this study, I present findings indicating that using a novel interaction technique can increase how much of the parameter space was explored and can increase participants' satisfication with the final design. Chapter 6 includes a two-part study that examines the differences in the search space enabled by choosing different granularities of control with the UI widgets. This study compares two different search space sizes of the same parameter space, while keeping the interaction technique the same. In this study, I present findings indicating that reducing the search space size by decreasing the granularity of control can decrease user satisfaction with their creative process. However, I demonstrate that users did see differences in how they explored the parameter space and recognized the potential benefits. In Chapter 7, I answer the research questions from Chapter 4 using the results from the studies done in Chapters 5 and 6. I also discuss future work and summarize the contributions of my dissertation.

CHAPTER 2: BACKGROUND

In this chapter, I provide a wide range of background literature that supports the work I will present in later chapters. I describe previous research of various interaction techniques and modalities that support the exploration of parameter spaces. I summarize research from behavioral economics and psychology that characterizes people's tendency to exhibit maximizing and satisficing behavior, as well as their inclination towards general satisfaction. I also present background literature where the authors have tried to bridge some of these topics. I have used these works to derive my study protocols and materials and ensure a strong foundation for my research.

2.1 Bimanual Interaction

Background literature on bimanual interaction has been included because it was the basis of BiCEP and SonicExplorer, two novel techniques that allow users to explore large parameter spaces more fluidly. There are three important aspects or categorizations of bimanual interaction to understand with respect to the current research: parallel versus serial interaction, symmetric versus asymmetric interaction, and direct versus indirect interaction.

2.1.1 Symmetric versus Asymmetric Interaction

Most early bimanual interaction work followed the paradigm of asymmetric interaction described by Guiard in his kinematic chain model [20]. In asymmetric interaction, the dominant hand does the detail work, while the non-dominant hand plays a supporting role. Typically, the non-dominant hand begins the interaction by setting the frame of reference in which the dominant hand works. Many of our daily activities can be characterized as asymmetric, but there are a variety of situations in which we use our hands more symmetrically, either temporally and/or spatially. Latulipe's work focused on modeling symmetric interaction in the human-computer interface, stating that symmetric interaction can be thought of as a superset for bimanual interaction styles [28]. Allowing symmetric interaction through the use of two identical devices and dual cursors affords the developer the ability to design an interaction either symmetrically or asymmetrically, which affords the user the ability to choose whether they want to interact more symmetrically or asymmetrically [28]. When using two different devices for an asymmetric interaction technique, only one cursor is provided by most modern operating systems and a second device is limited to a supporting role such as panning or zooming. This type of system does *not* allow the user to interact symmetrically. Both symmetric and asymmetric bimanual interaction can be used to support the user as they explore large, complex, parameter spaces where there can be variances in the work load for each hand. Supporting both interactions empowers the user to choose the interaction that they feel works best and to fluidly switch between different styles of interaction.

2.1.2 Parallel versus Serial Interaction

Typically, the benefits and costs of using a bimanual interaction technique are either physical or cognitive and can often be seen as the interaction actually plays out:

whether the two hands work simultaneously in parallel, or whether they work in serial. There can often be time saved if a task is unified bimanually and the user does not have to mode switch (e.g., from resizing an object to positioning that same object) by stopping their task and clicking on a button, object handle, or menu item. These interruptions can increase cognitive load, physical effort, or task time. Leganchuk et al. showed cases where even after removing the time necessary for a mode-switch in the unimanual control technique, the bimanual techniques were faster [33]. There can also be time saved with bimanual interactions due to parallel interaction allowing more to be accomplished at once. However, Balakrishnan and Hinckley showed that parallel interaction is most likely to occur when the task of the two hands is perceptually unified: the two hands have to be working together on a common task, otherwise interaction becomes serialized, with each hand taking turns to move [3]. Latulipe showed that a poorly constructed bimanual interaction design can lead to worse performance on a task than a traditional single cursor interaction, while a well constructed bimanual design can lead to significant performance improvements on that same task [30]. Both parallel and serial bimanual interactions can support the user as they explore large, complex, parameter spaces by decreasing the need for tedious mode switching. Both interactions empower the user to choose the interaction they feel is most appropriate for their current task.

2.1.3 Direct Bimanual Interaction

Multitouch interaction can be direct (such as on iPads, smart phones or interactive tabletops), or indirect, as when used with multiple devices attached to a laptop or desktop computer. One example of direct bimanual interaction is the work of Brandl et. al which demonstrates the advantages of two-handed interaction using pen and touch on a large, direct-input surface [7]. As part of their work, the authors also presented an HSV color picker that affords pen and touch control of the Hue and Saturation-Value, simultaneously. Direct interaction has the benefits of providing direct manipulation with no intermediary device, but can sometimes be ergonomically demanding and lead to issues of arm fatigue. Direct interaction can also have issues due to fingers and arms occluding the content. Indirect interaction has fewer ergonomic issues and there are no occlusion problems. Indirect interaction on the other hand affords purposeful manipulations of the mapping between device movement and cursor movement, such as cursor warping and dynamic control-display gains [18].

As tablets have become more commonplace, applications for devices such as the iPad have provided users with control over their creative content either on the tablet (e.g., DJ Mixing) or in combination with a PC (e.g., MIDITouch). These devices allow for direct manipulation of parameter space controls (e.g., sliders), enabling interaction that is not available using a standard keyboard and mouse controlled PC. For example, in the app NodeBeat HD, users manipulate sounds and beats by moving dots around the tablet screen. Users can move as many dots as they have fingers and do so in parallel. The direct manipulation also allows for DJ mixing where the user can move multiple dots in a rhythmic pattern that would be virtually impossible with a single point of control provided by a mouse.

2.1.4 Indirect Bimanual Interaction

Bimanual interaction can be indirect, i.e., using two mice, or two fingers on a touchpad or trackpad. Previous studies on bimanual input involving touchpads or trackpads [6, 9] have mostly involved two hands using separate and different devices in an asymmetric fashion [20]. The work of Moscovich and Hughes created multiple cursors for the fingers in contact with the large external trackpad, but was strongly dependent on interpreting gestures [39]. Hutterer worked on groupware support for multiple cursors at the window management system level, allowing multiple indirect devices to connect to the system [22].

My previous work includes the creation of a color selection tool that support bimanual interaction on the trackpad [19] and a sound exploration tool for exploring MIDI audio parameters using geospectral metering and bimanual interaction on the trackpad [1]. Both tools support indirect bimanual interaction and parallel exploration of multiple parameters at the same time. Bimanual exploration is relevant to exploration as it is an interaction technique that can change the speed and fluidity at which users explore a parameter space. Furthermore, if the interaction technique is more engaging, it could impact when participants decide to satisfice.

2.2 Exploring Possibilities

Schön describes three types of experiments for exploring ideas or actions: hypothesis-testing, move-testing, and exploratory [45]. Hypothesis-testing involves the comparison of competing hypothesis to see which is true. Move-testing involves testing whether a particular stimulus or action results in an expected change. Exploratory experimentation is characterized by a lack of any particular expected outcome. Many professional users are performing hypothesis-testing or move-testing when manipulating parameters in an interface, while many hobbiest and novice users perform exploratory experimentation when manipulating parameters in complex creative content generation interfaces.

The work of Terry et al. discusses the exploration of parameter spectrums using SideViews [52]. For example, SideViews shows the user an array of image previews along a parameter spectrum, demonstrating how changes to the parameter will affect the image. SideViews can show several previews to let the user see how blurry an image will get when the user changes the blur strength. In later work, Terry et al. demonstrate how users can explore design alternatives by simultaneously visualizing and comparing different manipulations of the same image using parallel pies [53].

Chaudhuri et al. created AttribIt, a system of navigating the parameter space of virtual models using semantic attributes [12]. For example, if the user is trying to make a dangerous looking animal, they simple increase the 'dangerousness' attribute using the appropriately named slider. The system is currently limited to only predefined models that have been ranked for their semantic attributes. While this allows for a broader exploration of the parameter space, it does not yet support fine-grained exploration. However, the author's experiments still suggest that AttribIt is an effective tool for allowing a novice to explore a vast combination of virtual components.

Noy et al. describe an approach to measuring creative leaps during a study where participants discovered creative shapes while exploring a design space of shapes [41]. During the study, participants would linger in a particular area of the design space.

They would look for nearby related design shapes, ‘scavenging’, and then move to another area of the design space, ‘creative leaps’.

2.3 Satisficing and Maximizing

Herbert Simon coined the term for the strategy of decision making known as *Satisficing* [48, 49], a combination of the words ‘satisfy’ and ‘suffice.’ He suggested that the combination of limited information and limited memory makes evaluating all possible outcomes difficult. As humans, we often employ a satisficing strategy by considering multiple alternatives until one of those alternatives meets a threshold that is deemed acceptable (for a given context).

Besides choosing options, satisficing includes choosing particular actions. Bendor et al. informally describe satisficing as existing in two parts [4]. After performing a particular action, if that action results in an outcome that meets or surpasses the user’s expectations, the user will continue to perform that same action. Otherwise, if the outcome of the action falls below their expectations, they will look for and try a new and different action.

Schwartz provides four lessons on what to do when faced with a situation where there is an overwhelming amount of options [46]. One of them is “Learn[ing] to accept ‘good enough’ ” where you find the option that best meets your requirements and avoid thinking about the best option, not unlike satisficing. Other lessons include restricting the number of options available when possible, avoid thinking about the options you did not choose (or know about), and finally lowering your expectations.

People classified as satisficers can be characterized as having personalities that lend themselves to a lower threshold of acceptability, desiring to find a ‘good enough’ option. Opposite them, are the maximizers, users who want to have a higher threshold of acceptability, desiring to find the ‘optimal’ choice. Schwartz et. al developed a 13-item Maximization Scale for determining where on the spectrum between satisficer and maximizer a participant’s personality could be found [47]. This scale has since been refined to a 6-item scale [40]. According to Schwartz [46], among the thousands surveyed, eighty percent fell in a range between 2.5 and 5.5, with anything outside that range being extreme satisficers or extreme maximizers, respectively. Most of the participants (i.e., two-thirds) were outside the middle range of the scale (3.25 and 4.75). While Schwartz categorizes individuals into satisficers and maximizers, his results suggest that users could be divided into third categories.

2.3.1 Satisfaction

Using the Maximization scale, Schwartz et al. found maximizing often correlated with negative emotions (e.g. depression) and inversely correlated with more positive emotions (e.g., satisfaction, happiness, etc) [47]. When offered the opportunity, maximizers tended to sacrifice resources to try new options, but in the end were not as satisfied as satisficers who settled for a ‘good enough’ option [14]. Even in cases where the decision making strategies of maximizers had a measurably more ‘optimal’ option (e.g., higher payoff), maximizers were still less satisfied than satisficers [23]. Misuraca and Teuscher found that maximizers tended to underestimate the time spent making decisions. They postulate this is the consequence of the cognitive load maximizers bear as they process higher amounts of information to find the best option [38].

2.4 Understanding Explorability and Satisficing

My broader research goal is to understand how interaction techniques and interface designs that support explorability can impact satisficing. In the following sections, I discuss various interaction techniques that are commonly used for traversing parameter spaces (i.e., support explorability) and examples of where the interface has been shown to impact satisficing.

2.4.1 Explorability in Interfaces

When users engage in Schön's exploratory experimentation, their actions are mediated by the GUI controls, which will have an impact on the ease of exploration. Exploration is often put to use for tasks that are open-ended, where users do not have a clearly defined path to reaching a 'finished' creation or solution [52]. Simon describes the process of an architect designing a house and notes that the path an architect takes will not be known from the start [50]. Some of the information that impacts the architect's direction is only discoverable (often by accident) after a good deal of the problem space has already been explored. This suggests that exploring more of the design space may lead to more desirable end results. The importance of explorability in the interface is highlighted by Carroll et al., who found through their development of the Creativity Support Index that exploration is one of six orthogonal factors relevant to supporting creativity work [11].

Unfortunately, spending more time and cognitive resources on exploring a large parameter space is often impractical for users. For example, the Apple® Preview application provides an image color adjustment dialog. This dialog (see Figure 1(a)) has

9 standard UI sliders, each controlling a different parameter (e.g., exposure, sepia) and each parameter has approximately 200 possible values. The parameter space is therefore composed of $9^{200} = 5.12 * 10^{12}$ different combinations of values. Realistically, a subset of these combinations will be either *perceptually indistinguishable* or *inauspicious* (e.g., –100% exposure results in an all black image, rendering the usefulness of other parameters moot). However, even if those subsets are removed, it is still unlikely that users will be able to visit all of the remaining possible parameter combinations. The sheer immenseness of the possibilities available may be overwhelming, especially to novice users. This predicament can also be referred to as the 'tyranny of choice' [46]. This is especially a problem when designers assume that an application can overcome our physical limitations (i.e., visual awareness) with a well-organized display of information [54]. Just because the system is clean and allows the entire space of possibilities to be observably explorable, does not mean the user's cognitive abilities will be able to process (let alone transmit) all the visual information they are being presented with. There is also the problem of permanence of choice, some applications do not have a features such as a history or saved states that help temper the exploration of choices. To help them achieve their end goal and overcome the 'tyranny of choice,' users are forced to use strategies such as 'satisficing.'

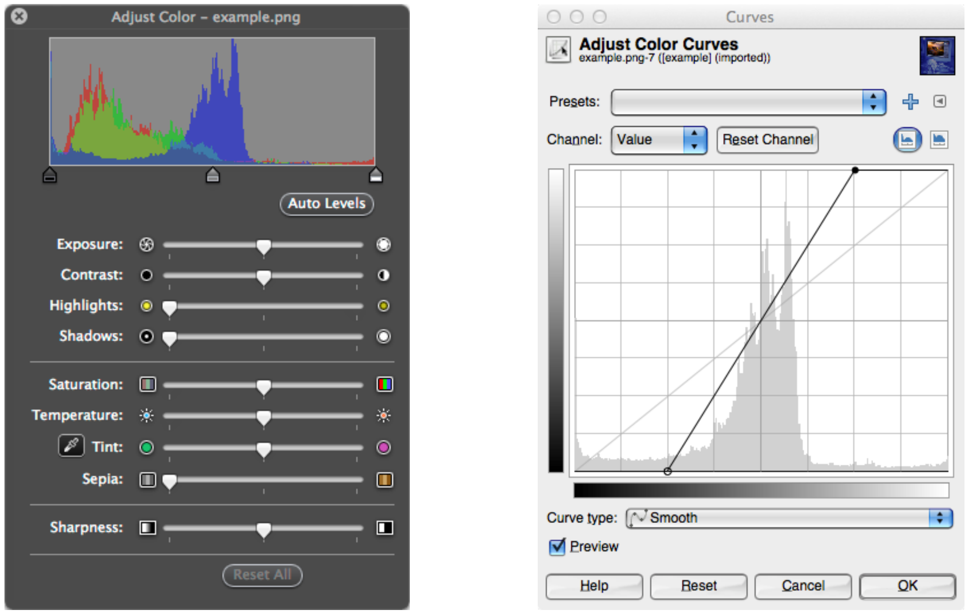
2.4.2 Satisficing in Interfaces

While satisficing is generally understood as a decision making approach to an optimization problem, in my work with Adams et al., I discuss the implications of satisficing in computer interfaces [1]. The interface we created leveraged bimanual interaction, color cues, and spatial memory to support the exploration of MIDI sound parameters. We posited that the limitation imposed by interfaces with a single-point of control (i.e., a single cursor) meant the user will explore less of the design space than if the user had multiple points of control (i.e., multi-cursor). Participants used the tool for creative exploration of sound parameters and provided feedback on their exploration. While we demonstrated how our interface supported sound-parameter exploration, we did not fully measure the amount of exploration completed.

At some point, most users decide that their created content meets some threshold of acceptability and stop exploring the parameter space within their task. What they consider ‘good enough’ will vary based on any number of factors: time, effort, compensation, personal preference, experience, intended audience, personality, whether or not they had an end goal pre-envisioned, etc. My goal is to understand the factors that contribute to satisficing and characterize the impact of the interface and interactions on this stopping point.

For example, if a task has a long response or feedback time (e.g., rendering time for photo filters), the user may stop exploring because they have simply run out of time to try out different ideas. There is also the possibility that the user stops after recognizing that further refinement of their design will have diminishing returns and therefore decides that any potential improvement on their work is not worth the additional effort required, a tradeoff that is considered when evaluating how well an interface supports creativity [11]. The work of Malacria et al. demonstrates this same phenomenon occurring when users ‘satisfice’ with respect to learning expert features of an interface (i.e., hotkeys), often choosing to stick to their more novice behavior (i.e., mouse clicking) [36]. Their application encourages users to use hotkeys to reduce how much users 'satisfice' on their performance when retaining novice behaviors. They postulate that at the core of this issue is the unknown costs or benefits of investing time and effort into learning the advances features, similar to the 'Results Worth Effort' factor in Carroll's work [11].

The work of Oulasvirta et al. found a difference in user satisfaction when they varied the number of search results in a time-restricted task [42]. Participants were shown the results of an internet search query and given 30 seconds to choose the best result. Of the 24 participants, 12 used an interface that would only display 6 result items, and the rest used an interface that displayed 24 result items. Participants who selected from 6 items were more satisfied and more confident that their selection was correct compared to the participants who selected from 24 items. They provide a table of variables present and controlled for in their experiment that could contribute to 'choice overload.' Some of these include: personality, domain knowledge, skill set, time in task, etc. [42]. While the authors used the Maximization Scale [47] as part of their study, none of their participants fell into the category of maximizers. While they were able to divide participants into sub-groups of satisficers, they comment that the homogeneity of their participants was a major factor in not finding any reliable differences based on their tendencies to 'satisfice.' The authors suggest that further analysis of this dichotomy in personalities could yield interesting results.



(a) Apple Preview's color adjustment dialog

(b) GIMP's color curves tool

Figure 1: Color adjustment tools from Apple Preview and GIMP

2.4.3 Interaction Techniques

The interaction technique used to traverse the parameter space could have a significant impact on exploratory satisficing. In the Apple Preview example (see also Figure 1(a)), each parameter is adjusted using a single dimension of control (e.g., a slider). Both gross exploration and fine tuning of designs often requires mode-switching back and forth between different sliders. One solution is to map the parameters to a multidimensional form of control, like a curves tool (see Figure 1(b)), which provides pixel level input and output adjustments (i.e., tones). The curves tool allows for a finer degree of control over parameters such as contrast. The trade-off is the loss of linear mappings for individual parameters and forcing the user to comprehend the computational model for the curves tool. For example, the curve (in this case a straight line) in Figure 1(b) increases the image contrast. But, without textual or strong visual indicators, only highly proficient users of the curves tool would be able to recognize or replicate this curve for modifying the contrast.

Latulipe et al. describe a technique for exploring tone-mapping using bimanual interaction (i.e., dual-mouse, dual-cursor) to control a spline curve [29]. The symTone technique enables fluid exploration of tone-mappings using two points of a spline curve as opposed to the traditional click-and-drag technique that requires mode-switching.

In my work with Adams et al., we created the SonicExplorer interface that combines multi-dimensional space (like the curves tool) and bimanual interaction [1]. SonicExplorer allows assigning up to 4 different sound parameters to each of the 4 dimensions of control: an X_1Y_1 controlled by the left hand and a X_2Y_2 controlled by the right hand. The interface is well constructed for bimanual interaction [30] and provides the parallel interaction that perceptually unifies the two hands working together [3]. This supports a fluid traversal of the parameter space of sound manipulations.

The previously mentioned work of Chaudhuri et al. has an interface that allows users to adjust the look and feel of a virtual model by adjusting a slider mapped to semantic attributes. The computer handles the computation of ranking the animals on “how dangerous they look” based on a database of attribute rankings for each model. This technique offloads cognition by giving the computer the task of ranking and modeling semantic judgements.

The previous interaction techniques have focused on traversing the parameter space, but design also includes the side-by-side consideration of alternatives. Our limited memory makes remembering our design process difficult. Users will often save different iterations of their work (e.g., as layers or files) that they can go back to for consideration. Therefore, alternative designs becomes another dimension of the parameter space, bringing with it the tediousness of trying to open, layout, and arrange the alternatives for comparison. Terry et al. created Parallel Pies for comparing particular combinations of parameters (i.e., alternatives designs) [53]. Alternatives are displayed in the various parts of the pie and as the user moves and rotates the pie, they can compare the alternatives in the same drawing space.

The interaction techniques mentioned above support the exploration of the parameter space through intuitive interaction and interface mapping or offloading the computational work to the computer. They also put less stress on the user by providing engaging and fluid interaction and offloading computationally intensive tasks to the computer. These traits help to reduce the role the interface plays in satisficing, so it originates mostly from external factors.

In the next chapter, I present a user study of the BiCEP color selection tool, which helped me understand some of the complexities of measuring and observing explorability and satisficing in interfaces. To help lay the groundwork for understanding and developing some preliminary metrics for explorabilty and satisficing, Chapter 4 highlights the methodology and constructs I developed to deepen my investigation of the relationship between explorability, satisficing, and satisfaction. I then utilize these methods and ideas in my studies presented in Chapters 5 and 6.

CHAPTER 3: BIMANUAL COLOR EXPLORATION PLUGIN

In this chapter, I describe the Bimanual Color Exploration Plugin (BiCEP), a tool that supports fluid color exploration, and a user study that used explorability as one of its metrics. The goal was to perform a usability study of a novel interaction technique for exploring color and demonstrate that users can explore colors more fluidly using BiCEP. While the results of the usability study support my hypotheses, it became clear during the analysis of the data that measuring exploration was more complex than simple usability metrics (e.g., time). This chapter sets the stage for the more in-depth studies of exploration done in Chapters 5 and 6.

3.1 BiCEP

BiCEP is a Mac OS X color picker plugin that enables dual-cursor color selection using two fingers on the trackpad [19]. There is no need for custom dual-cursor software that works external to an application or requires changing the source code. BiCEP is available to any application that uses the system color picker.

BiCEP allows for the traversal of a very specific and unified parameter space, the Hue, Saturation, and Brightness (HSB) color space. These three parameters can be controlled in different ways, including but not limited to sliders that control separate color parameters (Figure 2a), a color spectrum that unifies all the HSB parameters (Figure 2b), or a 2D Hue and Saturation color wheel and 1D Brightness

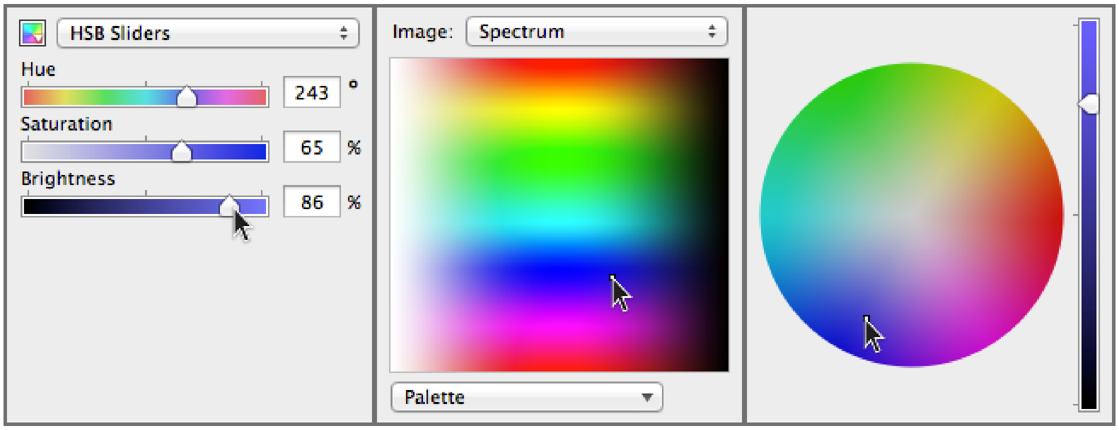


Figure 2: Native Mac OS X HSB color selectors (left to right) a) components, b) spectrum, and c) color wheel and brightness bar

Bar (Figure 2c). In section 2.4.3, I explained that separate sliders require tedious mode switching between the parameters and therefore inhibit explorability. While the color spectrum mimics the unity of the curves tool in section 2.4.2, the spectrum maps a two-dimensional controller (i.e., the cursor) onto a three dimensional space, which could potentially increase the cognitive load placed on the user.

Unlike other previous dual cursor applications [31, 29], no additional hardware (i.e., a second computer mouse) is needed. When a user places two fingers on the Apple trackpad, they control two cursors within the plugin (see Figure 3). The right cursor, controlled by the index finger of the right hand, moves within the bounds of a color wheel, adjusting the hue and saturation values. The left cursor, controlled by the index finger of the left hand, moves within the bounds of a rectangular bar, adjusting the brightness value. The arrangement of the color wheel on the right and the brightness bar on the left follows a right-handedness mapping, where the non-dominant hand does a less detailed task. The dominant hand controls the color wheel which has two spatial dimensions (X and Y), while the non-dominant hand controls

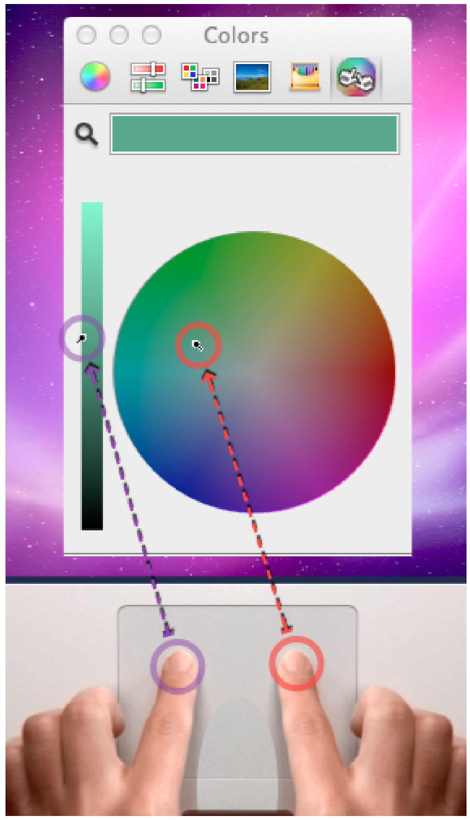


Figure 3: Mapping of left and right fingers on an Apple Trackpad to the BiCEP cursors displayed in the brightness bar and color wheel, respectively

the brightness bar which has a single spatial dimension (Y). At the very top there is a color bar that represents the currently selected color, a combination of the hue and saturation of the color wheel with the brightness of the color bar. The magnifying glass at the top-left corner is a color eyedropper tool created by the operating system and is a required component of all Mac OS X 10.8 color picker plugins.

3.1.1 BiCEP Interaction

While interaction with BiCEP starts with placing two fingers on the trackpad, I did not want to require users to always have their two index fingers on the trackpad in order to remain in control of BiCEP. Therefore, I came up with an appropriate metaphor: the bicycle handle-bar (see also Adams et al. [1]). While riding a bike, both hands can steer the bike by holding each end of the handlebar, but the cyclist can also steer the bike with only the right hand, allowing the left hand to let go of the handle bar. The cyclist can then return the left hand back onto the handle bar and let go with the right hand. While riding, as long as one of the hands is on the handlebar, the cyclist remains in control. However, once both hands let go of the handle bar, steering is not something they can do.

This metaphor is translated to BiCEP interaction in the following way: once the user places two index fingers on the trackpad, dual-cursor control in BiCEP is engaged and the user can drive two separate cursors inside the plugin window (see Figure 3). The user can then lift one of the index fingers, at which point the cursor that finger was controlling is anchored in place and the user can still use the other finger to move the other cursor. The user can then put the index finger back down to control both cursors again. The user can then lift the other index finger, all the while remaining in control of BiCEP. Only after the user lifts both index fingers will dual cursor mode disengage. When dual cursor mode is off, a finger touching the trackpad controls the system cursor. While controlling the BiCEP cursors, the system cursor is hidden to prevent distraction and anchored to the center of title bar of BiCEP. The system cursor is technically still 'active' and could bring another window into focus if it leaves the BiCEP window. In theory, the system cursor could be anchored anywhere in the window where there is no actionable mouse events (e.g., hovering over a button display pop-up hints). However, in practice the center of the title bar is relatively easy to find and does not require extraneous calculations whenever resizing the window.

The handle-bar metaphor also allows clutching: by keeping the right finger on the

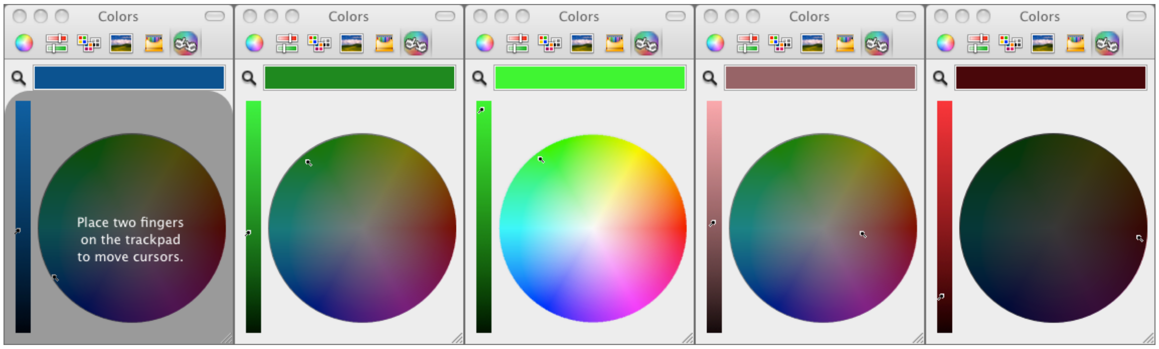


Figure 4: Frames of interaction with BiCEP: a plugin for the Mac OS X color picker

trackpad, the left cursor will stay in place as the user lifts and repositions their left finger off/on the trackpad. This is similar to clutching the mouse, where the mouse is lifted off the surface and replaced on the surface to a more ergonomic position.

Figure 4 shows different frames of interaction with BiCEP where the HSB parameters of the color wheel, brightness bar, and current color are each updated to match the locations of the BiCEP cursors. The first frame (from the left) provides instructions to the user to place two fingers on the trackpad to start interacting with BiCEP. The second frame shows the user has started their color search on a dark green. In the third frame, the user has selected a bright green. The user has performed a serial cursor interaction by deciding to only change the brightness, leaving the hue and saturation the same. In the fourth frame, the user has performed a gross color exploration by dramatically changing from their previous color (i.e., bright green) to a light shade of salmon. This was done by moving both the brightness bar cursor and the color wheel cursor in parallel. In the last frame, the user has refined their color to a dark maroon by again moving the two cursors in parallel. These interactions demonstrate that BiCEP supports the various bimanual interaction techniques outlined in Chapter 2 (e.g., symmetric, parallel, etc.)

3.2 Comparative Color User Study

To evaluate the usability of BiCEP and get feedback on the performance of dual cursor touchpad interaction, I performed a user study with two experimental sessions: 1) a repeated measures color matching experiment, comparing the speed and accuracy of BiCEP to the native single cursor color picker and 2) an open-ended coloring activity for measuring the tool as a support for creativity.

3.2.1 Participants

The study consisted of 16 participants (8 females) ranging from 18 to 24 years old with a few months to 5 years of experience using a MacBook. To reduce any effects from participants being unfamiliar with the native operating system and hardware, I recruited participants with experience using a MacBook. Other recruitment criteria included normal color vision and right handedness. Left handed users were not included because of the inherent asymmetrical interaction design layout of the controls in BiCEP. The left hand (the non-dominant for right-handed users) will control one dimension of color with the brightness bar, while the right hand will control two dimensions of color with the hue and saturation wheel. Since the right hand will do more work, the asymmetrical design [20] of the interaction is more favorable to right handed users. While changing the layout BiCEP is relatively simple, excluding left handed users reduced the number of confounds that may have been introduced.

All participants received a $10 gift card as compensation for their time.

3.2.2 Controlled Environment

Human color perception will vary based on a number of different external factors, such as surrounding colors, lighting and viewing angle [5]. It is important to control for all of these factors by maintaining a consistent experimental environment. Consistency is also a requirement to circumvent intrinsic differences in how gender [24] and language [17] influence color perception. In my experiment, I controlled environmental factors in the following ways:

- Room lighting was kept consistent by using the same experiment room with blackout curtains to block natural light and artificial lighting bright enough to see the screen clearly, but not so bright that objects were washed out.
- Screen glare was reduced by using a MacBook Pro with a matte screen, instead of a glossy screen.
- Maintaining the consistency of perceived screen colors between participants was supported by using the same color gamut and screen brightness.
- All participants used the same 15" MacBook Pro. At the beginning of the study, each participant was instructed to adjust the laptop screen to a comfortable angle and not to change the angle for the duration of the study.

3.2.3 Experiment Software

Participants completed the color matching task inside a custom application for the experiment (Figure 5), which consisted of a color selector (BiCEP or Native), two color wells, and a submit button. For the second part of the study, participants used a custom paint program designed for this study (Figure 6). This interface provided participants with access to only four common paint program features: bucket fill, eye dropper, undo fill, and redo fill.

3.2.4 Color Set

The color set used in the experiment came from the Macbeth Color Checker Chart [35]. The 24 Macbeth colors can be divided into office/nature colors, primary colors, and gray-scale colors. The colors were divided into two groups of 12 in a stratified random order. Each participant observed the same ordering of colors regardless of which color selector they used first. The original colors are provided in the XYZ color space. For my study, I used their approximate RGB values in the Adobe (1998) color gamut using an illuminate of D50. All the Macbeth colors fit into this gamut, thereby avoiding the color clipping issue discussed in Douglas et. al [16].

3.2.5 Procedure

The user study consisted of two phases: a color matching phase and a coloring phase. The experimental design of the color matching phase was based on previous color study designs [16, 21]. In the color matching phase, my participants were required to perform two sets of color-matching tasks, using BiCEP for one set and the native color selector for the other set. This color-matching task was performed within my experiment software, where participants were presented with one of two color selectors to use (Figure 5). The ordering of the dual cursor and single cursor color selection techniques was counterbalanced across participants.

Before beginning the color matching activity, I presented a short, scripted demonstration on how to use each of the color selectors. Participants were instructed to

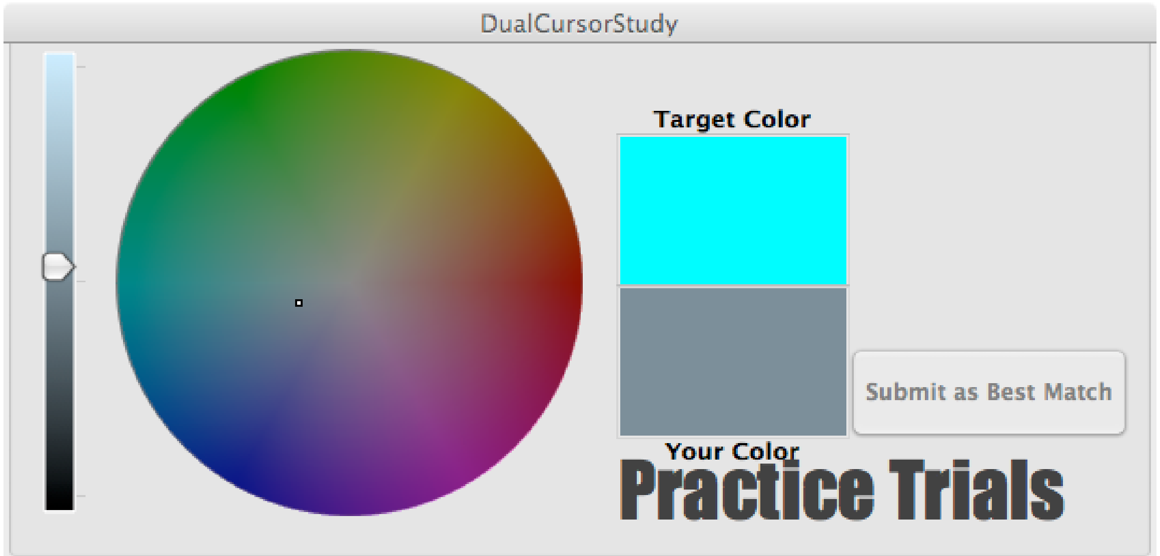


Figure 5: Matching task using the native color selection wheel and slider technique

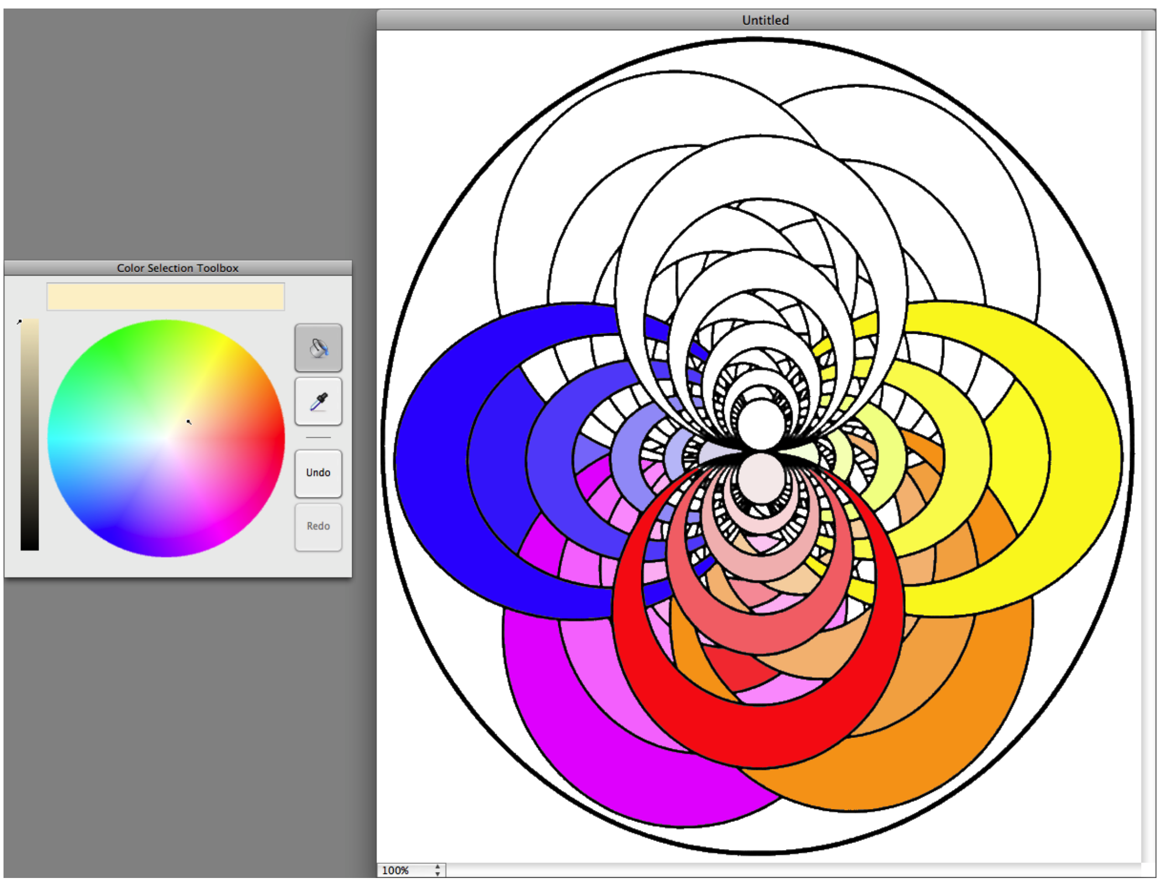


Figure 6: Experimental software for doing the coloring activity with BiCEP (on left)

closely match the color in the bottom well to the color in the top target well using the provided color selector. At the start of each condition block, participants were given 4 practice trials using colors different than those used in the actual trials. After each trial, there was a 3 second break, in which the experiment software repositioned the cursors, slider, and user color well to gray.

The second part of the study involved the use of BiCEP as a creativity support tool during a more natural creative activity. Using the custom paint program (Figure 6), participants were asked to spend about 15 minutes coloring a geometric design with colors they selected using BiCEP. The abstract geometric design was chosen so as not to lead the participant to 'typical' color selections such as a 'blue sky' and 'green grass', but rather to encourage them to openly explore colors. The fact that the fill-in cells of the design were side-by-side, also encouraged a natural color-pairing task, wherein the participant would likely try to find colors that pair in a pleasing way with colors they have already used in the design.

After completing the color phase, participants were allowed to save the image and have it optionally emailed to them. Participants then completed the Creativity Support Index [11] to rate how well BiCEP supported their creativity during the coloring activity. They also did the CSI paired comparisons of creativity factors for determining which factors were more important to the participants for this color task.

3.2.6 Analysis and Results

All but one participant completed the matching task within the time allotted. I removed that participant from my analysis, since the times for both conditions

Table 1: Descriptive statistics across the two interface types

Interface Type	Mean Time (s)	Std. Dev.	CIEDE2000 (units)	Std. Dev.
Native	36.9	24.26	4.337	2.48
BiCEP	31.9	24.62	4.339	2.53

were nearly two standard deviations from the average. The analysis involved three primary metrics: time to complete the color matches, accuracy of the color matches, and the number of colors visited by each participant. While measuring time for color matching is straightforward, measuring the accuracy of those matches is more complex. Colors were stored as RGB values, but people perceive color differences in a non-uniform manner, so using the euclidean distance between RGB values is not useful. Therefore, in order to compare the User Color to the Target Color, the RGB colors were converted into the XYZ colorspace with respect to the illuminate. Using the CIEDE2000 metric [34], the XYZ values can then be converted to the CIELAB color space, where color differences can be measured using euclidean distance. This is similar to the color comparison technique used by Douglas et. al [16].

I counted the number of colors visited per trial. I defined a visited color as a color that the participant viewed for at least 0.05 seconds and was perceptually different from other visited colors. The view time was loosely based on the 0.1 seconds reported by Miller as the ideal feedback/response time for computers [37]. I chose a slightly faster time (0.05 seconds) because the ideal feedback/response time can be contextual. I define two colors as perceptually different when their color difference is greater than the average distance between the recorded User and Target Colors for matched pairs across all participants.

For the analysis of the data, I performed a repeated measures ANOVA with between-subjects effects tests. Table 1 shows the average match times and accuracies. Participants matched colors more quickly using BiCEP (31.9 secs) than when they used Native (36.9 secs), $F_{1,343} = 5.181$, $p < 0.05$). There was no significant difference in accuracy between interfaces, interface ordering, and the interface type. The average color difference of matched colors across all participants was 4.30 units (SD = 2.50). Using this value as a filter for the color acuity of participants, I found a significant difference in the number of colors visited per trial based on Interface Type (BiCEP = 20, Native = 10, $F_{1,343} = 5.21$, $p < 0.01$).

These results prove that users were able to match colors more quickly with BiCEP, with an equal level of accuracy, while simultaneously viewing twice as many colors during the matching process. This shows that BiCEP supports both fast, targeted search, but also fluid exploration and higher coverage of the color space.

The coloring portion of the study was not comparative. For the sake of time, participants were only asked to engage in the coloring activity using the BiCEP technique. Participants completed the Creativity Support Index so that I could see how well the technique supports open-ended creative work. The CSI score for BiCEP was 77.01 (SD = 17.86) out of 100, which suggests that BiCEP supports creativity, the extend of which is made more clear when I compare BiCEP to another tool in Chapter 5. The individual categorical ratings are shown in Figure 7. Exploration was rated an average of 7.75 out of 10 and Results Worth Effort was rated an average 7.8 of out of 10. A pair-wise comparison of all 6 orthogonal factors reveals that Exploration and Results Worth Effort were rated as the more important to the task, and thus the

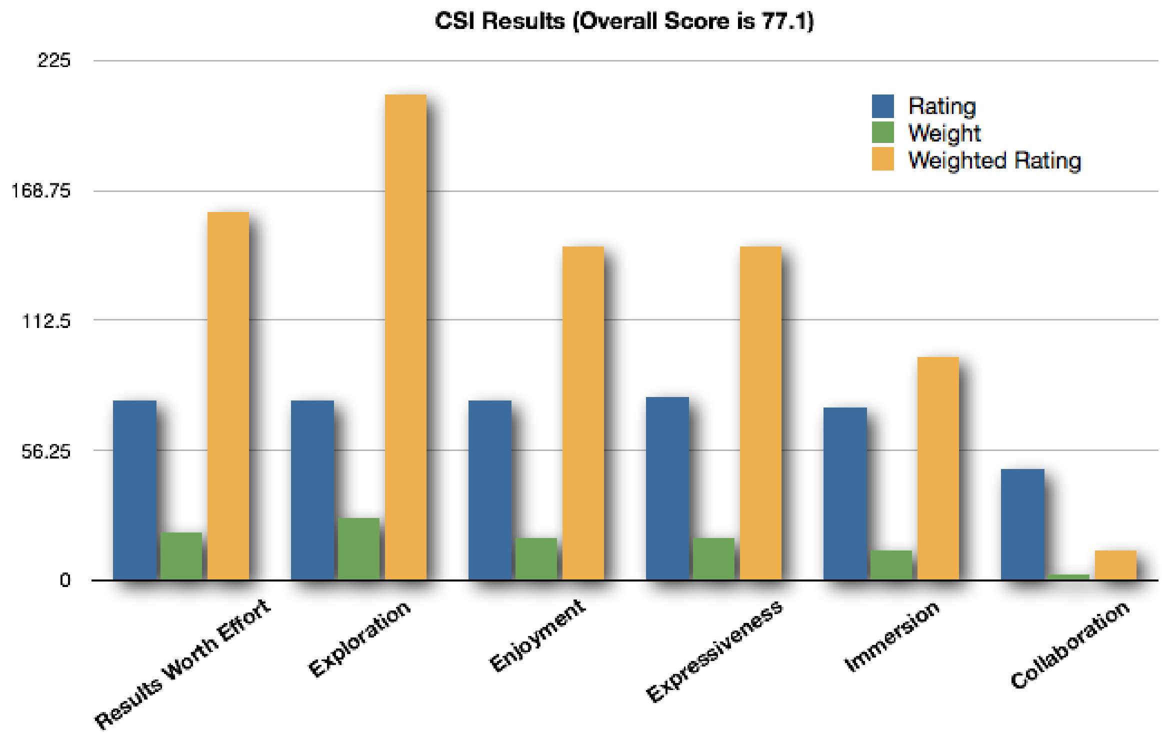


Figure 7: BiCEP's CSI score chart for the 6 orthogonal factors for evaluating creativity support tools

amount that these factors contributed to the overall ranking is high. This suggests BiCEP provides good support for explorability, and the rating for Results Worth Effort suggests that the participants felt that the effort they put into their designs are worth the outcome of their final designs.

3.3 Discussion

While my comparative study showed that BiCEP allowed faster color matching than the native color wheel interface, that measurement fails to tell the whole story. As with many novel interaction techniques, the advantages of BiCEP are not necessarily illuminated through standard measures of temporal performance and accuracy. By quantifying the exploration afforded by the interface, the benefits of the bimanual interaction technique are made explicit. Without the need to mode switch between the hue/saturation wheel and the brightness slider, the BiCEP exploration allows more colors to be explored with less effort and less interruption. This finding is not made explicit simply by measuring the time and accuracy of the color matching task.

3.4 BiCEP Summary and Future Work

With this study, I have contributed a detailed explanation of the importance of exploratory satisficing and argued that measuring explorability is a necessary first step in order to be able to develop measures of exploratory satisficing. I have shown through the BiCEP color plugin that interface design has a significant impact on explorability. I have demonstrated an overall methodology for measuring explorability. However, any particular measure of explorability needs to be developed within the specific context of the parameters being explored, because of variations in perceptual differentiation and the existence of interactional dependencies that make certain parts of any parameter space unusable. This notion is discussed in more detail in Chapter 4, Section 4.3 Explorability.

As part of my research goal, I want to find a way to quantify the level of exploratory satisficing. This notion has been implicitly measured in many HCI user studies: whenever users click a button to say they are done with a task, they are making a judgement about whether they have done enough. I aim to make the tradeoff explicit and quantify the level of satisficing as a ratio that reflects the parameter space.

CHAPTER 4: EXPLORABILITY, SATISFICING, AND SATISFACTION

4.1 Thesis Statement

Interaction techniques and size of search spaces (defined by the granularity of control) for the exploration of large parameter spaces can positively and negatively effect the amount of options users explore as well as the users' satisfaction with their creative process and final designs. Some of these effects impact Maximizers differently than Satisficers.

My goal is to better understand the relationships that exist between the user's actual and perceived exploration of the parameter space, and the satisfaction the user has with their design process and final result. Based on the literature I presented in Chapter 2, I also believe that characteristics of the user's personality (i.e., satisficer versus maximizer) can impact whether that user is more likely to satisfice early or be less satisfied with more exploration, creative process, or final designs.

The specific outcomes and results from my research into explorability, satisficing, and satisfaction are described in Chapters 5 and 6. To reduce repetitiveness, in this chapter, I establish a common knowledge base for understanding how I chose to study explorability, satisficing, and satisfaction. First, I introduce my research questions. Then, I describe a particular set of core questions, constructs, and mythologies that I utilized in my study of explorability, satisficing, and satisfaction. These topics have been marginally generalized for broader use outside the context of my dissertation, making this chapter a useful resource for other researchers investigating the level of exploration supported by their interfaces for navigating large parameter spaces.

4.2 Research Questions

Here are the research questions my dissertation aims to answer:

R1: How does changing the interaction technique used to navigate a parameter space impact people's satisfaction with their design process and final result?

R2: How does changing the search space of a set of parameters impact people's satisfaction with their design process and final result?

R3: How do Maximizers and Satisficers differ when is comes to their satisfaction with the design process and final results?

R4: How does the actual percentages of the parameter space explored impact the satisfaction users have with their process and final result?

R5: How does the perceived percentage of the parameter space explored impact the satisfaction users have with their process and final result?

R6: What are the differences in how parameter space interaction techniques and search spaces support creativity?

4.3 Explorability

Compared to satisficers, maximizers have a higher threshold of what they consider to be good enough and will often seek out as many new (and possibly better) options and alternatives as possible. One factor that may contribute to how much of the space they explore is how much they have (or have not) already explored. Knowing how much of the space was explored can be beneficial or detrimental to their process. If the quantity of options already explored is a key factor in their decision making process, maximizers may take less time to make a choice having explored a certain quantity of choices (often because the quantity was more than usual or otherwise available). If the quality of the option is a key factor in their decision making process, maximizers may take more time to explore new options since they are covering more ground in a shorter amount of time.

4.3.1 Quantifying Explorability

To quantify explorability, we need to understand the ratio of how much was explored in relation to the entire parameter space that can be explored. Thus, we need to identify the numerator (the number of parameter combinations visited) and the denominator (the number of parameter combinations available). However, there are two issues with producing these numbers.

The first issue is that not every combination of parameter choices will be perceptually different. Making a minor adjustment in any given parameter may not lead to a change that a user is able to detect. Figuring out the number of perceptibly different options in a multi-dimensional parameter space is tricky. It is doable in a well-understood domain such as color spaces, but in other areas, we may not know how much change in a given parameter leads to a perceptibly different outcome as the user changes other parameters. While it is possible to determine this by running psychometric studies, we may not want to do that level of anaylsis for every possible type of parameter space.

The second issue in quantifying explorability is that some parameter options have interactional dependencies that can lead to pointless exploration of some parts of a parameter space. For example, when using the BiCEP or Native color picker, moving the brightness slider to the bottom of the brightness bar sets all brightness values in the hue/saturation color wheel to black. While the color black may be useful, exploring the color wheel at this brightness value is a waste of time, since all the colors look perceptually the same (i.e., black). Another example is the font dialog box in most word processors: setting the foreground color of the font to be the same as the background color is an allowable setting, but doing this makes all other options (typeface, bold, italics, type size, etc.) completely irrelevant. Thus, it would not make sense to count all of those options as ones which need to be explored. This shows that calculating the denominator, the number of parameter combinations that are relevant and valid, is not a straightforward mathematical calculation.

One solution to address both issues is to have the user save milestones of the design process (e.g., interesting images, unique sounds, creative designs). This technique was used by Noy et al. for tracking and comparing images created during the design process [41]. We can use this data to evaluate a sub-section of the larger parameter space. This 'sub-space' can then be used to make generalized conclusions about the entire parameter space of possibilities.

4.3.2 Self-Reported Explorability

Self-reported measures of ‘exploration’ do not paint the entire picture of how a participant may mentally perceive their own exploration of a parameter space, nor do they describe how the mental perception of exploration can impact a participant’s decision making process. These topics go beyond the scope this dissertation. My goal with the self-reported measure of exploration is to get insights into the mental processes, using the self-report as a comparative metric between conditions and in conjunction with the other measurements of explorability.

To measure how much of a parameter space the user self-reported themselves as “having explored,” I developed the following question that is answered using a scale of 0% - 100%.

Exploration: E1: “Given the huge number of possible [choices], it is impossible to explore all the [combinations]. What percentage of the [options] do you think you considered while using the provided [application]?”

The words in brackets [ ] can be adjusted to match the context of the parameter space explored, see examples in Chapters 5 and 6.

4.3.3 Supporting Quality Explorability

Thus far we have discussed measuring explorability, but it is also important to understand the factors that can contribute to the quality of the exploration (or design process). The quality of explorability in an interface depends on the interaction technique used for traversing the parameter space and the size of the search space available to the user. These two factors (i.e., interaction technique and search space size) serve as my independent variables in Chapters 5 and 6, respectively.

Switching between individual sliders allows individual control of parameters, but requires frequent switches between the sliders to explore the combined parameter space. Deciding how many values a slider represents can also impact the user experience. Sliders will often represent either a relatively few number of options or hundreds of options. If you are using a slider to control the music volume, the feedback rate is real-time so you will ideally not spend much time using the slider. On the other hand, in the case of some photoshop filters, each parameter's slider value can have a drastically different visual effect that may take a few seconds to render at each step along the slider. Other aspects of the interface can provide support to enhance the quality of explorability: history capture and browsing interfaces can help users investigate and reflect upon their past explorations. Unlimited undo/redo stacks can reduce the costs associated with traversals of a parameter space.

4.4 Satisficing

I focus on two elements related to satisficing. The first is the act of satisficing, where someone determines that their work is 'good enough,' with 'good' relating to the level of satisfaction with the final product and 'enough' relating to how much the product meets a participant's threshold of acceptability. The second element has to do with the threshold of acceptability. Some personalities may be inclined to perfectionism or high achievement, while others may be more easy-going. Both of these elements are covered in the next two sections.

4.4.1 Experimental Design

In order to measure Satisficing, the experiment needs to have an observable event that can indicate that a participant has decided their work is *good enough.* There are two main components to my approach for measuring satisficing: 1) provide a context in which the participant has constraints that elicit satisficing and 2) ask the participant about when and why they decided their work was 'good enough.' See Chapter 6 for an example of an experiment with both components.

Time is often a principal constraint that influences when and why a participant chooses to satisfice. Some participants may decide their design is 'good enough' and submit their design before their time is up. The time-stamp metric is not a direct measure of satisficing behavior, as it suffers from construct validity. In my work, I have made attempts to encourage participants to invest in the design task. Even so, I cannot completely rule out the possibility that submitting early is actually the participant deciding they just want to finish the experiment as fast as possible. However, when participants use all of the trial-time, I have an indication that they were not finished exploring the design space, possibly because they felt their design had not met their threshold of acceptability (see Section 4.5 Satisfaction). This time-measurement, when accounting for external factors (e.g., block ordering, task repetition, etc.), can indicate the amount of satisficing that has occurred when comparing two applications.

4.4.2 Maximization Scale

Schwartz's Maximization Scale [46, 40] is used to determine where on the spectrum of satisficers versus maximizers a participant's personality is located. Participants rate themselves on a 7-point Likert scale for a series of 6 statements indicating how much they agree or disagree with each (1 – completely disagree, 7 – completely agree). These statements measure maximization along three different dimensions: alternative search, decision difficulty, and high standards. I have made small modifications to some of these statement in order to modernize them (e.g., the original statement used the wording "renting a video"). The statements are as follows:

Alternative Search:

MS1: When I am at home or in the car listening to the radio, I often check other stations to see if something better is playing, even if I am relatively satisfied with what I am listening to.

MS2: No matter how satisfied l am with my job, it is only right for me to be on the lookout for better opportunities.

Decision Difficulty:

MS3: I often find it difficult to shop for a gift for a friend.

MS4: Choosing a video to rent/stream is really difficult, I am always struggling to pick the best one.

High Standards:

MS5: No matter what I do, I have the highest standards for myself.

MS6: I never settle for second best.

The maximization score is the average score for all questions. These scores help categorize users and serve as a factor to understanding the variances in satisfaction and exploration that could be the result of personality traits. The categorization method I used is as follows:

Two-Tier Categorization: Using the maximization scale, participants scoring higher than 4 are categorized as maximizers and those scoring less than 4 are categorized as satisficers. From the thousands of participants recruited, Schwartz found that 80% of users fell between 2.5 and 5.5 [46]; extreme maximizers (5.5) and extreme satisficers (2.5) were less common (10% at each extreme).

4.4.2.1 Contextualized Maximization Scale

I strongly believe that our inclination to maximize or satisfice is highly contextual. Depending on the context we may act as Maximizers or Satisficers. This may especially be the case for those who fall in the very middle of the spectrum. For example, someone may satisfice when it comes to their choice of entree, but maximize when selecting their beer or wine pairing.

To this effect, I developed the Contextualized Maximization Scale (CMS), which is a 3-item scale that shortens and contextualizes the Maximization scale. This 3-item scale was developed using the structures and categories in creating the more general Maximization scale, as well as the scale-shortening methodology described in the work of Nenkov et al. [40]. The three items are rated on a 7-point Likert scale of agreement. The following is an example of the CMS set in the context of color selection, this example is also used in the studies done in Chapters 5 and 6.

Alternative Search: CMS1: When selecting a color to use in a computer application, I will often try out different colors from across the spectrum.

Decision Difficulty: CMS2: Choosing the best color for my work can often be difficult.

High Standards: CMS3: When selecting a color, I often imagine what other colors may be a better choice.

4.5 Satisfaction

In order to measure a participant's level of satisfaction with the final design(s) created using an application, I generated a series of 7-point Likert scale statements of agreement/disagreement based off those used by Oulasvirta et al. [42], who found Maximizers and Satisficers had significantly different ratings for satisfaction and confidence. The statements are as follows:

Satisfaction: S1: "I am satisfied with my final design."

Suitability: S2: "I think my final design is suitable [to the task requirements]."

Carefulness: S3: "I made my final design carefully."

Satisfaction: S4: "I am satisfied with my design process with the provided [software]."

Immenseness: S5: "I felt overwhelmed by the [choices] offered by the [software]."

The words in brackets [ ] can be adjusted to match the context of the study, see the separate examples in Chapters 5 and 6.

The rating of the satisfaction claim is indicative of where their final design fell in accordance with the participant's threshold of acceptability. The rating of suitability allows me to tease out differences in satisfaction with the participant's final creation versus their satisfaction with how their design fit the task prompt (if applicable). The rating of the carefulness claim is indicative of the participant's level of engagement, or interest in the task [44]. I have added an additional statement type: Immenseness. The rating of the immenseness claim is indicative of how overwhelmed the user may have felt by the number of choices made available through the interface.

4.6 Creativity Support

Carroll et al. [11] specifically include 'exploration' as a factor that contributes to supporting creativity, while 'satisfaction with the design process' is included indirectly through other creativity factors mentioned in their work. While my work does not directly contribute to creativity research, I did not want to ignore these relationships that may have bearing on the design process of users.

4.6.1 Creativity Support Index

In part one of the Creativity Support Index (CSI) [11], participants respond to two sets of agreement statements related to each of the following factors: Results Worth Effort, Exploration, Collaboration, Immersion, Expressiveness, and Enjoyment. Participants answers these questions after every interface they use during the study. In part 2, participants do paired-rankings of the factors above. While not all interfaces will support Collaboration, the most recent version of the CSI has a 'not applicable' mechanism to handle these cases. Three of the most relevant factors include:

Results Worth Effort: What I was able to produce was worth the effort I had to exert to produce it.

Exploration: The system was helpful in allowing me to track different ideas, outcomes, or possibilities.

Expressiveness: I was able to be very creative while doing the activity inside this system or tool.

Results Worth Effort gets at the question of satisficing when it comes to the amount of work required to achieve a result. Exploration measures explorability at a macro level. While my work focuses on the explorability of the parameter space, the participant's evaluation of the overall exploration that is supported by the interface is also important. Expressiveness relates to the satisfaction users will have with their design process. I used the CSI to compliment the measure I have developed.

4.7 Approaches to Parameter Space Explorability

There are two approaches I considered when thinking about easing the exploration of parameter spaces. The first approach was changing the Interaction Space. This is where the physical interaction, be it a device or interaction technique, changes. The second approach was changing the Search Space, or changing from fine-grained to course-grained control of UI widgets used for controlling parameters. Interaction Space and Search Space will be covered in Chapters 5 and 6, respectively.

4.8 Summary

In this chapter, I have outlined some of the constructs and methodologies that I used for studying explorability, satisficing, and satisfaction. While this chapter is by no means a catch-all framework for measuring these topics, it does lay a foundation for my research. In Chapter 5, I incorporate the metrics outlined in this chapter in order to investigate how changing the interaction technique impacts parameter space explorability and satisfaction. In Chapter 6, I take the framework outlined in this chapter and refine them based on the findings from Chapter 5. I then used the refined framework to investigate how changing the search space impacts parameter space explorability and satisfaction.

CHAPTER 5: EXPLORABILITY AND SATISFICING IN DIFFERENT INTERACTION SPACES

In this chapter, I describe a study measuring how changes in the interaction space can impact explorability and satisficing. I conducted this study as a follow up to the BiCEP activity described in Chapter 3. More specifically, I measured differences in creativity support, explorability, satisficing, and satisfaction between BiCEP and the native Mac OS X color picker (i.e., color wheel and brightness bar).

5.1 Motivation

The color-matching task in Chapter 3 is a controlled task with a correct answer. In the following study, participants took part in a coloring activity, an open-ended, creative task, where these is no 'correct' answer. While I had participants do a coloring task as part of the study in Chapter 3, they only did so with BiCEP. Therefore, I was not able to make comparisons between the BiCEP color picker and the Native color picker for the creative activity. The following study expanded my understanding of how changing the interaction technique can impact explorability and satisficing.

5.2 Experimental Design

The experiment was a repeated measures design where participants engaged in two trials of the same coloring activity, one with BiCEP and one with a Mac OS X native color picker. In this study, I collected data on creativity support, satisfaction, completion time, level of exploration, and interface preference.

I recruited 27 participants (19 females) ranging between the ages of 18 to 44 years old[1] with a few months to over 5 years of experience using a MacBook. These were a completely different set of participants from those in Chapter 3. They also ranged from inexperienced to advanced users of graphic design software. Other recruitment criteria included normal color vision and right handedness (see rational in Chapter 3 Section 3.2.1 Participants).

All study participants received a $10 gift card as compensation for their time. Three participants received a bonus $10 gift card based on votes from 100 Amazon Mechanical Turk users who were the judges of the final designs. Each Mechanical Turker was compensated $0.10 for their time spent voting (less than 2 minutes).

5.2.1 Questionnaires and Surveys

The demographic questionnaire at the start of the study gathered background information (e.g., age, sex, ethnicity), years of experience with MacBooks, and familiarity with image editing tools (e.g., Instagram, Photoshop, etc.). The follow-up questionnaire at the end of the study asks the participant about which color plugin (BiCEP or Native) the participants preferred and why. In addition to these short questionnaires, I used three different surveys from the literature to measure the following: whether a participant's personality inclines toward satisficing or maximizing, the amount of creativity support offered by the BiCEP and Native color-pickers, and the satisfaction of the users with their final design and design process. Additionally, I used a

[1]The exact age of each participant was not collected. Instead, participants selected from a series of age ranges (e.g., 18 - 24).

time-measurement approach for determining the amount of satisficing that occurred while using the BiCEP and Native color pickers.

In addition to the surveys outlined below, for this study, I used the Maximization Scale (see Chapter 4 Section 4.4.2 Maximization Scale) and the Creativity Support Index (see Chapter 4 Section 4.6.1 Creativity Support Index).

5.2.1.1 Satisfaction

At the end of each condition block (i.e., after using the BiCEP or Native color picker), participants rated the following claims on a 7-point Likert scale:

Satisfaction: S1: "I am satisfied with my final design."

Carefulness: S3: "I made my final design carefully."

Satisfaction: S4: "I am satisfied with my design process with the provided color picker."

Immenseness: S5: "I felt overwhelmed by the choices offered by the color picker."

See Chapter 4 Section 4.5 Satisfaction for more details. Note that the Suitability claim (S2) was excluded from this study as the task prompt was very open, allowing participants to color the image any way they could and wanted to. Suitability would be relevant if there was a specific prompt that would put a constraint on the design space of possibilities. An experimental prompt is often used to put all participants in a similar mindset going in. In Chapter 6, I use a prompt, asking participants to make a DVD/Album Cover. By measuring Suitability, I can account for differences whenever a participant was unsatisfied with their final design, but still thought the design fit the prompt, and visa versa. These could be cases where the prompt may have inhibited creativity. However, in practice, I believe that constraints can encourage creative thinking and can also lead to the discovery of ways to circumvent the constraint.

5.2.1.2 Explorability

At the end of each condition block (i.e., after using the BiCEP or Native color picker), participants answered the following question using a scale of 0%-100%.

Exploration: E1: "Given the huge number of possible colors, it is impossible to explore all the color combinations. What percentage of the color space do you think you considered while using the provided color picker?"

See Chapter 4 Section 4.3.2 Self-Reported Explorability for more details.

5.2.1.3 Contextualized Maximization Scale

For this study, I used the Contextualized Maximization Scale described in Section 4.4.2.1. The scale was administered at the end of the study, rather than the beginning like the general Maximization Scale. This was done to avoid priming participants more than necessary about what I was measuring.

5.2.2 Procedure

At the start of the user study, I provided each participant with a scripted overview of what participation in the study would entail. This brief overview allowed me to inform the participants that their final designs would be submitted to a panel of judges for review. I also explained how the coloring interface and time constraints worked. After the introduction, each participant was asked to sit at a table with a 15"

MacBook Pro laptop with a matte screen to reduce screen glare. All tasks and surveys were completed on this laptop and the same laptop was used by all participants.

Once seated, the experimenter presented a scripted demonstration on how to use each of the color pickers (Native and then BiCEP). After describing the native color picker, the researcher asked each participant to adjust the color wheel and brightness bar using the system cursor. After describing BiCEP, the experimenter asked participants to adjust the color wheel and brightness bar using the dual-cursors and then use the system cursor to close the window. This ensured exposure to both color pickers and provided time to ask questions. All participants were able to get through this part without asking questions.

The rest of the study was completely automated and started with participants filling out a demographic survey and the Maximization Scale. The next portion of the study consisted of two counterbalanced condition blocks, one for each color selector. Each block consisted of three phases: a very short color matching phase, a coloring phase, and a post-survey phase. In phase 1, participants completed 4 color-matching tasks using the same color-matching interface described in Chapter 3 Section 3.2 Comparative Color User Study (see Figure 5). This gave the participants a chance to practice using each interface and provided our team with the data needed for estimating the users' perception of color difference, a parameter used to measure the number of unique (to the user) colors visited.

In phase 2, participants used a coloring interface similar to the coloring interface described in Chapter 3 Section 3.2 Comparative Color User Study. The key differences being the images used and support for both the BiCEP and the Native color picker (see Figure 8). The custom program has only four common paint program options: bucket fill, eye dropper, undo, and redo. Participants were asked to spend 10 minutes coloring in an image of an animal with a geometric backdrop using colors they selected with the provided color picker (Figure 8). After 10 minutes, a submit button became available for them to submit their final design. At the start of the study, participants were informed that they could take an additional 5 minutes to complete their designs, but at 15 minutes the application would close automatically. They were informed that their images were still saved and submitted even if they ran out of time.

The software logged how much longer they took to complete their coloring task after the submit button was available. I used this measurement as a way of understanding the participants' level of satisficing. Before each coloring activity phase, the participants were reminded that their final design would be submitted to a team of judges to vote on and the top three designs would receive a bonus gift card. This message was also displayed in the confirmation dialogue that appears when participants pressed the submit button (before time expired). This reminder was used to incentivize participants to fully engage in the task and reduce the effect of confounding variables that can influence a participant's decision to submit their designs early.

In phase 3, participants completed a post-block questionnaire consisting of: the Likert scale statements from the Creativity Support Index [11], the satisfaction statements, and the explorability question.

Participants then repeated phases 1-3 under the alternative condition (i.e., other color picker). Afterward, they completed the paired rankings of the CSI and filled out a short questionnaire about their preference of the color pickers used in this study.

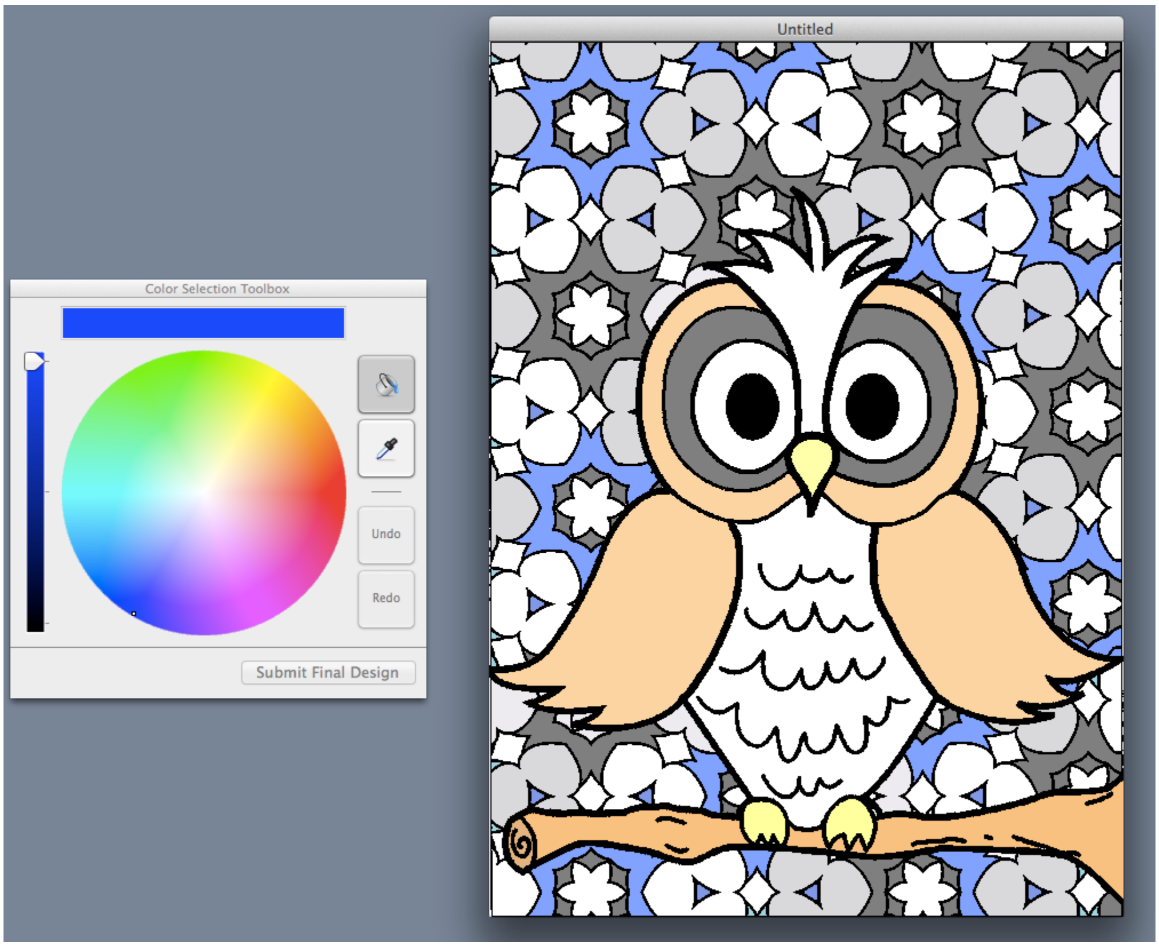


Figure 8: Experimental coloring software

I collected the final 54 coloring designs from all 27 participants and posted them on Amazon's Mechanical Turk. The Turkers were asked to select the 3 images they thought were the "coolest" from each of the two types of images (owl and monkey). By having them vote within the image types, I removed any bias they may have had towards one of the animals. The ordering of the images on the web page was randomized for every Turker. They did not have access to any identifiable information about the study participants and no other information was collected from the Turkers.

5.2.3 Hypotheses

This section contains my general hypotheses that guided the design and analysis of this study. In the analysis and results section, I will use the data from this study to test each hypothesis. Note that for H_3-H_5, a high Maximization Scale Score indicates the participant is a maximizer (more than 4.0) and a low score indicated the participant is a satisficer (less than 4.0). Though participants can be put into two categories (i.e., satisficer, maximizer), H_3-H_5 use the raw Maximization Scale score as a continuous variable in correlation tests.

H_1: Participants will explore a wider distribution of colors with BiCEP than with the Native color picker.

H_2: Participants will be more satisfied with how adequately they explored the color space using BiCEP over the Native color picker

H_3: Participant's Maximization Scale Scores will inversely correlate with their satisfaction with exploring the parameter space using BiCEP

H_4: Participant's Maximization Scale Scores will correlate with the amount of allotted time they use

H_5: Participant's Maximization Scale Scores will inversely correlate with the perceived percentage of color space explored

H_6: Participants will explore more unique colors with BiCEP than the Native color picker

H_7: Participants will be more satisfied with their results from BiCEP versus the Native color picker

H_8: Participants with higher exploration (percentage of color space covered) will have a higher user satisfaction with the design process

H_9: Participants with higher exploration (percentage of color space covered) will have a higher user satisfaction with the final product

H_{10}: Participants will perceive having explored a higher percentage of the design space with BiCEP than with the Native color picker

H_{11}: The CSI factor Expressiveness will be higher in BiCEP than the Native color picker

H_{12}: The CSI factor Exploration will be higher in BiCEP than the Native color picker

H_{13}: The CSI factor Results Worth Effort will be higher in the BiCEP color picker than the Native color picker

H_{14}: Participants will prefer to use BiCEP over the Native color picker

The following are additional hypotheses that are specific to participants classified as either Maximizers or Satisficers:

$H_{M.1}$: For Maximizers, Maximization scores will correlate with BiCEP satisfaction

$H_{M.2}$: For Maximizers, Maximization scores will inversely correlate with satisfaction with the Native color picker

$H_{M.3}$: Maximizers will use all the allotted time for their trials.

$H_{S.1}$: Satisficers will spend more time with BiCEP versus the Native color picker.

$H_{S.2}$: Satisficers' satisfaction will be higher overall compared to Maximizers.

5.3 Analysis and Results

I analyzed my data using the Restricted Maximum Likelihood (REML) method of Standard Least Mean Squares (SLS), which is similar to a Repeated Measures Analysis of Variance (ANOVA). REML-SLS is a mixed-model approach that better accounts for the variance and random effects intrinsic to my unbalanced data[2]. Reported averages were calculated using Least Mean Squares. Reported averages also include standard errors (SE), which takes into account the Standard Deviation and the sample size. When applicable, the change or numerical difference between conditions is reported using the standard delta notation (Δ). Effect sizes for correlations

[2]Certain pairs of data were not equally balanced (e.g., Maximizers versus Satisficers, BiCEP versus Native Preference, etc.). While some conditions had enough data points from which to drawn conclusions, others did not, in which case I provide an explanation.

were analyzed using Cohen's recommendations: R>0.2 (small effect), R>0.5 (medium effect), and R>0.8 (large effect) [13].

5.3.1 TouchPad Confounds

There were several participants who commented on the trackpad interaction being difficult to control or uncomfortable. All participants were MacBook users and therefore were accustomed to the nuances of using a trackpad, but some of them may have been more experienced trackpad users (e.g., three-finger drag, gestures, etc). A majority, but not all, of these participants stated they preferred Native over BiCEP. In this study, I changed the way participants used the trackpad, which seems to have had some bearing on the results. These participants are still included in the results, but have been labelled as having issues with the system. When it had bearing on the results, I indicate when there was a significant difference between those that had "trackpad issues" and those did not. The preference section towards the end of this chapter, covers this issue in more detail.

5.3.2 Actual Exploration

The percentage of the parameter spaced explored (actual exploration) is calculated as a ratio of the parameters explored to the size of the parameter space. Since I am dealing with color, I wanted this ratio to reflect that some of the colors are perceptually equivalent. Therefore, I developed and used the formula below, where n_{colors} is the number of unique and perceptually different colors that the participant visited, 255^3 is the size of the HSB parameter space in my interface, where each parameter (i.e., hue, saturation, brightness) could be set to a value between 1 and

255, and Δcp is the participant's threshold for perceiving coloring differences (derived from matching task).

$$Exploration \approx \frac{n_{colors}}{255^3} \times \Delta cp \qquad (1)$$

The inclusion of the color perception threshold is important as this factor impacts the size of the parameter space for each participant. The thresholds varied between participants, so while this equation represents an approximation of the total exploration that occurred. Even as an approximation, I believe the measurement can still be used as a comparative tool for evaluating the exploration done between BiCEP and the Native color picker.

5.3.3 Satisfaction

The difference in reported satisfaction with participants' final designs (SatFD) was statistically significant and was higher when participants used BiCEP (8.14 SE 0.57) than when they used the Native color picker (6.95 SE 0.57). Therefore, for H_7, I can reject the null hypothesis in favor of the alternative: H_7 : *Participants were more satisfied with their results from BiCEP versus the Native color picker.*

There was also an interaction effect between the interface used and whether participants had issues with the trackpad. A t-test revealed this effect was only statistically significant ($p<0.01$) when participants had trackpad issues and used the Native color picker. These participants reported a mean SatFD of 5.41 (SE 0.99), which was lower than the participants who used Native and did not have trackpad issues and the participants who used BiCEP. This is interesting because many of these users preferred the Native color picker to BiCEP. This could indicate that SatFD may not be closely related to interface preference.

The difference in reported satisfaction with participants' creative process (SatCP) was not statistically significant between interfaces. There were no interaction effects from the other measurements. Therefore, for H_8, I cannot reject the null hypothesis: $H_{8,null}$: *Participants with higher exploration (percentage of color space covered)* ***did not*** *have a higher user satisfaction with the design process.* Also, for H_2, I cannot reject the null hypothesis: H_2 : *Participants* ***were not*** *more satisfied with how adequately they explored the color space using BiCEP over the Native color picker.*

5.3.4 Exploration

The difference between the number of unique (and perceptually different) colors explored was statistically significant ($p<0.01$) and was higher when participants used BiCEP versus when they used the Native color picker (see Table 2). There were no other interaction effects. To test the reliability, I also calculated the number of visited colors per minute and found the difference in the number of visited colors per minute was statistically significant ($p<0.01$) and was also higher when participants used BiCEP versus when they used the Native color picker. Since color can be perceived differently for select users, I used my formula for actual exploration to compare the the interface and found that the percentage of actual exploration was higher ($p<0.01$) in BiCEP than the Native color picker (Δ 1.7×10^{-3}% SE 4.1×10^{-4}). Therefore, for H_6, I can reject the null hypothesis in favor of the alternative: H_6 : *Participants explored more unique colors with BiCEP than the Native color picker.*

Table 2: Descriptive statistics of the visited colors and colors/minute for each interface

	Visited Colors		Colors/Minute	
Interface	Mean	SE	Mean	SE
BiCEP	**59.60**	6.21	**4.61**	0.45
Native	**39.23**	6.21	**3.16**	0.45
				p<0.01

Actual exploration and reported satisfaction with participants' final designs (SatFD) were not correlated (R<0.1). There were no interaction effects from the other measurements. Therefore, for H_9, I cannot reject the null hypothesis: $H_{9,null}$: *Participants with higher exploration (percentage of color space covered)* ***did not*** *have a higher user satisfaction with the final product.*

The difference in perceived exploration was not statistically significant between interfaces. There was an interface ordering effect. A t-test revealed that participants reported themselves as having explored more using BiCEP versus the Native color picker, when they used Native first and then BiCEP (Δ = 17.62 SE 5.35, p<0.01). The t-test also revealed that participants rated themselves as having explored more using BiCEP in Block 2 than in Block 1 (Δ = 25.38 SE 11.01). There was no significant difference between the two color pickers when participants used BiCEP first and then Native. There were no other interaction effects.

This suggests that when participants use Native then BiCEP, the difference in their exploration may have been more apparent. Participants are then able to better compare the differences in exploration and report a higher exploration in BiCEP. Therefore, for H_{10}, I can reject the null hypothesis in favor of a conditional alternative: H_{10} : ***When participants used the Native color picker first,*** *they perceived*

Table 3: Reported perceived exploration between interfaces and blocks **p<0.05**

Interface	Block	Mean (%)	Std. Error
Native	1	50.15%	7.79
	2	45.46%	7.79
BiCEP	1	**42.38%**	7.79
	2	**67.77%**	7.79

having explored a higher percentage of the design space with BiCEP than with the Native color picker.

Perceived exploration correlated with the number of unique colors visited ($R \approx 0.28$) and with the number of unique colors visited per minute ($R \approx 0.30$). Using Cohen's guidelines, since $0.2 < R < 0.5$, the effect sizes are considered small. This suggests that the participants' perceptions of their exploration were slightly in line with some of my measures of their actual exploration. However, their perceived exploration percentage did not correlate with their actual percentage. Actual percentage takes into account their color acuity, which participants are likely not taking into consideration when self-reporting their perceived exploration.

In order to compare distributions of explored colors between the BiCEP and Native color pickers, I placed each of the HSB values into a $6 \times 6 \times 6$ matrix of bins. This allowed me to compare the differences in the number of 'bins' of the parameter space that were explored. The 6-bin method was decided after trying different sizes of bins from 5-10. Part of the goal of using the binning method was to make the analysis as cognitively manageable as possible. The 6-bin trial showed promise by not having so many bins that the data was spread thin, but at the same time allowed some patterns to be clearly visible.

BiCEP covered 16 bins (9%) that Native did not cover. Native covered 7 bins (4%) that BiCEP did not cover. See Heat Maps in Figures 9, 10, and 11. I did not count the bins where brightness was 0, as these colors are all black and indistinguishable from one another. One noticeable pattern is that when BiCEP covered areas the Native did not cover, the coverage would often include neighboring bins, as opposed to a single bin here and there. This could indicate that participants were searching around particular colors. Based on this analysis, for H_1, I can reject the null hypothesis in favor of the alternative: H_1 : *Participants explored a wider distribution of colors with BiCEP than with the Native color picker.*

5.3.5 CSI Scores

While the mean score for BiCEP was higher than the Native color picker, the scores for the CSI were not significantly different between interfaces, Native: 68.30 (SE 3.11) and BiCEP: 72.90 (SE 3.11). As a reminder, a higher CSI score indicates better creativity support. A t-test revealed a statistically significant interaction effect between the interface and block condition, where for the participants that used BiCEP in Block 1, BiCEP was rated higher than the Native color picker in Block 2. There were no other interaction effects.

The rest of the analysis of the CSI Scores uses standard mean calculations. All the other analyses in this chapter use least mean square calculations, but for this part of the analysis, I have chosen to follow the procedure described in [11]. Any minute variations between the reported means are not reporting errors but artifacts of the different mean calculations.

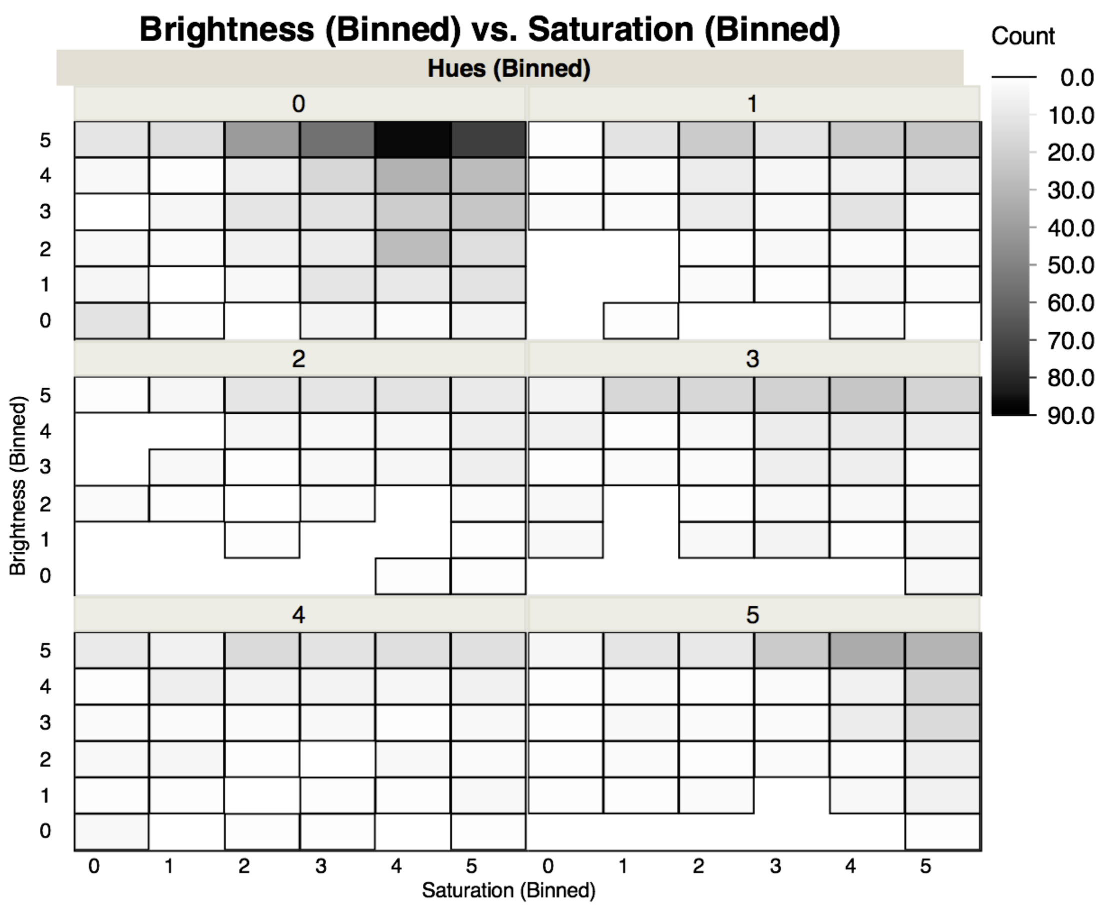


Figure 9: Heatmap of binned HSB values for Native, white squares contain no data

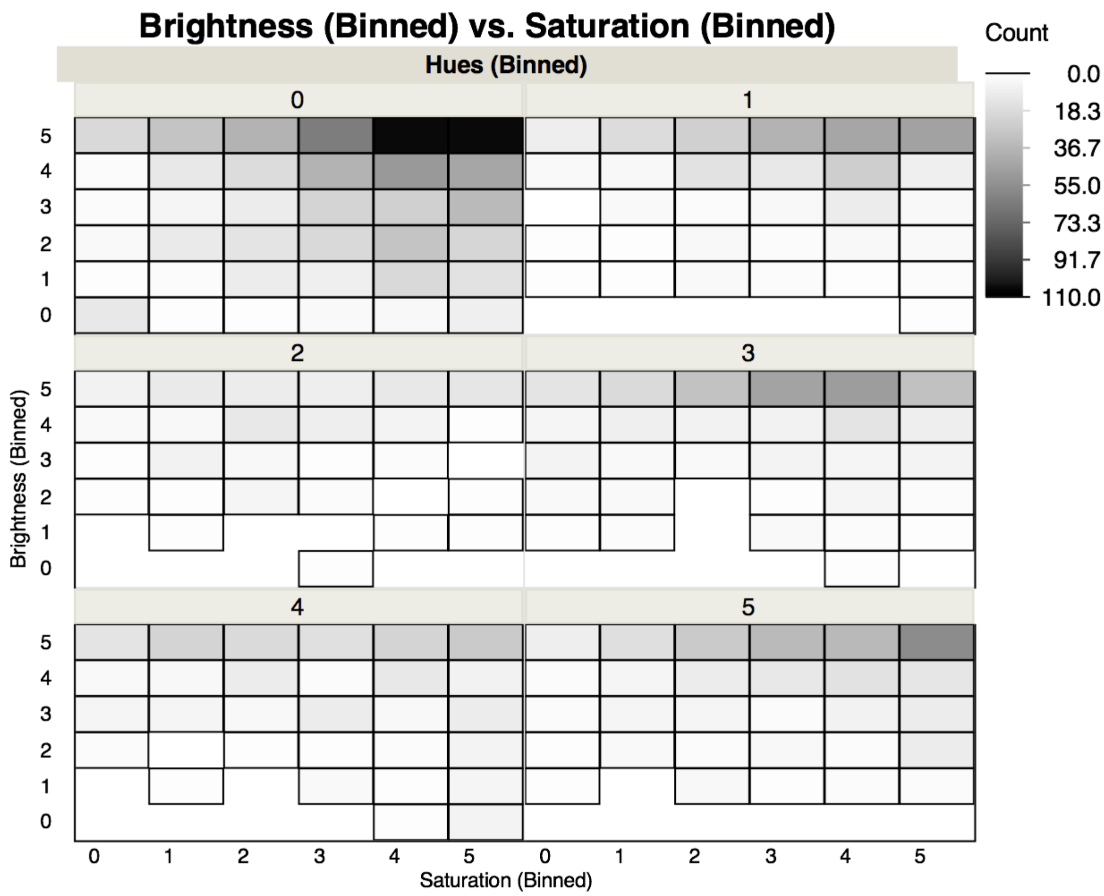


Figure 10: Heatmap of binned HSB values for BiCEP, white squares contain no data

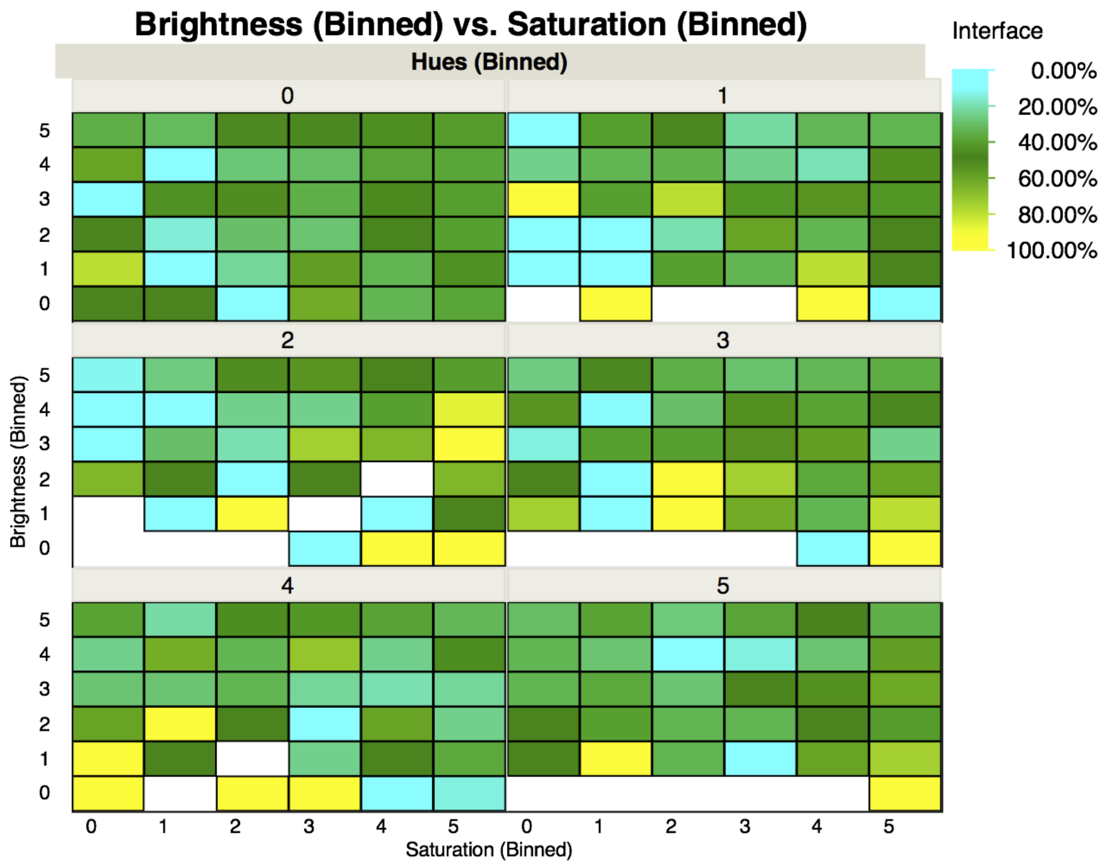


Figure 11: Heatmap of HSB values for BiCEP and Native. Yellow indicates 100% of the values in the bin are from Native. Cyan indicates 0% of the values in the bin are from Native (i.e., they are all from BiCEP). Shades of green indicate a combination of values from Native and BiCEP. White squares contain no data from either interface.

The overall CSI score for BiCEP was 73.03 (SD = 15.31) out of 100 and the overall CSI score for the Native color picker was 68.17 (SD = 17.20). The individual categorical ratings are shown in Figures 12 and 13. A pair-wise comparison of all 6 orthogonal factors reveals that Exploration and Expressiveness were rated as more important to the task, and thus the amount that these two factors contributed to the overall ranking is high. This suggests the Native color picker and BiCEP generally provide good support for explorability and expression.

Furthermore, the differences between the Native color picker and BiCEP were statistically significant for one of the orthogonal factors: Exploration. BiCEP received a higher rating for Exploration than the Native color picker. A t-test reveals a statistically significant ($p<0.01$) interaction effect between interface and block conditions, where participants that used BiCEP in Block 1 reported a higher rating for Expressiveness than the Native color picker. There were no statistically significant differences for the other orthogonal factors.

The results show that when participants used BiCEP first and then the Native color picker, the overall creativity ratings decreased. This finding suggests creativity rating decrease when there is reduction in control from manipulating all three colors dimensions at once (i.e., using the bimanual interaction provided by BiCEP) to only being able to control two or one color dimensions (i.e., using the color wheel and brightness slider in the Native color picker). Since exploration and expressiveness were rated as more important to the task, it is possible participants saw the interaction technique as allowing them to express themselves more because they had greater control over their exploration of the parameter space.

Therefore, for H_{11}, I can reject the null hypothesis in favor of the alternative, but only under certain conditions: H_{11} : ***When participants used BiCEP first****, the CSI factor Expressiveness was higher in BiCEP than the Native color picker.* For H_{12}, I can reject the null hypothesis in favor of the alternative: H_{12} : *The CSI factor Exploration was higher in BiCEP than the Native color picker.* For H_{13}, I cannot reject the null hypothesis: $H_{13,null}$: *The CSI factor Results Worth Effort* ***was not*** *higher in the BiCEP color picker than the Native color picker*

5.3.6 Maximization Scale Scores

There was no correlation (R<0.1) between the Maximization Scale scores and the satisfaction of the participants. Therefore, for H_3, I cannot reject the null hypothesis: H_3 : *Participant's Maximization Scale Scores* ***did not*** *correlate with their satisfaction with exploring the parameter space using BiCEP.*

Maximization scores did not correlate with the amount of time allotted. There were no interaction effects from the other independent variables. Therefore, for H_4, I cannot reject the null hypothesis: $H_{4,null}$ *Participant's Maximization Scale Scores* ***did not*** *correlate with the amount of allotted time they used.*

One thing that impacts the time spent coloring is the number of colors painted with. For example, a participant that paints all the background shapes green will take less time to color than a participant who paints the background with different patterns of color. After looking at the correlation between the number of colors used in the image and the total coloring time, I found a small effect size between the two variables (R=0.38). There was also a small effect size (R=0.45) between the actual

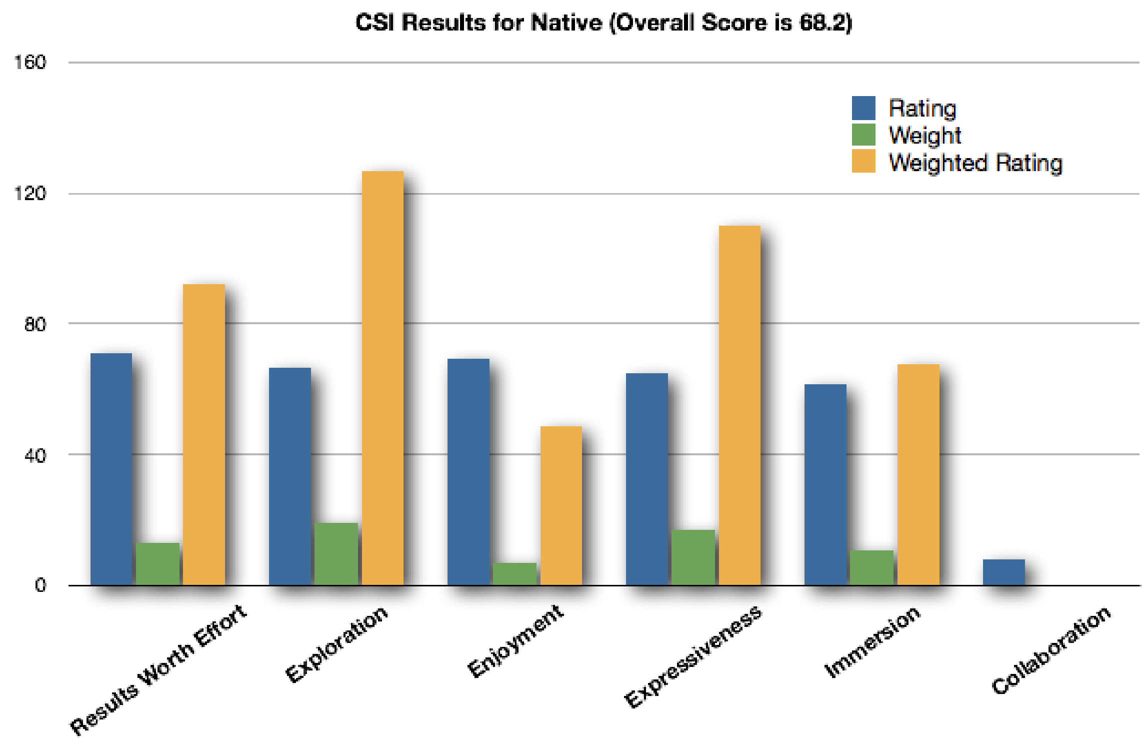


Figure 12: Native color picker's CSI scores for the 6 orthogonal factors for evaluating creativity support tools

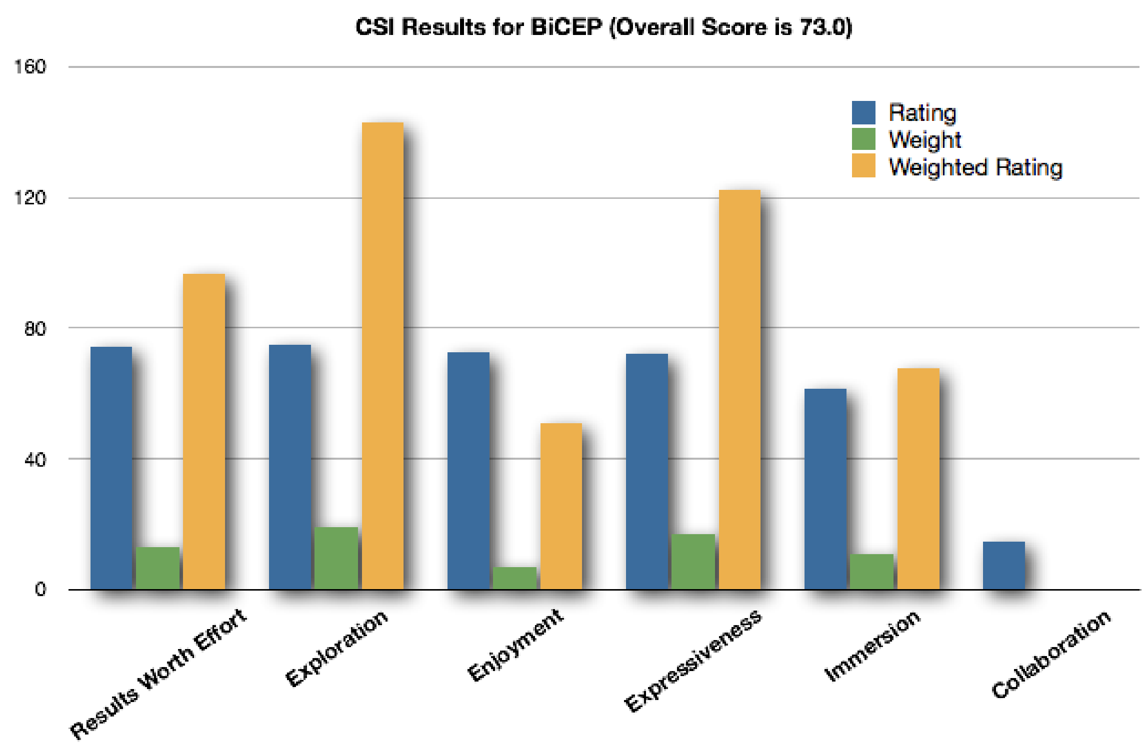


Figure 13: BiCEP's CSI scores for the 6 orthogonal factors for evaluating creativity support tools

percentage of the parameter space explored and the total time taken to complete the task. To round out the relationships, I also found a small effect size (R= 0.21) between the actual percentage explored and unique colors painted with. While other factors (e.g., external or unmeasured) may have contributed to the total time taken, the data suggests that both the exploration that occurred and the number of options used in the final design contributed to the total amount of time taken by participants.

The Maximization Scale and Perceived Exploration were correlated (R=0.22), with a small effect size. It should be noted that a Lack-Of-Fit test indicated a strong possibility (Prob>F <0.01) of a Type I error. This test suggests there is a strong probability that other independent variables may be missing from this correlation. None of the other measures I collected adjusted this fit. Even though there was a correlation, the reliability of this correlation is weak. Therefore, for H_5, I cannot reject the null hypothesis on the basis of reliability: $H_{5,null}$: *Participant's Maximization Scale Scores* ***did not*** *correlate with the perceived percentage of color space explored.*

5.3.6.1 Maximizers and Satisficers

The following analysis looks specifically at differences between Maximizers and Satisfiers, as well as phenomena that only occurs within a particular group. As a reminder, participants are divided into groups based on their responses to the Maximization scale. Because there were only a few Satisficers in my study (N=5), I have chosen to not include any analysis that involved correlations specific to Satisficers; the small number of Satisficers make the correlations less reliable and furthermore makes it difficult to defend any potential generalizations about Satisfiers.

Table 4: Distribution of interface preferences for Maximizers and Satisficers

	BiCEP	Native	Total
Maximizers	16	6	22
Satisficers	2	3	5
Total	18	9	27

For Maximizers, the maximization scores inversely correlated with participants satisfaction with their creative process while using BiCEP. An R value of 0.33 indicates a small effect size. There was no correlation with satisfaction with the final design. Therefore, for $H_{M.1}$, I can reject the null hypothesis, but in favor of an alternative with an inverse relationship: $H_{M.1a}$: *For Maximizers, Maximization scores* ***inversely*** *correlated with BiCEP satisfaction.*

There were no correlations between maximization scores and participants' satisfaction with the Native color picker. Therefore, for $H_{M.2}$, I cannot reject the null hypothesis: $H_{M.2,null}$: *For Maximizers, Maximization scores* ***did not*** *correlate with satisfaction with the Native color picker.*

Table 5 shows TimeOuts, the number of times participants used the maximum amount of time allowed (15 minutes), and Submits, the number of times participants used the submit button to indicate they were finished with their design. The difference between interfaces was not significantly different for Maximizers or Satisficers. Both Maximizers and Satisficers tended to submit their designs before the 15 minutes were finished. Therefore, for $H_{M.3}$, I cannot reject the null hypothesis: $H_{M.3,null}$: *Maximizers* ***seldom used*** *all the allotted time for their trials.* Also, for $H_{S.1}$, I cannot reject the null hypothesis: $H_{S.1,null}$: *Satisficers* ***did not*** *spend more time with BiCEP versus the Native color picker.*

Table 5: Distribution of TimeOut and Submit events for Maximizers and Satisficers

	TimeOuts	Submits	Total
Maximizers	8	36	44
Satisficers	3	7	10
Total	11	43	54

There were no statistically significant differences in satisfaction between Maximizers and Satisficers. There were no interaction effects from block or interface conditions. For $H_{S.2}$, I cannot reject the null hypothesis: $H_{S.2,null}$: *Satisficers' satisfaction* ***was not*** *higher overall compared to Maximizers.* This result is a bit surprising, as based on the literature, Satisficers should overall be more satisfied than Maximizers. There is the possibility that the lack of Satisficers created more variance in the data, therefore making statistical significance difficult to achieve.

Between Maximizers and Satisficers, the difference in the number of visited colors per minute was statistically significant, with Maximizers exploring more colors per minute than Satisficers (Δ 1.71 SE 0.80). Maximizers also visited more colors per minute using BiCEP than Native (Δ 1.41 SE 0.37 $p<0.01$). This suggests that Maximizers did seek out more options and were able to explore those options more fluidly using BiCEP than the Native color picker.

5.3.7 Other

The interface preferences of participants overall is reported in the 'Total' row in Table 4. While more participants (66%) preferred BiCEP to the Native color picker, the percentage is only marginally greater than chance (50%). Therefore, for H_{14}, I can reject the null hypothesis in favor of the alternative, but do so cautiously: H_{14} : ***66%*** *of participants preferred to use BiCEP over the Native color picker.*

I collected data on each participant's experience with graphic design software and found no correlation with other metrics presented earlier (e.g., SatCP, SatFD, Maximization scale, unique colors visited, CSI scores, etc). Part of the reason could be that over 50% of participants categorized their graphic design experience as "Intermediate." The question was on a four point scale: Little/None (N=5), Beginner (N=2), Intermediate(N=14), and Advanced(N=6). I would not assume such a high bias towards intermediate experience with graphic design software, which indicates more questions should have been asked in order to validate and confirm the level of graphic design knowledge and experience.

I have omitted any analysis concerning the contextualized maximization scale, as there were not enough satisficers to make any generalizable claims between the two groups of participants.

5.3.8 Free Response

There were various themes that emerged from the free response section of the questionnaire participants took at the end of the study.

Brightness: Participants reported on how their interaction with color brightness was different between BiCEP and Native. P26 stated they used brightness more often with BiCEP: "Since it was much easier to change the brightness of the chosen color, I found myself doing so more often." P26 also stated they used brightness less with Native: "having to click somewhere else again just made it a bit more inconvenient to change brightness, so I only did it when I really needed." P09 stated they were "able to control the hue and the brightness simultaneously." Controlling brightness was not alway something participants wanted to think about, P27 stated: "I could select the color first and then worry about the brightness."

Wider Color Range: Participants reported exploring a wider range of colors with BiCEP. P23 stated "I found myself using broader spectrums of color while in the other picker I was staying in a couple of ranges without even thinking of it because of having to separately change that aspect." P30 stated: "it was easier to find a wider range of color options."

Gradients: Participants reported that BiCEP helped try using gradients in their design. P28 stated "The dual cursor system allowed me to test different colors as a gradient. It was smoother and faster to select from different shades of a color than clicking and dragging. It allowed me to continuously see the different color options as opposed to restricting me to where I had clicked or dragged the cursor to." P07 stated "It is easy to pick colors especially that subtle varying colors." Using gradients with the Native color picker requires additional mode-switching as stated by P07, "[I would] have to click and check, click and check all the time."

Dual-Cursor Familiarity: One hurdle in our study was that some participants commented on how the Bimanual interaction changed the way they were accustomed to using the trackpad and/or system cursor. P24 stated that BiCEP "ended up switching the way I was using the trackpad, which made it a tad uncomfortable for me." P01 stated "Having to control both hue and saturation with the dual cursor got to be difficult because I was having to focus on both aspects at the same time.” However, P10 was able to overcome this hurdle stating “I found it difficult at first but i later realized that it was much comfortable to use that than the other one.”

Precision: Some participants commented on the precision of BiCEP. P15 stated “it was difficult to match the precision of the single cursor tool with the dual cursor tool.” This is an expected observation from participants, as they are controlling all three values, meaning smaller cursor changes results in bigger color changes.

Wider versus Narrow Search: Participants commented that BiCEP allowed them to explore a wider range of colors. However, they also commented on how Native allowed them to have a more narrow control of the color ranges. As stated above, one participant said that they did not want to think about brightness. They felt like Native allowed them to focus just on hue or saturation and that there was no need to touch the brightness bar. The three dimensional explorability in BiCEP may not be appropriate in situations where the user wants a more constrained exploration (just two or one dimensional spaces). Both P19 and P15 stated advantages of both, but still preferred Native.

5.3.9 People’s Choice

The results of the voting for ‘coolest’ image by Mechanical Turkers are shown (in order) in Figure 14. While an image created by the Native color picker took first place, the other top five images were created using BiCEP. Interestingly, the first and second image were both created by the same participant. This suggests that the

Figure 14: Top 5 most creative designs (left to right) as judged by Mechanical Turkers, N: Created using Native color picker. B: Created using BiCEP. Note images 1 and 2 were created by the same participant

creative capacity of the participant outweighed the effects of using BiCEP versus the Native color picker. While rating something as 'cool' is only a small part of creativity, the results suggest that images created using BiCEP are more likely to be judged as 'cooler' by online voters.

5.4 Discussion

Here I discuss some of the interesting relationships between findings as well as the limitations and confounds that were present in my data.

5.4.1 Exploration

Not only did some participants feel like they explored a wider range of colors, they did in fact explore a wider distribution and a higher percentage of the parameter space using BiCEP over Native. Exploration however was not always directly related to satisfaction: Higher exploration did not always result in higher/lower satisfaction. This indicates other factors contributed to participant satisfaction.

5.4.1.1 Expressiveness and Exploration

Expressiveness was the other orthogonal factor that participants felt was important to the coloring task. Expressiveness was not measured outside of the CSI survey. Even with high exploration, if the participants were not able to find a suitable color (or combination of colors), this could have impacted expressiveness negatively. While participants were able to color however they wanted, they were restricted to the images we provided and could only paint using the bucket fill. It is unclear how these constraints may have impacted expressiveness. However, high exploration could give them more options to choose from, but from the literature, more options does not necessarily increase expressiveness.

5.4.2 Native Support

Dual-cursor interaction is not natively supported on most laptops and personal computers. The dual-cursor interaction in BiCEP is confined to the window of the BiCEP interface. Participants had to switch from dual-cursor mode to single-cursor mode whenever they wanted to paint using the color they selected. They must then switch back to dual-cursor mode when they are want to select a new color. It is also important to note that if dual-cursor was natively supported, participants would not need to clutch, they could control the brightness bar with either cursor. Future work could investigate how participants use natively supported dual-cursors to control the interaction. This is where the cursor constraints of BiCEP do come in handy. They maintain the cursors in a localized space, so they do not get lost on the screen, which often occurs using a single cursor.

The non-native support for dual-cursor interaction also means participants were accustomed to two-fingers on the trackpad being used for scrolling and any more than two-figures on the trackpad will be used for gestures. The other issue was having to make sure all our participants did use their other hand to click the trackpad. I have observed pilot participants control the system cursor with their right hand and then use their left hand to click. The technical work-around I am using for supporting dual-cursor interaction using a MacBook trackpad may require additional time for the user to become familiar with the interaction technique. While participants received time to become accustomed to the dual-cursor interaction during the matching task, it may have been too controlled as the matching task did not have the same dynamics of interaction as the coloring activity (e.g., window dragging, eyedropper, undos, painting in a different window, etc).

5.4.3 Satisfaction and Preference

While I did not find a difference between interfaces for participant satisfaction with their design process, I did find that users were more satisfied with their results from BiCEP. 44% of participants preferred the Native color picker to BiCEP. Which tells me that satisfaction with the final design does not necessarily mean a participant would prefer to use a particular interface. Many of the themes from the free response have to do with the process of choosing colors and less to do with what they created using the interfaces. This indicates that just because a participant creates a more satisfying image, it does not mean that the participant would prefer to use the interface that helped them create the ‘cool’ image. This leads to me believe that interface preference is more related to the creative process and experience of using the interface (e.g., most of the participants that had trackpad issues, preferred the Native color picker.)

5.5 Conclusion

In this Chapter, I have shown how changing the interaction used to explore the parameter space can impact the satisfaction participants have with their final designs. I have also demonstrated that BiCEP supports a fluid exploration of the HSB parameter space, allowing participants to explore unique and perceptually different colors in total and per minute, as well as explore a larger percentage of the perceivable parameter space. I have also found that participants who scored higher on the Maximization scale were less satisfied with their creative process using BiCEP. The results also indicate a tie between exploration, expressiveness, and the creative process. The results here indicate that changing the interaction technique can support the creative process, but that other factors such was familiarity, can negatively impact user satisfaction.

CHAPTER 6: EXPLORABILITY AND SATISFICING IN DIFFERENT SEARCH SPACES

In Chapter 5, I presented a study where I changed the interaction technique while holding the size of the search space constant. In this chapter, I describe a study where I changed the size of search space by manipulating the granularity of control while holding the interaction technique constant between two similar interfaces (each with a different search space size) for manipulating image properties.

6.1 Motivation

Exploratory satisficing plays a role in interfaces and technologies where it is hard to evaluate all possible outcomes with our limited human memory. As a result, we search through alternatives until one meets an acceptable threshold. Consider trying to select a font for a poster. For most people, there are only a hundred or so fonts, but for many designers there are thousands. In either case, going through, selecting, and applying a font can take time, so at some point designers may say that the choice is 'good enough'. The font is just one design consideration (or parameter) amongst many: colors, alignments, visuals, etc. My broader research goal is to understand the role of interaction techniques and user interface design on satisficing and satisfaction. In the previous chapter, bimanual interaction was applied to alleviate the tedium associated with exploration of large design spaces. In this chapter, I present a study that investigates a completely different approach to the problem: reducing the size of the space. This work looks to understand the relationships between satisficing and explorability, in this case, the granularity of control of slider or how many options a slider provides you when trying to manipulate an image with controls such as: brightness, blur, sharpen, etc. Below I describe the two image manipulation tools that I used in my studies.

6.2 Creamy and Crunchy Sliders

There are two versions of my image manipulation tool. The Creamy Sliders Interface (see Figure 15) allows for fine-grained control of image properties using the provided sliders. Fine-grained control allows a property like brightness to be set to all the integer values between -100% and 100%. The Crunchy Sliders Interface (see Figure 16) allows for course-grained control of the properties. In course-grained control, instead of 200 values, brightness can only be set to 5 pre-defined value settings: low (-100% or very dark), high (100% or very bright), and three in the middle (i.e., -50%, 0%, 50% or darker, original, brighter).

The semantic contrast between the names of Creamy Sliders and Crunchy Sliders was inspired by peanut butter. Like creamy peanut butter, Creamy Sliders are smooth, making the interaction fluid across the continuous range of slider values. Like crunchy peanut butter, Crunchy Sliders have bumps (or invisible stopping points), making the interaction much more discrete and not as fluid moving, since the sliders stop at (invisible) tick marks.

The interfaces also allow the user to collect "saved image states" as they manipulate the image. These states are represented by a collection of thumbnails on the left side

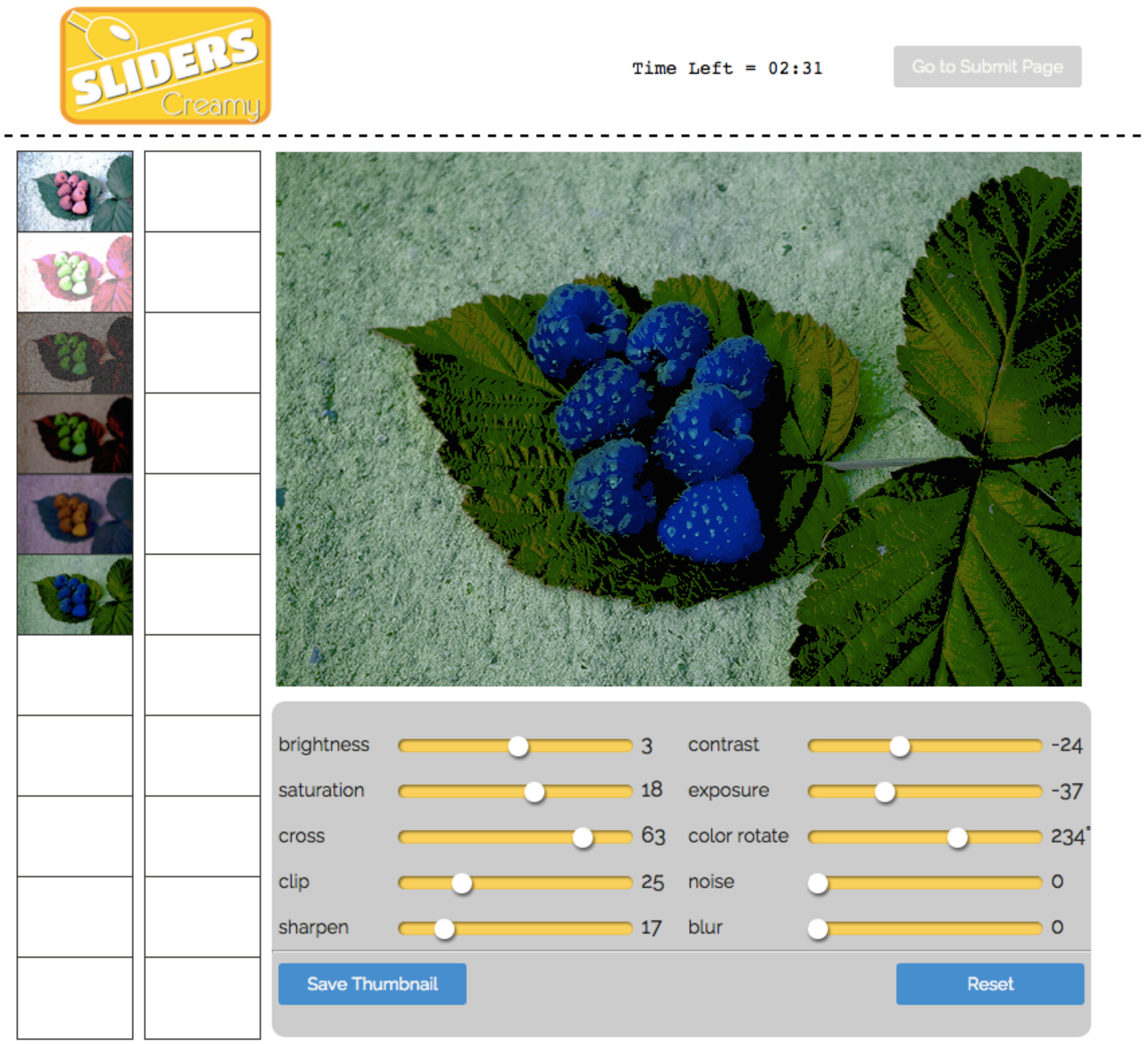


Figure 15: Creamy Sliders image manipulation tool: Allows the user to choose 100 different levels [0,100] for sliders that are additive only (e.g, sharpen, noise, blur) and 200 different levels [-100,100] for sliders that allow positive or negative values (e.g., contrast, brightness)

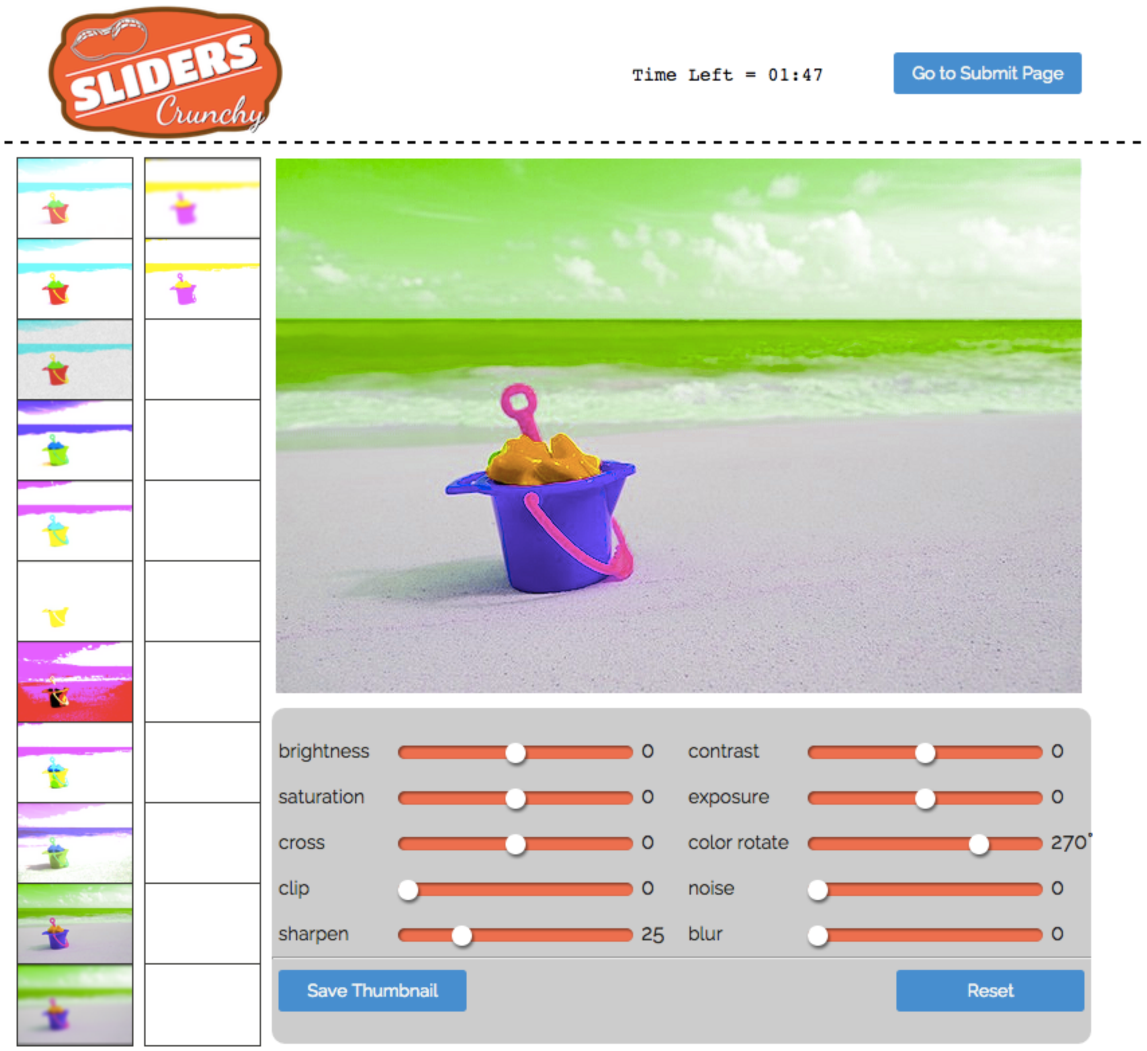


Figure 16: Crunchy Sliders image manipulation tool: Allows the user to choose 5 different levels for sliders that are additive only (e.g, sharpen, noise, blur) and 5 different levels for sliders that support positive or negative values (e.g., contrast, brightness)

of the interface. Users can return to any of these states by clicking the thumbnails. These states also serve as markers of explored points in the parameter space.

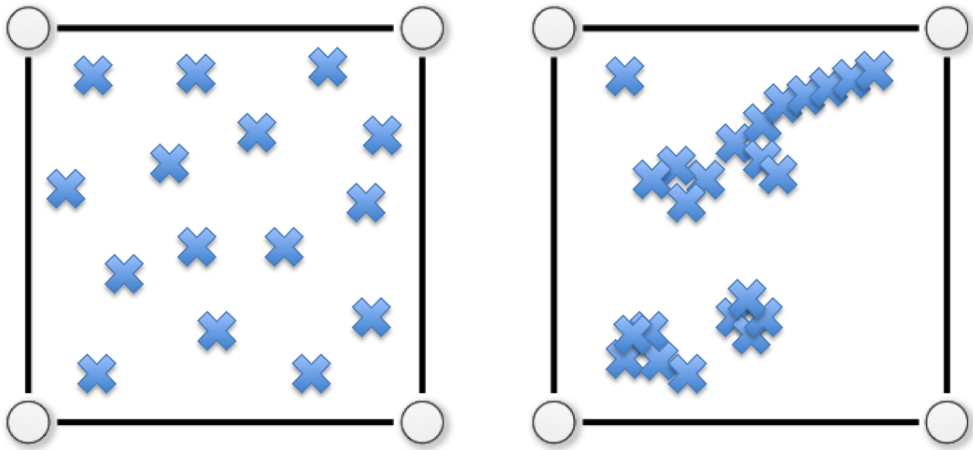
6.2.1 Explorability

Creamy and Crunchy Sliders differ in how they enable exploration of the parameter space. If we assume that D, the dimensionality of the space of possibilities, is equal to N, number of parameters available, then for either of the slider interfaces (with $N = 10$ different sliders) the parameter space of possibilities exists in a 10 dimensional space. Most people are unlikely to be capable of fully comprehending a $10D$ parameter space. To make this easier, in Figure 17, I have created visual aids of the $10D$ space projected onto $2D$ planes (i.e., a square). The $\times$'s represent hypothetical points in the parameter space that a user may visit. This visual aid also demonstrates the possible differences in users' movements through the parameter space based on the control interface (i.e., Creamy and Crunchy Sliders).

6.2.2 Coverage of Parameter Spaces

Crunchy Sliders give users course-grained control over the parameters. This constrains the user from spending lots of time refining a single idea and supports the user in exploring a more diverse set of possibilities across the entire parameter space (see Figure 17(a)). I refer to this as a *Uniformly Distributed Coverage of the Parameter Space*. The Crunchy Sliders interface was designed to approach the problem of exploring large parameters spaces with the assumption that an increase in the diversity of ideas explored will lead to a more desirable outcome.

Creamy Sliders give users fine-grained control over the parameters. This enables



(a) Uniformly distributed coverage of parameter space

(b) Non-uniform coverage of parameter space

Figure 17: Coverage of parameter space

the user to fine-tune their ideas or try completely new ideas (see Figure 17(b)). With the lack of constraints, the user may explore fewer unique ideas. I refer to this as a *Non-uniform Coverage of the Parameter Space.*

6.3 Mechanical Turk Study

I recruited 64 Mechanical Turk users in an online study to understand the broader relationships between explorability, satisficing, and satisfaction in the interface. I chose to use Mechanical Turk over a traditional laboratory study because Mechanical Turk supports high-volume participant recruiting, is less susceptible to population coverage error, and has a lower risk of contaminated subject pools, dishonest responses, and experimenter effects [43]. In addition to the Mechanical Turk users, I recruited 4 additional participants to do the same study in a controlled laboratory setting, where I did follow-up interviews with them. In the following sections, I describe the methodology and experimental procedure.

6.3.1 Questionnaires and Surveys

The demographic questionnaire at the beginning of the study gathered background information including participants' familiarity with image editing tools (e.g., Instagram, Photoshop, etc.). The follow-up questionnaire at the end of the study asked the participant about which interface (Creamy or Crunchy) they preferred and why. The Maximization Scale (see Chapter 4 Section 4.4.2 Maximization Scale) and the Creativity Support Index (see Chapter 4 Section 4.6.1 Creativity Support Index) were also administered during this study. In the next few sections, I provide details concerning the scales and questions that were tailored to this study.

6.3.1.1 Contextualized Maximization Scale

As stated in Chapter 5, I strongly believe that maximization behavior is highly contextualized. For this study, I used the same 3-item scale from Section 4.4.2.1 that shortens and contextualizes the Maximization scale to the domain of this study. Creating a contextualized scale for this domain (image manipulation) would have been difficult since many of the participants may not do image editing on a regular basis. Selecting a color (the context for the 3-item scale) is related to image manipulation and is a task that many users will have performed more frequently. The 3-item scale was included at the end of the Maximization Scale.

See Chapter 5 Section 4.4.2.1 Contextualized Maximization Scale for more details.

6.3.1.2 Satisfaction

At the end of each condition block (i.e., after editing an image with either Crunchy or Creamy), participants selected their final design from the set of saved thumbnails and then rated the following claims on a 7-point Likert scale:

Satisfaction: B1: “I am satisfied with my final designs.”

Suitability: B2: “I think my final designs are suitable as Album/DVD covers.”

Carefulness: B3: “I made my final designs carefully.”

Satisfaction: B4: “I am satisfied with my design process with the [Creamy/Crunchy] sliders.”

Immenseness: B5: “I felt overwhelmed by the choices offered by [Creamy/Crunchy] sliders.”

See Chapter 4 Section 4.5 Satisfaction for more details.

6.3.1.3 Explorability

At the end of each condition block (i.e., after editing an image with either Crunchy or Creamy), participants answered the following question using a scale of 0% - 100%.

Exploration: B6: “Given the huge number of combinations of slider positions, it is impossible to explore all combinations. What percentage of those combinations do you think you explored using [Creamy/Crunchy] sliders?”

See Chapter 4 Section 4.3.2 Self-Reported Explorability for more details.

6.3.2 Procedure

Participants viewed the study as a posting on Amazon's Mechanical Turk. These postings are called HITs (Human Intelligence Tasks). The HIT displays a description of the study, the expected time commitment to complete the study, compensation, and the informed consent. In the description, participants were asked to use Google Chrome or Mozilla Firefox to ensure browser compatibility. Included on the HIT page was a tool for helping participants know whether their browser supported features required for the study (e.g., javascript, cookies, local storage). Participants could then accept the HIT and agree to participant in the study.

After agreeing to participate, participants were taken to web pages containing a demographic questionnaire and a pre-survey (i.e., the Maximization Scale survey and Contextualized Maximization Scale survey).

Participants were told to use the Creamy and Crunchy interfaces to "Design an album/DVD cover for a band or movie that you like." Participants used 2 images that I pre-selected. The participants used the Creamy/Crunchy web tools to modify properties of the image (e.g., brightness, contrast) and apply filter effects (e.g., color mix, sharpen). One image was used in the Creamy Interface and the other image was used in the Crunchy Interface. While the image ordering did not change, the ordering of the interfaces was counterbalanced. Participants used both versions in a single sitting, with again the interface ordering counterbalanced between participants. Participants were encouraged (in the instructions) to save thumbnails as a way of saving and reverting back to previous work.

Each participant was asked to spend at least two minutes changing each image with a maximum time of five minutes allowed. After two minutes, a button became available for them to submit their image. Upon clicking the submit button, a confirmation box appeared with a reminder about the bonus. At five minutes (or when the participant submitted), the interface displayed the current image and any saved thumbnails. The participant then selected one of those images for submission. The rationale behind such a short amount of time is to elicit satisficing behavior from participants, where they must decide when their designs are 'good enough' given the time constraints (see Chapter 4 Section 4.4 Satisficing). During pilot studies, five minutes was an adequate amount of time for participants to complete the task. After each condition (interface version) participants filled out the Creativity Support Index (CSI) [11] and a short survey on satisfaction and exploration.

At the end of the study, participants completed a brief follow-up questionnaire and the last part of the CSI, where they do pair-wise rankings of the creativity factors.

To encourage participants to engage fully in the study, I asked participants to design a album or DVD cover for either a band or movie that they like. Additionally, participants were told that they would be evaluated on the quality of their work, based on that evaluation they would be paid a bonus of \$1.00 in addition to their compensation of \$4.00. The monetary value of the compensation was based on the minimum hourly wage in the United States in 2014. I defined (unbeknownst to the participants) the "quality of work" as having moved a slider at least once. The specific wording and metric were chosen as a way to decrease the need for overly complex deception strategies. In the end, all participants received \$5.00. The task prompt and monetary bonus incentive were designed to encourage participants to be more active and to engage fully in the study.

6.3.2.1 Non-Mechanical Turk Studies

The 4 non-Mechanical Turk users performed the same study outlined above with two major differences. After the study, I conducted one-on-one semi-structured interviews with each participant to ask them in-depth questions about their experiences. These interviews took approximately 15 minutes to complete. Additionally, participants receive a larger compensation for their time (i.e., a $10 gift card). The following are examples of questions I asked in the interview:

1. In your own words, how would you describe the differences between the Crunchy and Creamy slider interfaces?
2. Were you able to achieve the goal of making an Album/DVD Cover?
3. How might these differences have [inhibited/supported] your goal of making an Album/DVD Cover?
4. When did you choose to save thumbnails?
5. Did you ever have to undo an action? When?
6. Did you ever use the submit button? Why?
7. Were you aware of how much time you were spending on creating your design?
8. What other sort of [benefits/drawbacks] (if any), do you think were offered by the [Crunchy/Creamy] interface?

6.3.3 Images

All participants were provided with two different images (see Figure 18), one in each block of the study. The selection process for these two images was well thought-out and took into account several design considerations. I first created a pool of images that contained a dominant object that falls into one of six categories: Food, Artifact, Clothes, Vehicle, Animal, Furniture. The core semantic category consideration for the six categories was that they were mutually disparate such that no one category is a subcategory of another. I avoided selecting images with an overbearing visual appeal or familiarity. Additionally, each image falls into one of the following three categories pertaining to the dominant object:

Isolated: The dominant object is solitary and the sole focal point of the image.

Grouped: The dominant object is co-located with similar objects and the group is the main focal point of the image.

Scenic: The dominant object is situated within a broader photographic perspective and easily discernible from the rest of the image content.

To avoid fatigue effects from a long study, I narrowed down the pool of images to two images. These images are from different semantic categories and object dominance. Both images also contain at least one vibrant color and a layer of natural tones. One observation made by a digital-arts colleague was that the images are "dominantly defined by contrasting colors, green-red contrast largely define both [images]". This contrast in turn helps to reduce variability in the study. It is difficult to fully control

(a) Food-Grouped: Berries

(b) Artifact-Scenic: Bucket

Figure 18: Images provided to participants during study

for this type of variability, as even the image of the beach is more saturated and has a higher contrast. However, the main goal was to chose images that varied enough that putting each through filters would yield interesting results. These images also look 'fine' by themselves. If an image looks really good, it sets a high standard for the user. If the image looks poor, the user may not be as engaged. I have described the semantic categorization and selection process to emphasis that the selection process was not arbitrary or random.

6.3.4 Hypotheses

This section contains my general hypotheses that guided the design and analysis of this study. In the analysis and results section, I will use the data from this study to test each hypothesis. Note that for H_3-H_5, a high Maximization Scale Score indicates the participant is a maximizer (more than 4.0) and a low score indicated the participant is a satisficer (less than 4.0). For H_3-H_5, I use the raw Maximization Scale score as a continuous variable in correlation tests.

H_1: Participants will cover a more even distribution of the parameter space with Crunchy than with Creamy

H_2: Participants will be more satisfied with how adequately they explored the design space using Crunchy over Creamy

H_3: Participant's Maximization Scale Scores will inversely correlate with their satisfaction with the percentage of the design space explored using Creamy

H_4: Participant's Maximization Scale Scores will correlate with the amount of allotted time they used

H_5: Participant's Maximization Scale Scores will inversely correlate with the perceived percentage of design space explored

H_6: Participants will save more thumbnails (i.e., more unique results) with Crunchy than Creamy

H_7: Participants will be more satisfied with their final designs from Creamy versus Crunchy

H_8: Participants with higher exploration (percentage of space covered) will have a higher user satisfaction with the design process

H_9: Participants with higher exploration (percentage of space covered) will have a higher user satisfaction with the final product

H_{10}: Participants will perceive having explored a higher percentage of the design space with Crunchy than with Creamy

H_{11}: The CSI factor Expressiveness will be higher overall in Creamy than Crunchy

H_{12}: The CSI factors Exploration will be higher in Crunchy than in Creamy

H_{13}: The CSI factors Results Worth Effort will be higher in Crunchy than in Creamy

The following are additional hypotheses that are specific to participants classified as either Maximizers or Satisficers:

$H_{M.1}$: Overall, Maximizers will prefer Creamy to Crunchy

$H_{M.2}$: Maximization scores will correlate with Creamy satisfaction

$H_{M.3}$: Maximization scores will inversely correlate with Crunchy satisfaction

$H_{M.4}$: Overall, Maximizers will use all the allotted time to complete the task

$H_{S.1}$: Overall, Satisficers will prefer Crunchy to Creamy

$H_{S.2}$: Satisficers will spend less time with Creamy versus Crunchy

6.4 Analysis, Results, and Short Discussion

I analyzed my data using the Restricted Maximum Likelihood (REML) method of Standard Least Mean Squares (SLS), which is similar to a Repeated Measures Analysis of Variance (ANOVA). REML-SLS is a mixed-model approach that better accounts for the variance and random effects intrinsic to my unbalanced data[3]. Reported averages were calculated using Least Mean Squares. Reported averages also include standard errors (SE), which takes into account the Standard Deviation and

[3]While certain pairs of data were not equally balanced (e.g., Maximizers versus Satisficers, Creamy versus Crunchy Preference, etc.), each condition had enough data points from which conclusions could be drawn.

the sample size. When applicable, the change or numerical difference between conditions is reported using the standard delta notation (Δ). Effect sizes for correlations were analyzed using Cohen's recommendations: $R>0.2$ (small effect), $R>0.5$ (medium effect), and $R>0.8$ (large effect) [13].

Group W used Creamy Sliders in Block 1 and Crunchy Sliders in Block 2. Group X used Crunchy Sliders in Block 1 and Creamy Sliders in Block 2. Unless otherwise noted, all differences between condition are significant ($p<0.05$). Some of the results included extend beyond the hypotheses described earlier in this chapter. Because of the large number of different combinations of variables, in each of the following subsections, there is a brief discussion on the possible implications of and reasons for that described result. The discussion section at the end of this chapter is a high level summary of all these points.

6.4.1 Participant Removal

The data from 10 of the 64 Mechanical Turk participants were not included in the results of this experiment. Five of these participants had corrupted or incomplete logs. The other five participants failed my quality control tests (e.g., a Likert scale statement with instructions to leave the slider in the middle). The analysis of the studies done on Mechanical Turk uses the data collected from the remaining 54 participants. I did not include the log data from the 4 non-mechanical turk participants in the larger pool; these participants were treated as separate, but related case studies.

6.4.2 Preferences and Interviews

In the final questionnaire, all participants were asked about their preferences between Crunchy and Creamy Sliders and asked to provide an explanation. The explanations and non-mechanical turk participants' responses to the semi-structured interview questions were analyzed using thematic analysis [8]. Anecdotes from the explanations and interviews have been included throughout the discussion sections below. Notation for Mechanical Turk participants is $\mathrm{MP}_{n,m}$, where n is the participant number [01-54] and m is their classification ($M =$ Maximizers, $S =$ Satisifcers), e.g., $\mathrm{MP}_{21,s}$ is Mechanical Turk Participant number 21 who I classified as a Satisficer.

6.4.2.1 Interviewed Participants

All interviewed participants were classified as Maximizers, although two of the participants were close to the borderline between Satisficers and Maximizers. The scores for my interviewed participants are presented in Table 6. Since all my participants scored higher than 4.0, they would be classified as Maximizers accounting to Schwartz [46]. The literature on Maximizers suggests that Maximizers are high achievers (e.g., college students), therefore my university's subject pool is likely to have been saturated with Maximizers.

When a score is higher than 5.5, the participant is considered an extreme maximizer (e.g., usually 10% of samples). One participant scored over 5.5 (5.6) and another came close with a 5.3. The other two participants had scored relatively similar scores as well (4.2 and 4.3). Similar to the work of Oulasvirta et al. [42], I performed a '*post-hoc* median split' in order to create two groups of Maximizers: Low and High. I

Table 6: Maximization scores and interface ordering for the interviewed participants

Interface Ordering	Participant ID	Maximization Score
Creamy, Crunchy	$\mathrm{NP}_{55,HM}$	5.3
	$\mathrm{NP}_{56,LM}$	4.2
Crunchy, Creamy	$\mathrm{NP}_{57,LM}$	4.3
	$\mathrm{NP}_{58,HM}$	5.6

performed an ANOVA on the Maximization Scale Scores and the sub-groups. The difference between the groups was significant ($p<0.05$), which suggests the groups are distinct enough (i.e., very little overlap in variance) to be used for comparison. Since my sample size was small ($N=4$), the claims I make here are not generalizations about how Satisfiers versus Maximizers approached the study. I use the low/high groups to look at how the results from the interviews may have varied based on where participants scored along the spectrum of Maximization Behavior.

The notation for non-Mechanical Turk participants (e.g., lab-study with interviews) is as follows: $\mathrm{NP}_{n,m}$, where n is the participant number [55-58] and m is their classification (LM = Low-End Maximizers, HM = High-End Maximizers), e.g., $\mathrm{MP}_{55,HM}$ is non-Mechanical Turk Participant number 55 who I classified as a High-End Maximizer. The numbering of the non-Mechanical Turk participants starts at 55 in order to give all participants a unique numerical identifier.

6.4.3 Interfaces and Ordering in Mechanical Turk Studies

The following is an analysis of the data collected from the Mechanical Turk studies. When necessary, I have included anecdotes from the interviews done with non-Mechanical Turk participants. I used these anecdotes to clarify or reinforce my findings, but do not include the non-Mechanical Turk participants in the pool of data.

6.4.3.1 Total Task Time

The average total time spent on the design task was significantly different between blocks ($p<0.01$): 4.69 mins (SE 0.11) for Block 1 and 4.23 mins (SE 0.11) for Block 2. There were no significant interaction effects overall between the interface used and the block number ($p=0.61$). A t-test revealed that participants spent more time using Crunchy Sliders (4.80 mins, SE 0.16) when it was the first interfaced used (i.e., Block 1) than Creamy Sliders in Block 2 (4.19 mins, SE 0.16) and Crunchy Sliders in Block 2 (Group W) (4.25 mins, SE 0.177). There was no significant difference of total time between Creamy Sliders in Block 1 and the other conditions.

The shorter task completion time in Block 2 versus Block 1 could indicate one or more effects. For example, learning effects could have occurred as $NP_{57,m}$ stated that in Block 1, he tinkered with the blur filter since he did not fully understand how it worked. After 'tinkering', he did not feel he would need the blur slide, and he stated he did not use it in Block 2. In Block 1, participants are discovering the features of the interfaces for the first time, while in Block 2, the newness of the image manipulation system has worn off. Finally there could have been fatigue effects, where participants just wanted the study to be over with. Even though I offered incentives to keep participants engaged, like any study, there could have been a few participants that tried to finish the study as quickly as possible.

The results of the t-test suggests that the ordering of the interfaces could have an effect on the Task Time. While using Crunchy in Block 2, participant $NP_{56,m}$ stated "I was waiting for it to be over." Other participants stated that they found Crunchy

Sliders limiting ($NP_{55,m}$) or that Crunchy offered less options than Creamy ($NP_{56,m}$). This effect could have been exacerbated in Group W where participants used the fine-grained control of Creamy, and then felt like that control was taken away when they used Crunchy during the second block. This is in contrast to Group X who used Crunchy first, unaware that the interaction with Creamy Sliders would offer more fine-grained control. This notion is supported by $NP_{58,m}$ who stated "[with Crunchy Sliders] you don't have much choice, and you have to work with what you have."

6.4.3.2 CSI Scores

The scores for the CSI were significantly different ($p<0.01$) between the two interfaces; Creamy Sliders: 71.95 (SE 2.72) and Crunchy Sliders 62.64 (SE 2.72). There was a significant block effect: Block 1: 70.79 (SE 2.72) and Block 2: 63.64 (SE 2.72). As a reminder, a higher CSI score indicates better creativity support. The interaction effect between the interfaces and blocks was not statistically significant ($p\approx0.06$). Because the p-value was near the threshold of 0.05, I further invested this phenomenon by performing a t-test on this interaction effect and found that the CSI Score for Crunchy Sliders in Block 2 was significantly lower than the other three conditions. There was no significant interaction effects between the other three conditions (Creamy Sliders in Block 1, Creamy Sliders in Block 2, and Crunchy Sliders in Block 1).

The difference in CSI scores between interfaces suggests that the Creamy Sliders interface better supports creativity compared to Crunchy Sliders for the task of image manipulation. The higher CSI scores in Block 1 could suggest that by Block 2, participants had gained a better understanding of how to rate the tools, and were therefore more critical. The t-test result showing that the CSI score for Crunchy Sliders in Block 2 was significantly lower, could suggest that reducing the accessible parameter space influenced how critical users were of how well an interface supports creativity. This effect is discussed later on in this chapter.

The overall CSI score for Creamy was 71.9 (SD = 17.82) out of 100 and the overall CSI score for Crunchy was 62.2 (SD = 22.99). The individual categorical ratings are shown in Figures 19 and 20. A pair-wise comparison of all 6 orthogonal factors reveals that Exploration and Expressiveness were rated as more important to the task, and thus the amount that these two factors contributed to the overall ranking is high. This suggests the Creamy Sliders and Crunchy Sliders generally provide good support for explorability, and the rating for Expressiveness suggests that the participants felt that they were able to express themselves using the interface to create their final designs.

The differences between Creamy and Crunchy were statistically significant ($p<0.01$) for three of the orthogonal factors: Exploration, Enjoyment, and Expressiveness. Creamy Sliders received a higher rating for each of these factors. This suggests that overall, for the the task of image manipulation, Creamy Sliders better supported creativity than Crunchy Sliders. For H_{11} I can reject the null hypothesis in favor of the alternative: H_{11} : *The CSI factor Expressiveness will be higher overall in Creamy than Crunchy.* For H_{12}, I can reject the null hypothesis, but in favor of the inverse alternative: H_{12a} : *The CSI factor Exploration was rated higher for Creamy Sliders than Crunchy Sliders.* For H_{13}, I cannot reject the null hypothesis: $H_{13,null}$: *The CSI factor Results Worth Effort was similar for Crunchy Sliders and Creamy Sliders.*

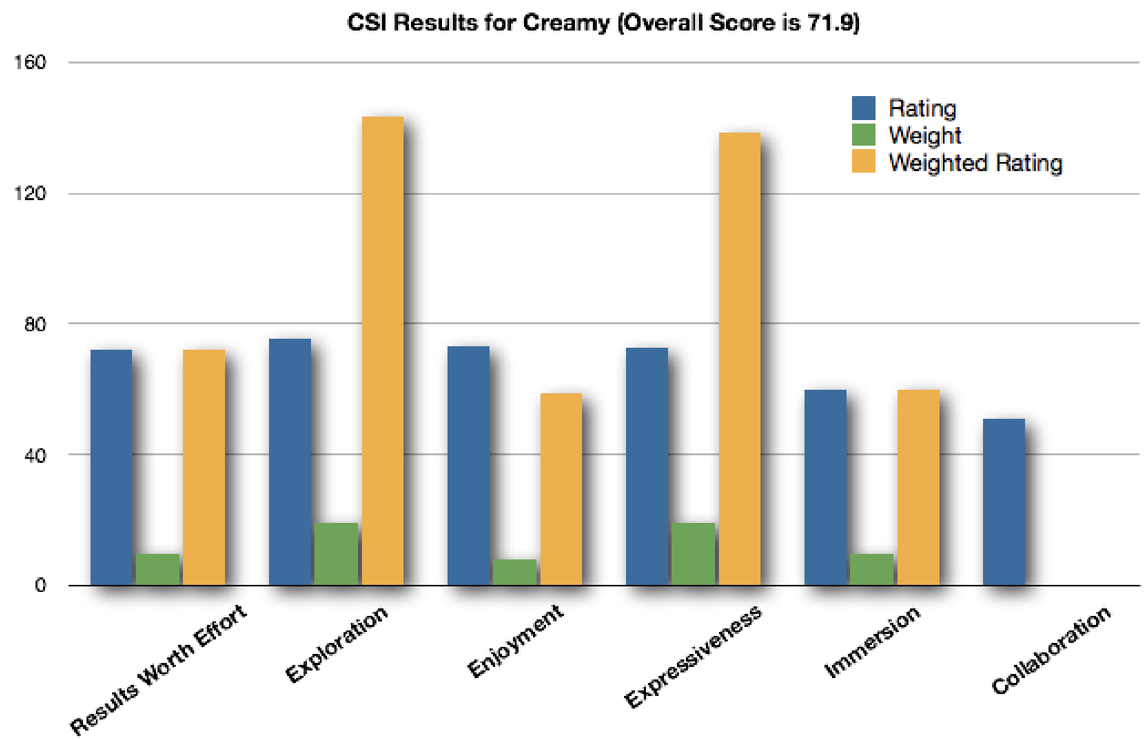


Figure 19: Creamy Slider's CSI scores for the 6 orthogonal factors for evaluating creativity support tools

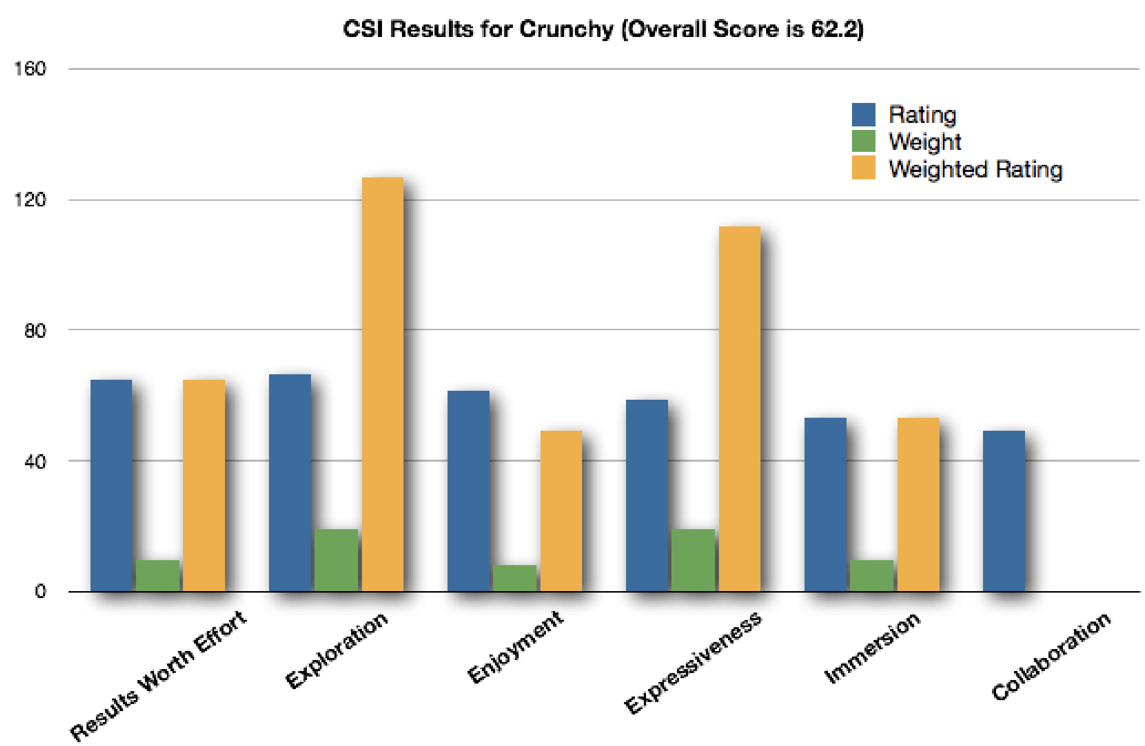


Figure 20: Crunchy Slider's CSI scores for the 6 orthogonal factors for evaluating creativity support tools

6.4.3.3 Immenseness

The rating of the immenseness of the parameter space was significantly different between interfaces: Creamy Sliders: 3.14 (SE 0.34) and Crunchy Sliders: 2.35 (SE 0.34). There were no significant block effects or interaction effects between the interface and block. A t-test of the interaction effect revealed that the rating for Creamy Sliders in Block 2 was significantly higher than the ratings for the other three conditions. There was no significant pairwise difference between the other three conditions (Creamy Sliders in Block 1, Crunchy Sliders in Block 1, and Crunchy Sliders in Block 2).

The difference in Immenseness ratings between interfaces suggests participants felt the size of the parameter space offered by Creamy Sliders was much more immense compared to Crunchy Sliders. The t-test result indicates that the immense rating for Creamy Sliders in Block 2 was significantly higher, likely due to the expansion of the accessible parameter space. All interviewed participants (NP_{55-58}) discussed the large number of options, possibilities, or choices offered by Creamy Sliders.

6.4.3.4 Satisfaction with Creative Process

The ratings of SatCP were significantly different between interfaces: Creamy Sliders: 7.51 (SE 0.34) and Crunchy Sliders: 6.47 (SE 0.34). There was an effect between the interface and block. I further investigated this phenomenon by performing a t-test on this interaction effect and found that the SatCP rating for Crunchy Sliders in Block 2 was significantly lower than the other three conditions. There were no other significant differences between the other conditions (Creamy Sliders in Block 1, Creamy Sliders in Block 2, and Crunchy Sliders in Block 1).

The difference in SatCP ratings between interfaces shows participants felt more satisfied with their creative process using Creamy Sliders compared to Crunchy Sliders. The t-test result that the SatCP rating for Crunchy Sliders in Block 2 was significantly lower could suggest that a reduction of the accessible parameter space reduces user satisfaction were with their creative process. ($NP_{57,m}$) stated "I felt I was able to be more creative with Creamy."

6.4.3.5 Satisfaction with Final Design and Suitability

The rating of SatFP was not significantly different between interfaces or blocks. A t-test reveals that the SatFP rating for Crunchy Sliders in Block 2 was significantly lower than the other three conditions. There was no significant difference between the other three conditions (Creamy Sliders in Block 1, Creamy Sliders in Block 2, and Crunchy Sliders in Block 1).

Similarly, there were no significant differences between interfaces and block conditions for the rating of how suitable a participant's final design was to the task prompt. A t-test revealed that the only significant difference between interface and block conditions was that Crunchy Slider in Block 2 was rated significantly lower then Creamy Sliders in Block 1. It should be noted that this rating was not significantly different than Crunchy Sliders in Block 1 (i.e., no ordering effect) and Creamy Sliders in Block 2 (i.e., no effect from interface).

While the t-test shows that the SatFP and Suitability ratings for Crunchy Sliders in Block 2 were significantly lower, in either case, there were no overall significant differences between the interfaces and blocks. This could suggest that without the reduction in parameter space, participants might feel the same level of satisfaction and suitability of their final design regardless of the interface they used. $NP_{57,m}$ stated:

> "[With both Creamy and Crunchy] I think I [reached the goal] with my sense of cool. They are nice looking quality images. [...] I felt I had the best image that I created. I was satisfied with my work. I thought the judges would like it."

Even though $NP_{57,m}$ preferred Creamy Sliders, he still indicated that both of the images he created were suitable for the task, were 'cool' to him, and that the judges might like them as well. Therefore, for H_7, I cannot reject the null hypothesis: $H_{7,null}$: *Participants* ***will not be more/less*** *satisfied with their final designs from Creamy Sliders versus Crunchy Sliders.*

6.4.3.6 Care

There was no significant differences between interface and block conditions for the rating of how much care a user put into their own design. The average 'care' users reported was 8.1 (SD 1.80). A t-test did not yield any new information. This suggests that overall most participants took a similar level of care in completing the task, which indicates the results presented thus far are reliable.

6.4.3.7 Other Hypothesis

I collected data on participants' satisfaction with their creative process and ratings of exploration from the CSI Scores, but did not measure their satisfaction with the exploration of the design space directly. Creamy Sliders was rated higher in both cases. This suggests that I can reject the null hypothesis, but in favor of the inverse relationship. H_{2a} : *Participants were **less** satisfied with how adequately they explored the design space using Crunchy over Creamy.*

6.4.4 Exploration in Mechanical Turk Studies

I have separated the analysis of exploration here because "exploration" is the leading research topic of this dissertation and because the following analysis required careful detailing. Exploration was measured in three ways: saved thumbnails, self-reported, and logged interface data. Each of these measures are covered in the sections below. Following the analysis of the exploration measures, I have also included correlations between Exploration and other measures collected in my study.

6.4.4.1 Exploration: Saved Thumbnails

There was no significant difference between interface or block conditions for the number of saved thumbnails. A t-test did not show any interaction effects between conditions either. This suggests that the interface did not have an impact on the number of saved thumbnails. Therefore, for H_6, I cannot reject the null hypothesis: $H_{6,null}$: *Participants **did not** save more thumbnails (i.e., more unique results) with Crunchy Sliders than Creamy Sliders.*

Participants could save thumbnails of their work during the task. These serve as milestones of exploration. These indicate that the state of the image was important to the participate, but the thumbnails alone do not tell me why it was important. Later in this chapter, I discuss the decision making processes participants went through to decide when to save thumbnails, and how this reflects parameter space exploration.

Table 7: Reported perceived exploration across interfaces and blocks ($p<0.05$ in **bold**)

Group	Block	Interface	Mean(%)	Std. Error
W & X	1 & 2	Creamy	**41.47**	3.79
		Crunchy	**48.10**	3.79
W	1	Creamy	44.18	5.26
	2	Crunchy	49.57	5.46
X	1	Crunchy	**46.61**	5.26
	2	Creamy	**38.77**	5.46

6.4.4.2 Exploration: Self Reported

The reported perceived exploration of the parameter space was significantly different between interface conditions: Creamy Sliders: 41.47% (SE 3.78) and Crunchy Sliders: 48.09% (SE 3.78). There were no significant block effects or interaction effects between the interface and block conditions. Further investigation using a t-test revealed that the condition for Group X (Crunchy, then Creamy) was only one pair of conditions that had a significant difference: Crunchy Sliders in Block 1: 49.57% (SE 5.46) and Creamy Sliders in Block 2: 38.77% (SE 5.46), see Table 7. The differences between interfaces for Group W were not significant. However, the mean differences, for both Groups, were in the same direction (with Crunchy Sliders having a higher reported percentage of perceived exploration).

These results suggest that participants did feel as though they were exploring more of the parameter space when using Crunchy versus when they were using Creamy. The t-test results suggest that using Crunchy first may have some effects on users' perceptions of the Creamy interface. Therefore, for H_{10}, I can reject the null hypothesis in favor of the alternative: H_{10} : Participants perceived having explored a higher percentage of the design space with Crunchy Sliders than with Creamy Sliders.

6.4.4.3 Exploration: Interaction Logs

I took two different approaches to analyzing the log data of the user's exploration of the parameter space. The first was a ratio of visited sliders values to the total possible slider values, which is the percentage of the parameter space explored by the participant (Actual Exploration).When calculating the percentages, the numerator in the Creamy Sliders condition was higher than than the numerator in the Crunchy Sliders condition. The second approach was to compare the distribution of slider values for searched using Creamy Sliders versus Crunchy Sliders.

For the ratio comparison, I calculated an approximation of the percentage of the parameter space users actually explored. These are approximations as they account only for the percentage of each unique slider position that were explored. The numerator here also differed between Creamy Sliders and Crunchy Sliders. This is different from the percentage of the parameter space that is exponential to the number of sliders and slider value accessible. Even though it is an approximation, the differences were significant enough to warrant their inclusion here.

There was a significant block effect ($p<0.01$) where participants explored more of the parameter space (in either Creamy or Crunchy Sliders) in Block 1 (52.22%, SE 1.85) than in Block 2 (43.29%, SE 1.85). There was also an interface effect ($p<0.01$) where participants explored more of the parameter space using Crunchy Sliders (71.24%, SE 1.85) than with Creamy Sliders (24.27%, SE 1.85). There were no other interaction effects between other conditions. No correlation was found between Actual Exploration and the Perceived Exploration in the previous section.

Finding that participants explored more using Crunchy was expected as the parameter space was dramatically reduced in size to allow users to explore more of the space. The block effect does show that in the first Block, participants were likely getting acquainted with sliders of either interface. By the second block, they did not need to explore what the sliders did as much. Not having an interaction effect between the interface and block conditions suggests that reduction in the accessible parameter space could have little to no impact on the actual exploration done.

For the comparison of distributions, I started with simple descriptive statistics (see Table 8) for the values for each parameter (i.e., sliders) in Creamy Sliders and Crunchy Sliders. A RM-ANOVA of the standard errors showed a significantly higher ($p<0.01$) SE for Crunchy Sliders (2.21) versus Creamy Sliders (1.83). While this is not a traditional way for using standard errors, it does demonstrate a high variance in the distribution of the data. A high variance indicates a more even distribution across the values of the parameters. Therefore, I reject the null hypothesis in favor of the alternative: H_1 : *Participants covered a more even distribution of the parameter space with Crunchy than with Creamy*

The analysis above is an indirect measurement of the distribution (i.e., it is an analysis of the descriptive statistics derived from the dataset). To better understand if the distribution of exploration did differ between Creamy Slider and Crunchy Sliders, I started with an ANOVA of the values for each parameter in Creamy Sliders and Crunchy Slider. One of the limitations of applying an ANOVA is that is assumes my data has a normal distribution. Some parameters, like Brightness or Saturation, are pixel manipulations that can be positive or negative. These parameters start at

Table 8: Comparison between the mean slider values for each parameter between interfaces (see Table 9 for significance tests)

		Creamy		Crunchy	
Slider Name	Range of Values	Mean	SE	Mean	SE
Brightness	[-100,100]	2.37	2.17	0.94	2.65
Contrast	[-100,100]	3.32	2.00	4.86	2.58
Exposure	[-100,100]	-4.92	2.38	-2.49	2.86
Saturation	[-100,100]	11.88	2.76	8.74	3.27
Cross Process	[-100,100]	2.65	2.47	5.02	3.04
Hue Rotation	[0,100]	43.32	1.07	47.19	1.44
Clipping	[0,100]	33.78	1.63	36.73	1.95
Noise	[0,100]	26.70	1.84	30.01	2.04
Sharpen	[0,100]	39.89	1.67	45.72	1.93
Blur	[0,20]	6.22	0.30	6.69	0.34

zero and can be moved in the positive or negative direction (See Table 8. These values could have a normal distributions that starts in the middle, but this is not a guarantee. Other parameters, like Noise or Blur, are more additive manipulations where the pixel values changed based on neighboring pixels. Additive manipulations can only be added to the base image, therefore the pixels only move in a positive direction. Additive manipulations are a bit more destructive and tend to deviate the most from the base image. These parameters may not have a normal distribution either. To ensure reliability in my results, I compared the distributions using analyses that were non-parametric, where each method does not assume my data has a normal distribution. The Kruskal-Wallis One-Way ANOVA [27], the Kolmogorov-Smirnov Test (KS Test) [51] [26], and Statistical Energy goodness-of-fit test [2] are three methods for determining if two samples come from the same distribution, i.e., if they come from the same distribution, they are likely to have similar distributions. The mean comparisons are in Table 8 and the results of the significance tests are in Table 9.

Table 9: Different tests for significance when comparing the distributions of slider values between interfaces

	ANOVA (P-Value)	Kruskal-Wallis ANOVA (Prob $>\chi^2$)	Statistical Energy (P-Value)	KS Test (P-Value)
Brightness	no sig.	no sig.	no sig.	p<0.001
Contrast	no sig.	no sig.	no sig.	p<0.001
Exposure	no sig.	no sig.	p<0.05	p<0.001
Saturation	no sig.	no sig.	p<0.05	p<0.001
Cross Process	no sig.	no sig.	p<0.05	p<0.001
Hue Rotation	p<0.01	no sig. *0.097	p<0.01	p<0.001
Clipping	p<0.05	no sig.	p<0.01	p<0.001
Noise	p<0.05	no sig.	p<0.01	p<0.001
Sharpen	p<0.01	0.007	p<0.01	p<0.001
Blur	no sig.	no sig.	p<0.01	p<0.001

The KS Test reveals that the distribution difference for between each of the sliders was significant. Crunchy Sliders was derived from the distribution of Creamy Sliders, but during the data collections process, this difference is more pronounced as the distribution Creamy Sliders is going to be less dense than Creamy Sliders, where users can select the parameters that are not available in Crunchy. The Statistical Energy Test is probably the outlier of other test in terms of methodology. Unlike χ^2-tests and other methods that rely on binning, the Statistical Energy Test does not use an arbitrary method of binning the data [2]. This test also reveals significant differences between the parameter explorations of Creamy Sliders and Crunchy Sliders. Brightness and Contrast however, were not found to be significantly different. This suggests that in both interfaces, the values for Brightness and Contrast were distributed equivalently. These two parameters are two of the most prevalent and familiar image properties (e.g., monitors and televisions allow users to adjust these properties). In both interfaces, these two parameters were the top-most parameters.

The ANOVA and Kruskal-Wallis ANOVA painted much different pictures. These ANOVAs suggests overlaps between the distributions for each slider. Kruskal-Wallis ANOVA shows only the difference in Sharpen was significant (and is supported by the regular ANOVA). Kruskal-Wallis ANOVA did not show significant for Hue Rotation, but was noticeably low. The regular ANOVA shows that most of the differences between additive parameters were significant. As I stated earlier, these parameters are a bit destructive, especially at higher values. The mean values (assuming normal distribution) are actually lower than 50 (the median of the range), suggesting that the higher ranges were not fully explored. Compare these means to the means of brightness, contrast, etc., where the mean value from the data is close to the median value for the range. The mean values for additive parameter are higher for Crunchy Sliders. Crunchy Sliders forced participants to explore higher values (i.e., more destructive) parameters. This in turn caused the medians to be higher consistently and significantly. This suggests that when participants used Creamy Sliders, if they had an idea of how much change occurs to the image when moving parameter, they might stop moving the slider before the image gets too 'messed up.'

I am cautious to say that the analysis and results suggest anything more. The issue with multivariate data is the difficulty in comprehending the analytic distribution. My data has many dimensions (i.e., parameters), and therefore claims like "X had a larger distribution than Y" or "Y has a denser distribution than X" are not easy to extract or infer from the data I collected in my study. To my knowledge at the time this dissertation was written, there is no methodology that can accurately represent and compare the analytic distribution of non-parametric, multivariate data, at least with a high enough reliability and confidence of making the 'X' and 'Y' comparative claims mentioned earlier in this paragraph.

6.4.4.4 Exploration: Linear Relationships

Using a combination of linear regressions and ANOVAs, I analyzed the full-factorial relationships between perceived exploration, actual exploration, block, and interface.

Actual exploration (percentage of parameter space explored by the user) was inversely correlated ($R \approx 0.26$) with the user's reported satisfaction with their creative process (SatCP). Since $0.2 < R < 0.5$, the effect size between the actual exploration and SatCP is considered small. There were no interaction effects from any of the other conditions and SatCP. For H_8, I can reject the null hypothesis, but in favor of the inverse alternative: H_8 : *Participants with a higher exploration (percentage of space covered) had a* ***lower*** *user satisfaction with the design process.*

Actual exploration was not correlated directly with the user's reported satisfaction with their final design (SatFD). There was however an interaction effect between block, interface, and actual exploration. While using Crunchy Sliders in Block 2, actual exploration was inversely correlated ($R \approx 0.37$) with the user's reported satisfaction with their final design (SatFD). Using Cohen's guidelines, since $0.2 < R < 0.5$, the effect size between the SatFD and the actual exploration of Crunchy Slider of Block 2 is considered small. For the other block and interface conditions, R was less than 0.2, indicated that there was no effect between conditions. These findings are in agreement with previous findings where the condition of Crunchy Sliders in Block 2 was significantly different from the rest of the conditions. These findings were not strong, so for H_9, I cannot reject the null hypothesis: H_9 : *Participants with a higher exploration (percentage of space covered)* ***did not*** *have a higher/lower user satisfaction with the final product.*

Perceived exploration (self-reported percentage explored) correlated (R≈0.20) with the user's reported satisfaction with their final design (SatFD). Using Cohen's guidelines, since 0.2<R<0.5, the effect size between the perceived exploration and SatFD is considered small. There were no interaction effects from any of the other conditions. This suggests that when participants felt their exploration of the parameter space was high, their satisfaction with their final product would also be high. What is not clear is whether their SatFD caused their perceived exploration to increase or visa versa. Perceived exploration was captured in a post-questionnaire. The question about their perceived exploration was below the questions concerning satisfaction.

6.4.5 Maximizers and Satisficers

In the following section, I use a self-report metric of personality categorization to gain a deeper understanding of the results from Mechanical Turk that I have presented thus far. The analysis up to this point involved independent variables that I controlled in my study. The classification of participants as Maximizers or Satisfiers, is an extraneous variable. Since I was not able to control for this variable experimentally (and because some researchers criticize the use of self-report metrics for predicting the experimental results [15]), I am presenting this analysis separately. My analysis uses the maximization categorization as a tool for guiding and understanding the experimental results and is used in conjunction with other metrics.

The first step was to divide participants into two categories: Maximizers (N=37) and Satisficers (N=17). Participants were placed into categories based on the Maximization Scale Questionnaire, which is composed of statements of agreement/disagreement related to high standards, difficulty in making decisions, and searching for alternatives. Building upon the analysis and results presented in the previous section, the results reported below incorporate this metric.

6.4.5.1 Saved Thumbnails

There was no overall significant difference between the number of saved thumbnails for Maximizers versus Satisficers. A t-test revealed that Maximizers would on average save more thumbnails with the Crunchy Sliders interface (6.90, SE 0.78) compared to the Creamy Sliders interface (5.13, SE 0.78). No significant differences were found for Satisficers. There were no significant interaction effects between the remaining combinations of block, interface and maximization category.

All of my interviewed participants (all Maximizers) stated various reasons for choosing to save thumbnails:

$\mathrm{NP}_{55,HM}$: "Usually saved when [I] wanted to change the overall color."

$\mathrm{NP}_{56,LM}$: "Wild changes and may not be able to get back. [...] Like saving before an action."

$\mathrm{NP}_{57,LM}$: "Trying to make the best option I could create. Might not be the same as before, but I could go back. [...] one click and you really change an image. Not that one is bad, one is a better option."

$\text{NP}_{58,HM}$: "When I was about to do something that might screw up the image, when I was trying something in a different direction."

As $\text{NP}_{57,LM}$ describes his experience, it is possible that saving thumbnails relates to finding the best option available. My intention in allowing users to create thumbnails was for users to save ideas that they could select from on the design submission page. Instead, the interviewed participants described an experience where they used the thumbnails as a substitute for undoing changes to the image. This is similar to what Latulipe et al. found in using exploration areas as memory cues for re-finding [32]. $\text{NP}_{56,LM}$ stated that she "start[ed] using thumbnails to get past not having an undo option." She further stated that with Creamy's level of control: "it's difficult to go back. [...] Getting sliders back to where it was, chances are slim it will go right back there." It is possible that with the limited number of slider levels (i.e., 5) in Crunchy Sliders, participants may have felt it was easier to re-find or return to slider configurations they had before.

6.4.5.2 Total Task Time

There was no overall significant difference between the total task time for Maximizers and the total task time for Satisficers. A t-test revealed some interaction effects between Maximizers and Block. The average task time for Maximizers in Block 1 (4.71 mins, SE 0.13) was higher than the average task time for Maximizers in Block 2 (4.25 mins, SE 0.13) or Satisficers in Block 2 (4.20 mins, SE 0.20). I also did a t-test for the interaction effects across the blocks, interfaces, and maximization categories. This t-test revealed the following deeper relationships (see Tables 10 and 11):

Table 10: Total task time per condition: interfaces, blocks, and maximization category (significant differences are reported in Table 11)

Block	Category	Interface	N	Mean (mins)	Std. Error
1	Maximizers	Creamy	21	4.67	0.18
		Crunchy	16	4.76	0.20
	Satisficers	Creamy	7	4.33	0.31
		Crunchy	10	4.84	0.26
2	Maximizers	Creamy	16	4.29	0.20
		Crunchy	21	4.21	0.26
	Satisficers	Creamy	10	4.11	0.18
		Crunchy	7	4.29	0.31

- The average time Maximizers spent using Creamy Sliders in Block 1 was significantly higher than the average time Maximizers spent using Crunchy Slider in Block 2.
- The average time Maximizers spent using Crunchy Sliders in Block 1 was significantly higher than the average time Maximizers spent using Creamy Sliders in Block 2 and Crunchy Sliders in Block 2.
- The average time Satisficers spent using Crunchy Sliders in Block 1 was significantly higher than the average time Maximizers spent using Crunchy Sliders in Block 2 and the average time Satisficers used Creamy Sliders in Block 2.

The items above fall in line with the results for Total Task Time and Blocks presented in Section 6.4.3 Interfaces and Ordering in Mechanical Turk Studies. In Table 11, most of the block effects seem to be occurring within Maximizers, and less so within Satisficers, and between Maximizers and Satisficers. This again is likely due to learning effects. Since there was no effect in the Satisficer condition, for $H_{S.2}$, I cannot reject the null hypothesis: $H_{S.2,null}$: *Satisficers spend* ***relatively the same*** *time with Creamy Sliders and Crunchy Sliders.*

Table 11: Differences of task time (Block 1 - Block 2) between interfaces, blocks, and maximization category. Standard Errors varied [0.18 - 0.40] and for clarity are not included. Comparison within Blocks had no significant differences, and has been excluded for brevity. The diagonal is the ordering effects

			Block 2			
			Maximizers		Satisficers	
			Creamy	Crunchy	Creamy	Crunchy
Block 1	Maximizers	Creamy	0.39	**0.46**†	0.56	0.37
		Crunchy	**0.48**	**0.56**	**0.66**	0.47
	Satisficers	Creamy	0.04	0.12	0.22	0.04
		Crunchy	0.55	**0.63**	**0.73**†	0.54

p<0.05 †p<0.01

Since most Maximizers desire to "explore all the potential choices that are available" (NP$_{55,HM}$), it is not surprising that they chose to use the entire time allotted. There could be several explanations for why Maximizers spent more or less time using Crunchy Sliders. NP$_{56,LM}$, who was on the borderline between Maximizer and Satisficers, stated that when it came time to submit a thumbnail she felt "I could have spent more time with Creamy." and that with Crunchy: "I've gone through all possible options." She also felt: "I was content with the image, nothing else I could do to improve it." This suggests that participants may see the task in two parts: 1) learning and using the interface and 2) completing the task. In the case of NP$_{56,LM}$, she felt nothing could improve on the image, but she could potentially spend more time using Creamy. This suggests that Satisficing occurred where the number of options selected was sufficient to get an image that did not need improvement, but with the knowledge that other images could exist, as NP$_{57,LM}$ states: "Not that one is bad, one is a better option."

Table 12: Distribution of TimeOut and Submit events for Maximizers and Satisficers

	TimeOuts	Submits	Total
Maximizers	38	36	74
Satisficers	17	17	34
Total	55	53	108

6.4.6 Timing Out

The number of participants that timed out (used all 5 minutes) versus the participants who clicked the submit button before 5 minutes were relatively equal for Maximizers and for Satisficers (see Table 12). Therefore, for $H_{M.4}$, I cannot reject the null hypothesis: $H_{M.4,null}$: *Overall, Maximizers* ***did not always use*** *all the allotted time to complete the task.*

6.4.6.1 CSI Scores

There was no overall significant difference between the CSI scores for Maximizers versus Satisficers. A t-test for interaction effects between Maximization category and block revealed that the average Maximizer CSI score in Block 1 (70.99, SE 3.30) was higher than the Maximizer CSI score in Block 2 (62.82, SE 3.30). No significant effects were found within Satisficers or between Maximizers and Satisficers. A t-test for interaction effects between the Maximization category and interface revealed two significant differences (see Tables 13 and 14):

- Maximizers gave Creamy Sliders a higher CSI score than Crunchy Sliders.
- There was no significant difference between the CSI scores Satisficers gave Creamy Sliders versus Crunchy Sliders. This suggests that Satisficers saw the creativity support offered by Creamy Sliders as equivalent to Crunchy Sliders.

- There was no significant difference between the CSI scores Satisficers gave Creamy Sliders versus the CSI score Maximizers gave Creamy Sliders. This indicates Satisficers and Maximizers had an equivalent view of creativity support for Creamy Sliders.

A t-test for the interaction effects across the blocks, interfaces, and maximization categories revealed that, on average, the CSI score Maximizers gave Crunchy Sliders was significantly lower score ($p<0.01$) in Block 2 than the average CSI score Maximizers gave Crunchy Sliders in Block 1 as well as the average CSI score Maximizers and Satisficers gave Creamy Sliders in Block 1 and Block 2. There was no significant difference between the average CSI score Maximizers gave Crunchy Sliders in Block 2 and the CSI scores Satisficers gave Crunchy sliders in Block 1 and Block 2.

The block effects from Section 6.4.3 Interfaces and Ordering in Mechanical Turk Studies can still be seen in these results. The more interesting result is that there was no significant difference within Satisficers. This could suggest that they perceived the creativity support offered by both systems to be similar. On the other hand, Maximizers had dramatically different scores on average for Creamy Sliders versus Crunchy Sliders. This is likely due to the reduction in parameter space for Crunchy Sliders, reducing the number of options available, and thereby, users' perceived ability to be creative with that interface.

6.4.6.2 Perceived Exploration

The reported perceived exploration (PE) of the parameter space was not significantly different between Maximizers and Satisficers. A t-test revealed there were

Table 13: CSI Scores of Creamy/Crunchy interfaces by Maximizers/Satisficers (significant differences are reported in Table 14)

Category	Interface	Mean (max 100)	Std. Error
Maximizers	Creamy	71.40	3.30
	Crunchy	74.58	4.90
Satisficers	Creamy	62.42	3.30
	Crunchy	64.68	4.90

Table 14: Differences of CSI Scores between interfaces and maximization category

	CSI_{Creamy} - $CSI_{Crunchy}$	SE
Maximizers	**8.98**	**3.55**
Satisficers	9.91	5.91
		p<0.05

some interaction effects with the interface condition, see Tables 15 and 16. The PE that Maximizers reported for Crunchy Sliders was higher than the PE Maximizers reported for Creamy Sliders. There were no significant differences between the PE Satisficers reported between Crunchy Sliders and Creamy Sliders, which suggests they saw themselves exploring relatively the same number of options in both. Other t-tests revealed no significant interaction effects between block, interface, and Maximizers and Satisficers.

The only significant difference was with Maximizers, who perceived themselves as having explored more the parameter space using Crunchy Sliders versus when they used Creamy Sliders. Based on the literature, Maximizers are more likely to be in tune with what options are available and what options are not available. In fact, $MP_{02,m}$ stated: "With [Crunchy Sliders], I felt that, eventually, all or most of the combinations would be created at some point or another." Here $MP_{02,m}$ noticed the reduction in the search space would make trying out the combinations more manageable or achievable. The difference within Satisfiers was not statistically significantly. Therefore, these

Table 15: Reported perceived exploration between interfaces, and Maximizers and Satisficers (significant differences are reported in Table 16)

Category	Interface	Mean (%)	Std. Error
Maximizers	Creamy	42.56%	4.65
	Crunchy	50.54%	4.65
Satisficers	Creamy	41.26%	6.90
	Crunchy	43.81%	6.90

Table 16: Differences of perceived exploration (PE) between interfaces and maximization category

	PE_{Creamy} - $PE_{Crunchy}$	SE
Maximizers	**-7.98**	**3.70**
Satisficers	-2.54	5.49

p<0.05

results suggest that Maximizers may be more likely to see the explorability differences. Satisficers, however, may have been less likely to notice the difference or they felt the exploration was less meaningful. In the literature, Satisficers were not inclined to think about other options that were not readily available in front of them.

6.4.6.3 Actual Exploration

There were no interaction effects between Actual Exploration, and Maximizers and Satisficers. This suggests that both sets of participants explored equal amounts of options. Given the short task time, it is likely that Maximizers were not able to continue exploring. As I stated earlier, there was no significant difference of task time between Maximizers and Satisficers. This suggests that given a limited amount of time, Maximizers and Satisficers may do the same amount of exploration.

6.4.6.4 Satisfaction with Creative Process

The average rating of Satisfaction with their Creative Process (SatCP) was not significantly different between Maximizers and Satisficers. A t-test revealed one interaction effect between blocks. Using either interface, Satisficers in Block 1 were more satisfied with there creative process than Maximizers in Block 2 ($\Delta 1.51$, SE 0.73). A t-test also revealed some interaction effects between interfaces. Maximizers were less satisfied with their creative process using Crunchy Sliders versus Creamy Sliders ($\Delta 1.11$, SE 0.48). A t-test also revealed one interaction effect between blocks, interfaces, and Maximizers and Satisficers. Maximizers gave Crunchy Sliders a significantly lower rating ($p<0.01$) in Block 2 than the average rating Maximizers gave Crunchy Sliders in Block 1 as well as the average rating Maximizers and Satisficers gave Creamy Sliders in Block 1 and Block 2. There were no other significant differences, see Tables 17 and 18.

The results of the SatCP rating are in line with the literature on how Maximizers and Satisficers differ in satisfaction, with Satisficers being on average more satisfied then their Maximizer counterparts. These results also show how Maximizers were less satisfied when there was a reduction in the parameter space using Crunchy. This is also in line with the literature that indicates Maximizers prefer to have more options.

6.4.6.5 Immenseness

The rating of the immenseness of the parameter space was not significantly different between Maximizers and Satisficers. A t-test revealed one interaction effect between blocks. The rating of immenseness was lower in Block 1 than Block 2 for

Table 17: Reported SatCP ratings between interfaces, and Maximizers and Satisficers (significant differences are reported in Table 18)

Category	Interface	Mean (max 10)	Std. Error
Maximizers	Creamy	7.36	0.41
	Crunchy	6.25	0.41
Satisficers	Creamy	8.06	0.61
	Crunchy	7.08	0.61

Table 18: Differences of SatCP rating between interfaces, and Maximizers and Satisficers

	SatCP_{Creamy} - $\text{SatCP}_{Crunchy}$	SE
Maximizers	**1.11**	**0.49**
Satisficers	0.98	.073
		p<0.05

Satisficers (Δ-1.46, SE 0.64). Using a t-test, no significant interaction effects were found between interfaces, and Maximizers and Satisficers. A t-test revealed one interaction effect between blocks, interfaces, and Maximizers and Satisficers. Satisficers gave Creamy Sliders a significantly lower rating in Block 2 than the average rating both Maximizers and Satisficers gave either interface in Block 1, as well as the average rating Maximizers gave Crunchy Sliders in Block 2 (see Table 19). There were no other significant differences.

These results suggest some interface ordering effects for Satisficers. When Satisficers used Crunchy Sliders in Block 1, they were on average more likely to rate it as more immense compared to other conditions. The lack of significant interaction effects, suggests, that without the ordering, participants may not have noticed how much more immense the parameter space was for Creamy sliders.

Table 19: Differences of immenseness rating between interfaces, blocks, and Maximization category. Standard Errors varied [0.57 - 1.23] and for clarity are not included

			Block 1				Block 2			
			Maximizers		Satisficers		Maximizers		Satisficers	
			Creamy	Crunchy	Creamy	Crunchy	Creamy	Crunchy	Creamy	Crunchy
Block 1	Maximizers	Creamy	0	-0.02	0.81	0.17	-0.58	0.71	**-2.03**	0.10
		Crunchy	0.02	0	0.83	0.19	-0.56	0.70	**-2.01**	0.12
	Satisficers	Creamy	-0.81	-0.83	0	-0.64	-1.39	-0.10	**-2.84**	-0.71
		Crunchy	-0.17	-0.19	0.64	0	-0.75	0.55	**-2.20**†	-0.07
Block 2	Maximizers	Creamy	0.58	0.56	1.39	0.75	0	1.30	-1.45	0.68
		Crunchy	-0.71	-0.70	0.10	-0.55	-1.30	0	**-2.75**	-0.62
	Satisficers	Creamy	**2.03**	**2.01**	**2.84**	**2.20**†	1.45	**2.75**	0	2.13
		Crunchy	-0.10	-0.12	0.71	0.07	-0.68	0.62	-2.13	0

$p<0.05$ $^{\dagger}p<0.01$

6.4.6.6 Satisfaction with Final Design

The rating of a participants' Satisfaction with their Final Design (SatFP) was not significantly different between Maximizers and Satisficers. A t-test revealed no significant interaction effects between blocks, and Maximizers and Satisficers. A t-test revealed no significant interaction effects between interfaces, and Maximizers and Satisficers. A t-test between all three variables revealed two interaction effects. Maximizers were more satisfied with their final design when they used Crunchy in Block 1 than they were when they used Crunchy in Block 2 (Δ1.91, SE 0.78, $p<0.01$). Satisficers who used Creamy in Block 1 were on average more satisfied than Maximizers who used Crunchy in Block2 (Δ2.38, SE 01.02, $p<0.05$). There were no other significant differences. These results suggest that without the ordering effects of using Creamy first, Maximizer may be just as satisfied using Creamy as they were with Crunchy.

6.4.6.7 Suitability

There was no significant difference between Maximizers and Satisficers for the rating of how suitable their final design was to the task prompt. A t-test revealed only one interaction effect between blocks, interfaces, and Maximizers and Satisficers: Maximizers rated their Final Design created using Crunchy Sliders in Block 1 more suitable to the task prompt than their Final Design created using Crunchy Sliders in Block 2 (Δ1.85, SE 0.80). No other significant interaction effects were found between blocks, interfaces, and Maximizers and Satisficers.

The Suitability results relate to earlier discussions where participants stated that both images were suitable to the task and that the judges would like them. The results suggest an ordering effect when Creamy Sliders was used first, resulting in a decrease in the participant's rating of suitability of the image created using Crunchy Sliders. This is compared to using Crunchy first (no ordering effects) where participants were not aware that there is a different interface that expands the size of the search space. There was no significant difference within Satisficers for the rating of Suitability. This could suggest that in this parameter space, the interface was not a significant influence on Satisficers' ratings of Suitability.

6.4.6.8 Other Factors

The was no significant difference between interfaces and blocks conditions for the rating of how much care a user put into their design. A t-test revealed no significant interaction effects between blocks, and Maximizers and Satisficers, no significant interaction effects between interfaces, and Maximizers and Satisficers, and no significant interaction effects between blocks, interfaces, and Maximizers and Satisficers.

I correlated the Maximization scores with participants' satisfaction with their creative process (SatCP) and final designs (SatFD) while using Creamy and Crunchy Sliders. There were no significant effects found (R<0.2). Therefore, for $H_{M.2}$ and $H_{M.3}$ (these include only Maximizers), I cannot reject the null hypotheses: $H_{M.2}$: *Maximization scores* ***did not*** *correlate with Creamy satisfaction* and $H_{M.3}$: *Maximization scores* ***did not*** *correlate with Crunchy satisfaction.* Additionally this means that, for H_3 (all participants), I cannot reject the null hypothesis: H_3*: Participant's Maximization Scale Scores* ***did not*** *correlate with their satisfaction with the percentage of the design space explored using Creamy Sliders.*

I correlated the Maximization scores with participants' satisfaction with the amount of time used and with the perceived exploration. There were no significant effects found (R<0.2). Therefore, for H_4 and H_5, I cannot reject the null hypotheses: H_4 : *Participant's Maximization Scale Scores* ***did not*** *correlate with the amount of allotted time they used* and H_5 : *Participant's Maximization Scale Scores* ***did not*** *correlate with the perceived percentage of design space explored.*

I collected data on each participant's experience with graphic design software and found no correlation with other metrics presented earlier (e.g., SatCP, SatFD, Maximization scale, exploration, CSI scores, etc). The question was on a four point scale: Little/None (N=13), Beginner (N=18), Intermediate(N=19), and Advanced(N=4). There were clearly not enough Advanced participants, which may have contributed a large amount of noise in my data, resulting in a lack of statistical significance.

6.4.7 Omitted Analysis

At the start of this chapter, I indicated that I would look at analyzing the data based on the Contextualized Maximization Scale (CMS), where the questions were in the same domain as the task. I used the scores from the CMS to re-categorize participants as Maximizers and Satisficers. Using a Contingency Analysis for comparing nominal data, I compared the categories based on the original Maximizations scale with the categories based on the CMS. The χ^2 for the Pearson test was 0.012 (Prob$>\chi^2$ = 0.91) and indicates that they are not significantly different. Because the scales were deemed similar, I did not perform the analysis using the CMS. This does not indicate that that CMS is not valuable. The use of a CMS requires valida-

Table 20: Distribution of interface preferences for Maximizers and Satisficers

	Creamy	Crunchy	Total
Maximizers	26	11	37
Satisficers	14	3	17
Total	40	14	54

tion through additional studies. However, by supplementing the Maximization scale with a CMS within the experimental domain, I believe it helps to ensure the reliability of categorizing users as Maximizers or Satisficers, especially when utilizing the Maximization scale (created for general use) in domain specific research studies.

I also indicated in Chapter 4 that I would look at analyzing my data based on tertiary categories. However, when participant were re-categorized using the tertiary categories (Maximizers, Negotiators, Satisficers), there were not enough satisficers (N=2). Any analysis performed would have been biased and any conclusions draw would have very limited applicability. Later in this Chapter, I analyze the interviews using sub-groups of Maximizers, which is similar to the proposed analysis but specific to the interview data.

6.4.8 Preference

The distribution of interface preferences for Maximizers and Satisficers are shown in Table 20. Overall 70% of Maximizers preferred Creamy Sliders. Therefore, for $H_{M.1}$:, I can reject the null hypothesis in favor of the alternative: $H_{M.1}$: *Overall, Maximizers preferred Creamy Sliders to Crunchy Sliders.* Overall 82% of Satisficers also preferred Creamy Sliders. Therefore, for $H_{S.1}$:, I can reject the null hypothesis, but in favor of the inverse relationship: $H_{S.1}$: *Overall, Satisficers preferred* ***Creamy Sliders to Crunchy Sliders.***

To find out which conditions were most likely to contribute to a Participant's preference of Crunchy Sliders versus Creamy Sliders, I performed a Nominal Logistic Fit for interface preference using the various data from this study. There were no variables or combinations of variables that could adequately predict the preference of participants. This suggest that preference of interface was highly subjective. After a thematic analysis, I found most of the preferences fell into two categories: creative output and creative process (which tied heavily with the interaction). Within these categories, I found four themes within the free-response questions for why they preferred Crunchy Sliders. The theme center on Crunchy Sliders, because Creamy Sliders is similar to the baseline model of interaction, as it follows a more familiar interaction style with sliders values.

Creative Process: Parameter Variations: Several participants said that working with the parameters using Crunchy Sliders was easier than Creamy Sliders. With Creamy Sliders, there are many values for each parameter that the user has to work through.

$MP_{08,M}$ "[made] work with the color variations that were presented more easily." $MP_{37,S}$ "I was able to compare and change options more easily and experiment when I had set values that the Crunchy Sliders had." $MP_{39,M}$ "The Creamy Sliders moved to freely and I was not able to save that many images. I lost control of the image to the point where it completely disappeared. [...] The Crunchy Sliders allowed me to control the output much better. [...]" $MP_{47,M}$ "They were a little more easier to control"

Creative Process: Design Variations: In Parameter Variations above, participants discussed how sliders moved and how the sliders let participants control parameters. On a similar note, some participants noted they were able to try out different designs or ideas. It is important to make this distinction as moving sliders is not the same as creating an actual design. As one participant pointed out, the limitation helped that participant think more creatively, which likely relates to the variations they saw in their final design.

$\mathrm{MP}_{04,M}$ "Limitations can help you think more creatively" $\mathrm{MP}_{36,S}$ "I could more quickly move through options, and more quickly get back to what I'd already done if I decided I liked the previous version better. I did like the greater range of options with the creamy sliders but I'd prefer it as an option to fine-tune the crunchy sliders after already getting the basic idea roughed out." $\mathrm{MP}_{39,M}$ "[With Creamy Sliders,] I was not able to save that many images. [...] I was able to save 22 thumbnails with [Creamy Sliders]."

Creative Process: Rapid Ideation: There were some participants that discussed the speed at which they came up with ideas, which is separate from the time it takes to complete the task. From their responses to the questionnaire, they preferred Crunchy because they either made one design very quickly, or they made several designs very quickly.

$\mathrm{MP}_{46,M}$ "The limited sliders allowed me to see all the options quicker - I did not need the amount of fine-tuning the free sliders provided." $\mathrm{MP}_{26,S}$ "There were fewer options, which aided coming up with a reasonably good picture quickly"

Creative Output: Better Results: For some participants, their preference of interface was related to what they were able to create. They felt the image produced using Crunchy Sliders was better than the image produced with Creamy Sliders. Part of the goal was to create a 'cool' image, so for these participants, their preference was largely related to their ability to achieve that goal.

$\mathrm{MP}_{05,M}$ "The result produced seemed better." $\mathrm{MP}_{06,M}$ "I really liked my results" $\mathrm{MP}_{29,M}$ "I was able to produce a better picture." $\mathrm{MP}_{41,M}$ "It seemed to produce better results for me."

6.4.9 Thematic Analysis of Interviews

After performing a thematic analysis of the interview responses, I came up with six themes specific to how Maximizers perceived the complexities of changing the search space and the impact on satisfaction and the creative process.

Move-Testing versus Exploratory Ideation: One participant ($\mathrm{NP}_{56,LM}$) stated "If I had a vision, I could have gotten to that point [with Creamy]. If I didn't have that vision, just exploring [would feel] overwhelming. ... Crunchy was easy to go through every possible option." This participant pointed out how Creamy Sliders is likely better for users who have an idea of what they are looking for (Move-Testing), while Crunchy Sliders may be better if you want to explore the space and try out the various options (Exploratory).

Varied Responses to Outcomes: The range of responses for Maximizers varied from $\mathrm{NP}_{55,HM}$ "neither outcome was worthy", $\mathrm{NP}_{57,LM}$ "Creamy let me be more creative," $\text{NP}_{57,LM}$ "[both were] nice looking quality images," and $\text{NP}_{58,HM}$ "I don't know if the [Creamy] picture was as cool." This could indicate a high variance of value between users. While some users preferred Creamy, they would state that the image created using Creamy was 'cooler.' $\text{NP}_{55,HM}$ even said that "not enough options were available" for both Creamy and Crunchy.

Value of Control: The value of the degree of control each interface offered cannot be understated. Even though some users acknowledged that Creamy sliders was "a little overwhelming" ($\text{NP}_{56,LM}$), Maximizers still emphasized the desire for more control "to explore all potential choices" ($\text{NP}_{55,HM}$) and because "Used to Photoshop to get exact result that you want" ($\text{NP}_{58,HM}$).

Thumbnail Exploration: $\text{NP}_{58,HM}$ pointed out that saving thumbnails of their work supporting her "try[ing] out a ton a different things in a really short time. A lot easier than having to save work and go into a new file to try something else out." It is a widely held practice in creativity support tools (e.g., Photoshop) to save iterations of work in case the user needs to go back to a previous design. Another practice is to copy the current state of the design and store different versions of designs in the same file, either in hidden layers or elsewhere on the main canvas. Participants used the save thumbnail feature before "wild changes" ($\text{NP}_{58,HM}$) or "screw[ing] up an image" ($\text{NP}_{58,HM}$). $\text{NP}_{57,LM}$ noted the risk inherent in exploration: "one click and you really change an image."

Time Exploring: Time was associated with "explor[ing] all potential options" ($\text{NP}_{55,HM}$). Two participants were aware of how much time was left, stating that they were "waiting for it [Crunchy] to be over" ($NP_{56,LM}$). The other two participants timed out. As mentioned before, $NP_{55,HM}$ wanted to try all potential options in both interfaces and $NP_{58,HM}$ said she timed out when using Crunchy, but when she used Creamy (in Block 2) she stated she "thought about hitting the submit button," but noted she ran out of time before she could act on the thought. In the literature, Maximizers are likely to not keep very good track of time. In this case, our Maximizers became aware of the time when they felt they were "satisfied with [their] work" ($NP_{56,LM}$) or when they had "gone through all possible options" ($NP_{57,LM}$).

Ordered versus Exploratory Search: The approach $NP_{58,HM}$ took in exploring the sliders was to "mess with one until I got where I wanted to," and then "go back if there's a different direction I wanted to go with it." $NP_{57,LM}$'s approach was different, he would "Tinker" with a slider he didn't fully understand and then go and try other ones. This demonstrates two ways of exploring the options that were made available in my interface.

6.5 Chapter Discussion

Here I discuss some of the interesting relationships between findings as well as the limitations and confounds that were present in my data.

6.5.1 Explorability

While Maximizers (like $MP_{02,M}$) saw that the Crunchy Sliders made trying each of the options more achievable, this difference was seen as limiting, as it meant the number of total creative possibilities was also decreased. I chose a dramatic difference in options per slider, from 100 options to 5 options. My rational was to create a parameter space that was obviously less vast. One participant wondered if the "frequency of intervals" (number of options available) could be increased in Crunchy. It could be that making the parameter space dramatically less vast was too drastic for participants. Some participants were able to figure out how many values were available in Crunchy Sliders and Creamy Sliders.

6.5.2 Reduction in Parameter Space

There were clearly some ordering effects, especially for Maximizers. This ordering effect is one-directional and occurs when there is a reduction in the parameter space that is provided by the interface. This reduction occurs when participants used Creamy Sliders in Block 1 and then used Crunchy Sliders in Block 2. The data does not suggest a counter/opposite effect, i.e., when there was an expansion of the parameter space (Crunchy first and then Creamy did), the data does not suggest users felt more liberated.

From the interviews, it is apparent that participants felt the reduction in parameter space in Crunchy Sliders limited their creativity. But this effect was only statistically significant when the Maximizers would use Creamy Sliders before Crunchy Sliders; it was not significantly different when observing Satisficers under the same condition. This suggests that Maximizers may have satisfied when the interaction felt like "you [didn't] have much choice, you have to work with what you have." This is an example of meta-satisficing with the interface, i.e., accepting the interface as is without wondering about what other interfaces are out there. In the case of Creamy then Crunchy, participants were likely to have been more critical of Crunchy because they had seen Creamy and therefore had a different interface to use as a comparison. When I had participants move from Creamy to Crunchy, I was taking away their ability to refine their ideas to a really high degree.

6.5.3 Familiarity

Several participants noted that the functionality provided by Creamy Sliders allowed for greater control of their final designs. Crunchy Sliders was designed to limit the user's ability to refine their designs, so they could try a wider variety of designs. The 'Crunchy' design decision had essentially taken away a core feature of photo editing tools, the ability to refine ideas; the importance of which became apparent through the results of this study.

6.5.4 What is an Option?

The difference between Creamy Sliders and Crunchy Sliders was designed to be along one spectrum: the number of design possibilities. However, the participants saw differences along various spectrums. In the anecdotes, there were several terms that were often used interchangeably by the participants: options, choices, ideas, and possibilities. In order to analyze the interview and free-response data, context was important to determining what sort of 'choices' participants were observing in the system. These terms would often be used by the participants to refer to: the number of values on the sliders, the number of sliders (i.e., properties and filters), the variations of a single image, or the number of different image design. The latter two I will cover more deeply in the next section.

This was also visible in how participants described their experiences. They rarely mentioned the number of created images as a limiting factor in their design process. The biggest limitation, as mentioned in a previous section, seemed to be the reduction in the size of the accessible parameter space. This suggests that from a interaction design perspective the number of 'choices' offered by an interface may be more closely related to the values users are manipulating in the interface and less closely with the number of design ideas they can explore. From a developer perspective, the two should be the same, after all manipulating values is how the user would go about creating different design ideas.

In order to further understand the results presented, it would be worth exploring how users of creativity support tools understand how their interaction with the values they manipulate impacts the different designs they can create.

6.5.5 What is Designing an Image?

From the interviews, it became apparent that there were two strategies participants used when designing the image. The first approach was to try all the sliders and then go back and use the sliders the participant felt would allow them to accomplish the image manipulation task. The other approach was to go through each slider until it was set to a value that closely matched what the user had envisioned. These two approaches emphasize how the participants engaged with the system. For those users that had an idea of what the final design could look like, the sliders became a tool for getting closer to that idea. On the other hand, if the participants had no idea in mind, they felt free to explore the sliders as a way of discovering different ideas.

These approaches were not fully studied in this experiment. However, I am still able to discuss the implications of these approaches as it related to interaction design.

It is possible that when users have a final design in mind, an interface like Crunchy Sliders limits their ability to be creative because it limits their ability to achieve their creative goals (where re-fining the idea is essential). On the other hand, if Crunchy Sliders is used as a tool for ideation, users may be more comfortable with delaying the refining process because they have not yet committed to an idea.

6.6 Future Work

In addition to the future work suggested in the discussion above, another study could be done where participants spend the first minutes of the study using Crunchy Sliders and save thumbnails along the way. Then, they switch to Creamy Sliders. This study is based on a suggestion from one of the interviewed participants who described how she would have liked to use Crunchy Sliders as she got to know the sliders and then switch to Creamy Sliders so she could get exactly what she wanted. In this study, participants could search very distinct designs in the parameter space, and then be able to load thumbnails later on for refining. They could also use Creamy Sliders to continue exploring other areas of the design space, this time having been primed by Crunchy Sliders to explore the search space more broadly.

6.7 Conclusion

In this chapter, I have described an experiment involving the Creamy Sliders and Crunchy Sliders Interfaces. I have demonstrated how Maximizers and Satisficers differ in how they see these tools supporting their creative process. I have also presented evidence that suggest participants meta-satisfice with the interface, that is participants are less critical of the Crunchy Interface, when it is the first interface they use and have not yet realized there has been a reduction in the accessible parameter space. I have also described several implications for interaction design, as well as presented some ideas of taking advantage of the benefits offered by reducing the parameter space, but also allowing users to refine their ideas when necessary.

CHAPTER 7: RESEARCH CONTRIBUTION

7.1 Thesis Statement

> *Interaction techniques and size of search spaces (defined by the granularity of control) for the exploration of large parameter spaces can positively and negatively effect the amount of options users explore as well as the users' satisfaction with their creative process and final designs. Some of these effects impact Maximizers differently than Satisficers.*

In this dissertation, I have presented a deep and rich investigation of the problem of exploration enormous parameter space in order to generate creative content.

7.2 Synopsis

In Chapter 1, I introduced the problem of exploring large parameter spaces and how users can become overwhelmed by the number of choices made available by the applications they use for creative content generation. In Chapter 2, I provided a review of the related research on interaction techniques that support explorability and maximizing and satisficing behaviors. In Chapter 3, I presented results from a user study that demonstrated how BiCEP better supports the exploration of the HSB parameter space than a single cursor color selector. This study lays the groundwork for understanding how parameter spaces are explored and one way to measure visited points in a parameter space. In Chapter 4, I provided an overview of the methodologies, constructs, and questions surrounding my study of explorability, satisficing, and satisfaction in parameter spaces. Chapter 5 and 6 include my research questions concerning parameter spaces and explains the studies that measured explorability, satisficing, and satisfaction when changing the interaction technique and search space, respectively. These measurements were used to answer a set of hypotheses concerning: real and perceived exploration of the parameter space, maximizers and satisficers, and satisfaction with the design process and final result.

7.3 Research Questions and Contributions

To recap the contributions of my dissertation, I answer the six research questions from Chapter 4 using the findings from the BiCEP study in Chapter 5 and the Creamy/Crunchy study in Chapter 6.

> *R1: How does changing the interaction technique used to navigate a parameter space impact people's satisfaction with their design process and final result?*

Changing the interaction technique can increase a user's satisfaction with the final result. However, user's satisfaction with the creative process did not seem to change when the interaction technique changed. The unfamiliarity of an interaction technique could change the creative process of users, which in turn offsetting the possible benefits provided by the (unfamiliar) interaction technique.

> *R2: How does changing the search space of a set of parameters impact people's satisfaction with their design process and final result?*

When there is a reduction in the search space, user satisfaction with their creative process can decrease. While user's may not be satisfied with their creative process, they may still be satisfied with their final designs. The results from my study are inconclusive, but from the perspective of my participants, they may not have always liked the constraints, but recognized there were some benefits to their search process and recognized that they still made too 'good enough' designs.

R3: How do Maximizers and Satisficers differ when is comes to their satisfaction with the design process and final results?

Maximizers, who are inclined to want more options, were different than satisficers in that Maximizers became less and less satisfied as they explored more and more color options. For the image manipulation task, Satisficers were more satisfied with their creative process than Maximizers even when the search space was changed. There were no statistically significant differences between how satisfied Satisficers were between conditions. This suggests meta-satisficing of the interface, where Satisficers will be satisfied with their process regardless of how the interface changes. Maximizers' satisfaction only differed when there was a reduction in the search space (i.e., using Crunchy Slider in Block 2), in which case Maximizers were more dissatisfied. In Block 1, Maximizers' satisfaction were similar between interfaces. If there was a reduction in explorable parameter space, Maximizers were less satisfied with the interface. However, if there was an increase in the explorable parameter space, their satisfaction did not increase, though they were still likely to prefer the interface that offered more options. This also suggests meta-satisficing of the interface, where users that have no knowledge of alternative forms of an interface, may just as satisfied with smaller explorable parameter space as with a larger explorable parameter space.

> *R4: How does the actual percentages of the parameter space explored impact the satisfaction users have with their process and final result?*
>
> *R5: How does the perceived percentage of the parameter space explored impact the satisfaction users have with their process and final result?*

For the image manipulation task, participants were able perceive the differences in exploration between the different search spaces of explorable parameter spaces. However, I found that higher explorations of the parameter space resulted in lower satisfaction with the creative process. Actual exploration did not have the same impact on satisfaction with the final design. This could indicate that explorability has a stronger relationship with the creative process than the creative content.

For the coloring task, no relationship was found between satisfaction and actual exploration. This could be due to inconsistencies between how participants perceived their exploration. There was only a difference in perceived exploration when participants used the more fluid interaction technique (i.e., BiCEP) after having used the familiar interaction technique (i.e., Native color picker). This could indicate that the differences in explorability are better perceived when participants can compare interfaces. While no direct relationships were found between the level of satisfaction and exploration, I did find an indirect relationship. Participants explored more unique colors using the more fluid interface (i.e., BiCEP) and were more satisfied with their final designs created using that interface. This indicates that there is more than just explorability that plays a factor in a user's satisfaction with their final design. This question could be addressed with help from the design community and additional interviews concerning users' satisfaction with their design process and final result.

> *R6: What are the differences in how parameter space interaction techniques and search spaces support creativity?*

Exploration and Expressiveness were the two orthogonal factors that were rated as most important to supporting creativity in both the coloring activity and image manipulation task. Changing to a more fluid interaction technique resulted in higher score for creativity support. Reductions the search space, resulted in a lower score creativity support. This finding becomes especially important for those creating interfaces where Exploration and Expressiveness will be important creativity support factors in the tasks enabled by the interface.

7.4 Discussion

In this section, I discuss topics that span the entire scope of the dissertation.

7.4.1 Crunchy Experiences in Interfaces

Users could drag or click along the slider to change the values in real time. However, when moving the 'blur' slider the canvas would take a little over a second to update. This is an artifact of the increasing computational complexity of doing a pixel level blurring of an image using the javascript language. This creates a Crunchy type of experience as users are not getting a fluid exploration of this slider. They can drag the UI slider, but they will not see the image update until they pause or stop dragging.

This point is important when considering that many of the popular filters in GIMP or Photoshop take longer than a second to compute. This delay is compounded when working with higher resolution images, taking minutes to hours to render a filtered image. While faster, multi-threaded processors could mitigate this issue, the fact is that many users are already having Crunchy Experiences in modern software applications. Some filters get around this issue by having thumbnails previews (low-res version of the image) or a small region-of-interest that takes less time to update. However, the final rendering may be different than the preview and the user may not discover the subtle differences until after they render the entire image.

7.4.1.1 Crunchy Movement

Interacting with BiCEP can also be considered a Crunchy Interaction. When moving the two cursors in BiCEP, the user can simultaneously change all three dimensions of the color space. While this allows the user to explore a broader range of colors, it has the added effect that small movements in BiCEP will change the color more than with Native. This is a Crunchy Interaction because BiCEP makes slightly larger movements around the parameter space. These movements are not the same as creative leaps across the interface, but they have been shown in my dissertation to encourage exploring more unique colors.

7.4.1.2 Crunchy Interaction

On the contrasting side to BiCEP having Crunchy Movements, the Native color picker has more of a Crunchy Interaction. The user must update the brightness slider and then mode-switch to the color wheel. This waiting period of having to control one component before you can move on to the next is a Crunchy Interaction at a macro level. Note that this means at a macro level Creamy Sliders and Crunchy Slider both had Crunchy Interaction, as users must adjust one slider before moving on to the next slider.

7.4.2 Defining Controls

I mentioned earlier that Crunchy Experiences already exist in some of our modern software application suites. Some users noted that while the understood the benefits of using Crunchy, they really wanted to have control over their experience, which is often not the case in many applications. However, some applications allow the user to define what controls they use to explore alternatives. For example, the architecture software, Grasshopper, allows the user to set which controls they use in their 3D models. These layouts can become very complex, but it is a level of complexity that seems necessary when considering that in a model of a bridge or building there are an overwhelming number of parameters that can be adjusted. Giving the user the power to define what controls they use, allows them to focus on those parameters that are most important to them. They can later refine what controls they have if necessary.

7.5 Future Work

In both studies, participants mentioned the benefits of using either interface. Several suggestions were made as to when each interface (i.e., BiCEP versus Native, Creamy versus Crunchy) would be more appropriate.

For Creamy/Crunchy, future studies could give participants the same task, but ask them to spent 5 minutes using Crunchy Sliders to generate ideas, and then switch to

Creamy Sliders in order to refine those ideas. Crunchy Sliders would allow participants to try out many different and unique ideas without getting overwhelmed with small refining. From my studies, it was obvious that taking away the ability to refine an idea was seen as detrimental to the task goal. This study allows participants to use both. The control conditions would be two separate groups of participants that use only one interface (i.e., Creamy Sliders or Crunchy Slider).

Based on the Creamy and Crunchy interfaces, an future interface could provide 'zoomable' sliders that enable users to change the size of the search space. These 'zoomable' sliders would allow the user to adjust how much fine and course grain control they would like to have over the sliders. One interesting side-effect of this additional feature could be that users will have to make a decision about what is a good zoom-level to work in, which adds to the number of decisions they are already having to make in the interface.

To better understand how BICEP can support creativity, I suggest conducting an in-the-wild study where the plugin is used by participants in the interfaces they regularly use. This would allow me to compare BiCEP to not just the native color picker but other color pickers like RGB sliders or the crayon box. Since there are some familiarity issues with Bimanual Interaction on the trackpad, transferring the BiCEP interface to a touch enabled device may provide new insight into how it can used in an environment where user's have multiple points of control is familiar.

The two orthogonal factors that were important to participants in both studies were exploration and expressiveness. This dissertation focused on variations in explorability, which only accounted for some of the differences in satisfaction. Future work should also look at further understanding the role of expressiveness, and what relationship exploration has on how much a user can express themselves.

7.5.1 Interaction Techniques and Search Space

In my dissertation work, I did one study where I changed the interaction technique but kept the size of the search space constant, and another study where I kept the interaction technique constant and changed the size of the search space. In future work, I could develop a prototype interface that supported both using a bimanual interaction technique for controlling multiple dimensions of parameters and allowing the user to control the size of the search space. One of the limitations of my work is that it does not address how there could be interaction effects between different interaction techniques and different sizes of the search space. For example, I could develop a version of the coloring activity that allowed the user to switch between using BiCEP and another color picker. The software could also allow the user to adjust the size of the HSB color space in BiCEP. They could explore all colors, their specific range of perceivably different colors, or a selection of 255 traditional colors.

7.5.2 Expanding These Results

The experimental results presented in this dissertation are limited by the number of participants. My sample sizes ranged from 15 to 54. Many behavioral and psychometric studies recruit over a thousand participants. This allows the researcher to tease out clusters of behavior that are more likely to appear when the level of noise in the data disappears. By recruiting a lager number of participants, differences between interaction techniques and search space sizes are likely to become more pronounced.

I could not accept many of my hypotheses, even though they were supported by the literature. In future work, similar experimental studies could be done with a larger population to see how the results differ. As I stated before, maximizing and satisficing behaviors are dependent on the context, which could explain the lack of statistical significant differences that were found in my data. With a higher volume of data, the issues of context will likely disappear. The larger dataset could make it easier to tease out the specific exploration and satisfaction differences within the Maximizer and Satisficer user-populations.

7.6 Conclusions

I have presented design implications that lay the groundwork for understanding the relationship between explorability, maximizing and satisficing behavior, and the satisfaction of users in regards to their design process and final result. The results of my study demonstrate the explorability can support satisfaction with the creative process, users are aware of the explorability offered by interfaces, users can report differences in how much exploration they do, and that Maximizers are more critical than Satisficers about the differences in explorability. The results of my study also indicate that each interaction technique and search space size offers benefits to the creative process of users.

For those doing research in interaction design, measuring the explorability of an interface can be a useful metric that goes beyond simple time and error metrics. Exploration is happening in our interfaces, but exploration has been always been well defined by researchers. I have demonstrated that exploration is complex from both an interaction design and cognitive perspective. Researchers must be clear about what types of exploration are enabled by an interface, how the user decides to interact with parameter spaces using the interface, and how the user perceives what their own exploration looks like.

I suggest that future designers create ways to allow users to control how they interact with parameter spaces, by either changing the interaction technique or changing the search space. My work can help designers better understand this relationship in an effort to support users in the face of overwhelming large parameter spaces that a user must navigate in order to make creative content.

REFERENCES

[1] A. T. Adams, B. Gonzalez, and C. Latulipe. Sonicexplorer: Fluid exploration of audio parameters. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 237–246. ACM, 2014.

[2] B. Asian and G. Zech. Statistical energy as a tool for binning-free, multivariate goodness-of-fit tests, two-sample comparison and unfolding. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 537(3):626 – 636, 2005.

[3] R. Balakrishnan and K. Hinckley. Symmetric bimanual interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '00, pages 33–40, New York, NY, USA, 2000. ACM.

[4] J. B. Bendor, S. Kumar, and D. A. Siegel. Satisficing: A'pretty good'heuristic. *The BE Journal of Theoretical Economics*, 9(1), 2009.

[5] R. S. Berns. *Billmeyer and Slayzman's Principles of Color Technology.* John Wiley & Sons, Inc, 3rd edition, 2000.

[6] F. Block and H. Gellersen. Two-handed input in a standard configuration of notebook with external mouse. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, NordiCHI '10, pages 62–71, New York, NY, USA, 2010. ACM.

[7] P. Brandl, C. Forlines, D. Wigdor, M. Haller, and C. Shen. Combining and measuring the benefits of bimanual pen and direct-touch interaction on horizontal interfaces. In *Proceedings of the working conference on Advanced visual interfaces*, AVI '08, pages 154–161, New York, NY, USA, 2008. ACM.

[8] V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

[9] W. Buxton and B. Myers. A study in two-handed input. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '86, pages 321–326, New York, NY, USA, 1986. ACM.

[10] S. K. Card, J. D. Mackinlay, and G. G. Robertson. The design space of input devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 117–124. ACM, 1990.

[11] E. A. Carroll, C. Latulipe, R. Fung, and M. Terry. Creativity factor evaluation: towards a standardized survey metric for creativity support. In *Proceedings of the seventh ACM conference on Creativity and cognition*, C&C '09, pages 127–136, New York, NY, USA, 2009. ACM.

[12] S. Chaudhuri, E. Kalogerakis, S. Giguere, , and T. Funkhouser. Attribit: Content creation with semantic attributes. *UIST 2013*, Oct. 2013.

[13] J. Cohen. *Statistical power analysis for the behavioral sciences*. Academic press, 2013.

[14] I. Dar-Nimrod, C. D. Rawn, D. R. Lehman, and B. Schwartz. The maximization paradox: The costs of seeking alternatives. *Personality and Individual Differences*, 46(5):631–635, 2009.

[15] M. de Vega, M. Marschark, M. J. Intons-Peterson, P. N. Johnson-Laird, and M. Denis. Representations of visuospatial cognition: a discussion. *Models of visuospatial cognition*, pages 198–226, 1996.

[16] S. A. Douglas and A. E. Kirkpatrick. Model and representation: the effect of visual feedback on human performance in a color picker interface. *ACM Trans. Graph.*, 18(2):96–127, Apr. 1999.

[17] A. Franklin, O. Wright, and I. R. Davies. What can we learn from toddlers about categorical perception of color? comments on goldstein, davidoff, and roberson. *Journal of Experimental Child Psychology*, 102:235–245, 2009.

[18] R. Fung, E. Lank, M. Terry, and C. Latulipe. Kinematic templates: end-user tools for content-relative cursor manipulations. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*, UIST '08, pages 47–56, New York, NY, USA, 2008. ACM.

[19] B. Gonzalez and C. Latulipe. Bicep: bimanual color exploration plugin. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*, CHI EA '11, pages 1483–1488, New York, NY, USA, 2011. ACM.

[20] Y. Guiard. Asymmetric division of labour in human skilled bimanual action: The kinematic chain as a model. *Journal of Motor Behaviour*, pages 486–517, 1987.

[21] P. Henry, S. Westland, and T. Cheung. An intuitive color-selection tool. In *Fourteenth Color Imaging Conference*, pages 144–147. Society for Imaging Science and Technology, 2006.

[22] P. Hutterer and B. H. Thomas. Groupware support in the windowing system. In *Proceedings of the eight Australasian conference on User interface - Volume 64*, AUIC '07, pages 39–46, Darlinghurst, Australia, Australia, 2007. Australian Computer Society, Inc.

[23] S. S. Iyengar, R. E. Wells, and B. Schwartz. Doing better but feeling worse looking for the best job undermines satisfaction. *Psychological Science*, 17(2):143–150, 2006.

[24] K. A. Jameson, S. M. Highnote, and L. M. Wasserman. Richer color experience in observers with multiple photopigment opsin genes. *Psychonomic bulletin review*, 8(2):244–261, 2001.

[25] E. Kang, E. Jackson, and W. Schulte. An approach for effective design space exploration. In *Foundations of Computer Software. Modeling, Development, and Verification of Adaptive Systems*, pages 33–54. Springer, 2011.

[26] A. N. Kolmogorov. Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, 4:83–91, 1933.

[27] W. H. Kruskal and W. A. Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260):pp. 583–621, 1952.

[28] C. Latulipe. *A Model for Symmetric Bimanual Interaction*. PhD thesis, The University of Waterloo, 2006.

[29] C. Latulipe, I. Bell, C. L. A. Clarke, and C. S. Kaplan. symtone: two-handed manipulation of tone reproduction curves. In *Proceedings of Graphics Interface 2006*, GI '06, pages 9–16, Toronto, Ont., Canada, Canada, 2006. Canadian Information Processing Society.

[30] C. Latulipe, C. S. Kaplan, and C. L. A. Clarke. Bimanual and unimanual image alignment: an evaluation of mouse-based techniques. In *Proceedings of the 18th annual ACM symposium on User interface software and technology*, UIST '05, pages 123–131, New York, NY, USA, 2005. ACM.

[31] C. Latulipe, S. Mann, C. S. Kaplan, and C. L. A. Clarke. symspline: symmetric two-handed spline manipulation. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, CHI '06, pages 349–358, New York, NY, USA, 2006. ACM.

[32] C. Latulipe, M. Youngblood, I. Bell, and C. Orlando. Tonezone: image exploration with spatial memory cues. *seventh ACM conference on Creativity and Cognition*, pages 427–428, 2009.

[33] A. Leganchuk, S. Zhai, and W. Buxton. Manual and cognitive benefits of two-handed input: an experimental study. *ACM Trans. Comput.-Hum. Interact.*, 5(4):326–359, Dec. 1998.

[34] M. R. Luo, G. Cui, and B. Rigg. The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color Research & Application*, pages 340–350, 2001.

[35] Macbeth. *Color Checker Chart*. Munsell Color, Baltimore, MD, 1990.

[36] S. Malacria, J. Scarr, A. Cockburn, C. Gutwin, and T. Grossman. Skillometers: reflective widgets that motivate and help users to improve performance. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*, pages 321–330. ACM, 2013.

[37] R. B. Miller. Response time in man-computer conversational transactions. In *AFIPS Fall Joint Computer Conference*, volume 33, pages 267–277. ACM, 1968.

[38] R. Misuraca and U. Teuscher. Time flies when you maximizemaximizers and satisficers perceive time differently when making decisions. *Acta psychologica*, 143(2):176–180, 2013.

[39] T. Moscovich and J. F. Hughes. Multi-finger cursor techniques. In *Proceedings of Graphics Interface 2006*, GI '06, pages 1–7, Toronto, Ont., Canada, Canada, 2006. Canadian Information Processing Society.

[40] G. Y. Nenkov, M. Morrin, A. Ward, B. Schwartz, and J. Hulland. A short form of the maximization scale: Factor structure, reliability and validity studies. *Judgment and Decision Making*, 3(5):371, 2008.

[41] L. Noy, Y. Hart, N. Andrew, O. Ramote, A. Mayo, and U. Alon. A quantitative study of creative leaps. In *International Conference on Computational Creativity*, page 72, 2012.

[42] A. Oulasvirta, J. P. Hukkinen, and B. Schwartz. When more is less: the paradox of choice in search engine use. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 516–523. ACM, 2009.

[43] G. Paolacci, J. Chandler, and P. Ipeirotis. Running experiments on amazon mechanical turk. *Judgment and Decision Making*, 5(5):411–419, Aug. 2010.

[44] C. Peters, G. Castellano, and S. de Freitas. An exploration of user engagement in hci. In *Proceedings of the International Workshop on Affective-Aware Virtual Agents and Social Robots*, AFFINE '09, pages 9:1–9:3, New York, NY, USA, 2009. ACM.

[45] D. A. Schön. *The reflective practitioner*, volume 1. Basic books, 1999.

[46] B. Schwartz. The tyranny of choice. *Scientific American*, 290(4):70–75, 2004.

[47] B. Schwartz, A. Ward, J. Monterosso, S. Lyubomirsky, K. White, and D. R. Lehman. Maximizing versus satisficing: happiness is a matter of choice. *Journal of personality and social psychology*, 83(5):1178, 2002.

[48] H. A. Simon. A behavioral model of rational choice. *The quarterly journal of economics*, pages 99–118, 1955.

[49] H. A. Simon. Rational choice and the structure of the environment. *Psychological review*, 63(2):129, 1956.

[50] H. A. Simon. The structure of ill-structured problems. In *Models of discovery*, pages 304–325. Springer, 1977.

[51] N. Smirnov. Table for estimating the goodness of fit of empirical distributions. *The Annals of Mathematical Statistics*, 19(2):279–281, 06 1948.

[52] M. Terry and E. D. Mynatt. Side views: Persistent, on-demand previews for open-ended tasks. In *UIST 2002*, pages 71–80. ACM Press, 2002.

[53] M. Terry, E. D. Mynatt, K. Nakakoji, and Y. Yamamoto. Variation in element and action: supporting simultaneous development of alternative solutions. In *CHI 2004*, volume 24, pages 711–718, 2004.

[54] D. A. Varakin, D. T. Levin, and R. Fidler. Unseen and unaware: Implications of recent research on failures of visual awareness for human-computer interface design. *Human–Computer Interaction*, 19(4):389–422, 2004.

APPENDIX A: EXPERIMENTAL MATERIALS FOR COLORING STUDY

Demographic Survey

1. What is your sex?
   - ○ Male
   - ○ Female

2. What is your age range?
   - ○ 18-24
   - ○ 25-34
   - ○ 35-44
   - ○ 45-54
   - ○ 55+

3. What is your ethnicity? (Check all that apply)
   - ☐ American Indian or Alaskan Native
   - ☐ Asian or Pacific Islander
   - ☐ Black or African American
   - ☐ Hispanic or Latino
   - ☐ White or Caucasian
   - ☐ Other [Specify...]

4. How long have you been using a MacBook?
   - ○ < 1 year
   - ○ 1-2 years
   - ○ 2-5 years
   - ○ 5+ years

5. How would you categorize your experience with graphic design software activities? (e.g., Photography, Print, 3D Models, etc.)
   - ○ Little to no experience
   - ○ Beginner (e.g., used 2-5 times)
   - ○ Intermediate (e.g., use occasionally)
   - ○ Advanced (e.g., use regularly)

[Submit]

Pre-Survey

Move the slider to indicate your agreement or disagreement with the following statements.

When I am at home or in the car listening to the radio, I often check other stations to see if something better is playing, even if I am relatively satisfied with what I am listening to.

Highly Disagree **Highly Agree**

No matter what I do, I have the highest standards for myself.

Highly Disagree **Highly Agree**

I often find it difficult to shop for a gift for a friend.

Highly Disagree **Highly Agree**

I never settle for second best.

Highly Disagree **Highly Agree**

No matter how satisfied I might be with my job, it would only be right for me to be on the lookout for better opportunities.

Highly Disagree **Highly Agree**

Choosing a video to rent/stream is really difficult, I am always struggling to pick the best one.

Highly Disagree **Highly Agree**

Submit

Post Questionnaire

Please move the slider to indicate your agreement or disagreement with the following statements

I am satisfied with my final design.

Highly Disagree **Highly Agree**

I made my final design carefully.

Highly Disagree **Highly Agree**

I am satisfied with my design process using the provided color picker.

Highly Disagree **Highly Agree**

I felt overwhelmed by the color choices offered by the color picker.

Highly Disagree **Highly Agree**

Given the huge number of possible colors, it's impossible to explore all the color combinations. What percentage of the color space do you think you considered while using the provided color picker?

(Please move the slider to indicate a value between 0% and 100%.)

0% **100%**

Submit

Final Questionnaire

Move the slider to indicate your agreement or disagreement with the following statements.

When selecting a color to use in a computer application, I will often try out different colors from across the spectrum.

Highly Disagree **Highly Agree**

Choosing the best color for my work, can often be difficult.

Highly Disagree **Highly Agree**

When selecting a color, I often imagine what other colors may be a better choice.

Highly Disagree **Highly Agree**

Which color picker did you prefer using today? *Please answer as honestly as possible*

- Single Cursor - Similiar to the built-in Mac OS X color picker
- Dual Cursor - Works when two fingers on placed on trackpad

Why? *Please answer as honestly as possible*

Enter your response here...

Submit

APPENDIX B: EXPERIMENTAL MATERIALS FOR SEARCH SPACE STUDY

Demographic Survey

1. What is your sex?
   - ◯ Male
   - ◯ Female

2. What is your age range?
   - ◯ 18-24
   - ◯ 25-34
   - ◯ 35-44
   - ◯ 45-54
   - ◯ 55+

3. What is your ethnicity? (Check all that apply)
   - ☐ American Indian or Alaskan Native
   - ☐ Asian or Pacific Islander
   - ☐ Black or African American
   - ☐ Hispanic or Latino
   - ☐ White or Caucasian
   - ☐ Other [Specify...]

4. How would you categorize your experience with graphic design software activities? (e.g., Photography, Print, 3D Models, etc.)
   - ◯ Little to no experience
   - ◯ Beginner (e.g., used 2-5 times)
   - ◯ Intermediate (e.g., use occasionally)
   - ◯ Advanced (e.g., use regularly)

5. How would you describe your use of Instagram (or similiar photo filtering and uploading app)?
   - ◯ I have never used it
   - ◯ I have stopped using it
   - ◯ I upload a photo around once a month or less.
   - ◯ I upload a photo at least once a week
   - ◯ I upload a photo almost (if not) daily.
   - ◯ I upload photos multiple times a day

[Submit]

Pre-Survey

Move the slider to indicate your agreement or disagreement with the following statements.

When I am at home or in the car listening to the radio, I often check other stations to see if something better is playing, even if I am relatively satisfied with what I am listening to.

Highly Disagree **Highly Agree**

No matter what I do, I have the highest standards for myself.

Highly Disagree **Highly Agree**

I often find it difficult to shop for a gift for a friend.

Highly Disagree **Highly Agree**

I never settle for second best.

Highly Disagree **Highly Agree**

No matter how satisfied I might be with my job, it would only be right for me to be on the lookout for better opportunities.

Highly Disagree **Highly Agree**

Choosing a video to rent/stream is really difficult, I am always struggling to pick the best one.

Highly Disagree **Highly Agree**

Deciding how to respond to all these statements is exhausting work, please leave the slider below in the middle.

Highly Disagree **Highly Agree**

When selecting a color to use in a computer application, I will often try out different colors from across the spectrum.

Highly Disagree **Highly Agree**

Choosing the best color for my work, can often be difficult.

Highly Disagree **Highly Agree**

When selecting a color, I often imagine what other colors may be a better choice.

Highly Disagree **Highly Agree**

Continue

Post Questionnaire

Please move the slider to indicate your agreement or disagreement with the following statements

I am satisfied with my final design.

Highly Disagree **Highly Agree**

I think my final design are suitable as backgrounds for a Album or DVD cover.

Highly Disagree **Highly Agree**

I made my final design carefully.

Highly Disagree **Highly Agree**

I am satisfied with my design process using the interface provided.

Highly Disagree **Highly Agree**

I felt overwhelmed by the design choices offered by the interface provided.

Highly Disagree **Highly Agree**

Given the huge number of combinations of parameter/filter settings, it's impossible to explore all combinations. What percentage of the color space do you think you considered while using the provided interfaces?

(Please move the slider to indicate a value between 0% and 100%.)

0% **100%**

Submit

Final Questionnaire

1. How motivating did you find today's task:
"Create a cool or interesting image that could be used as a background for an Album or DVD cover"

Very Un-motivating **Very Motivating**

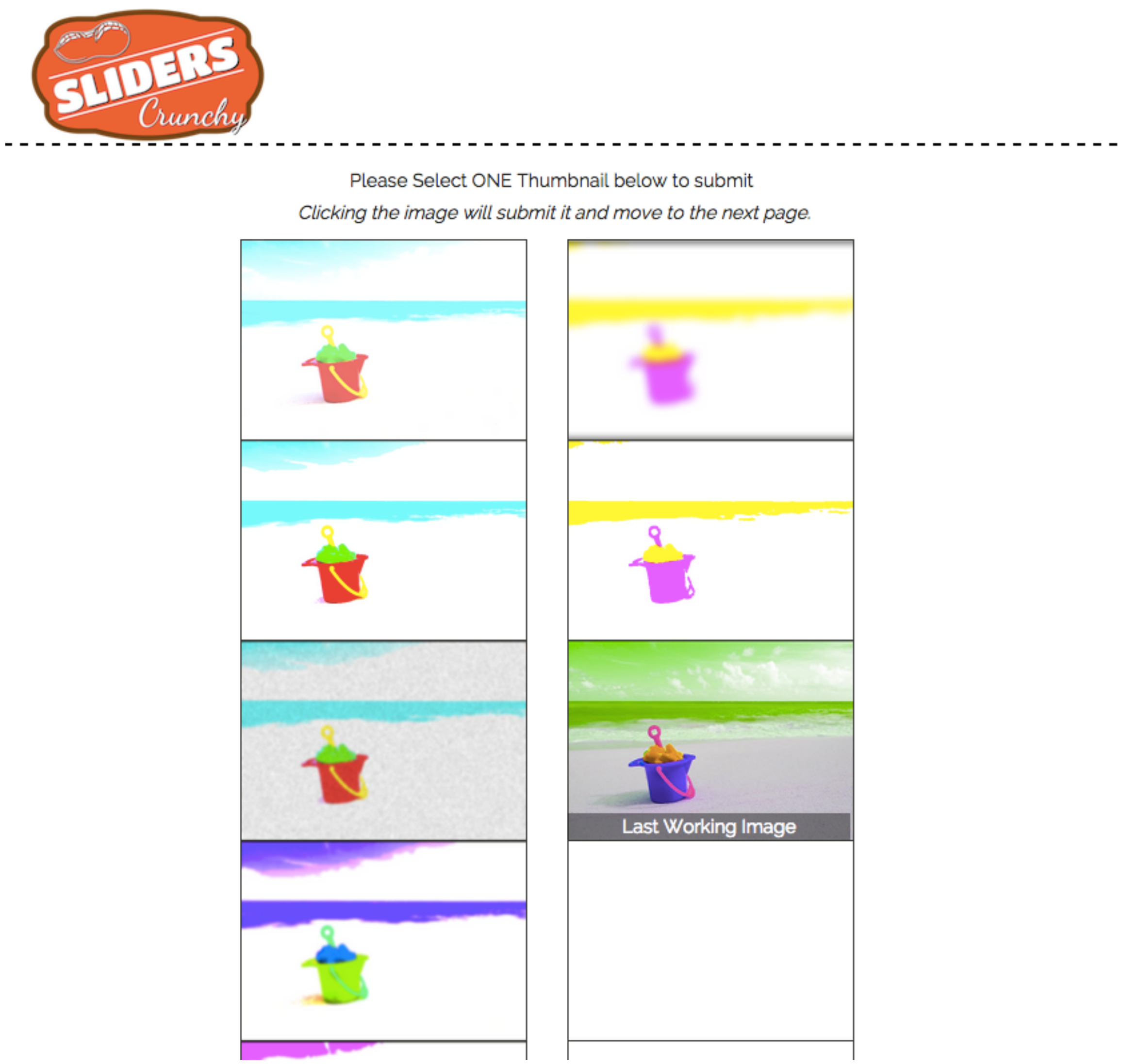
2.1 Which interface did you prefer using today?

- Creamy Sliders - Sliders moved freely
- Crunchy Sliders - Sliders were limited

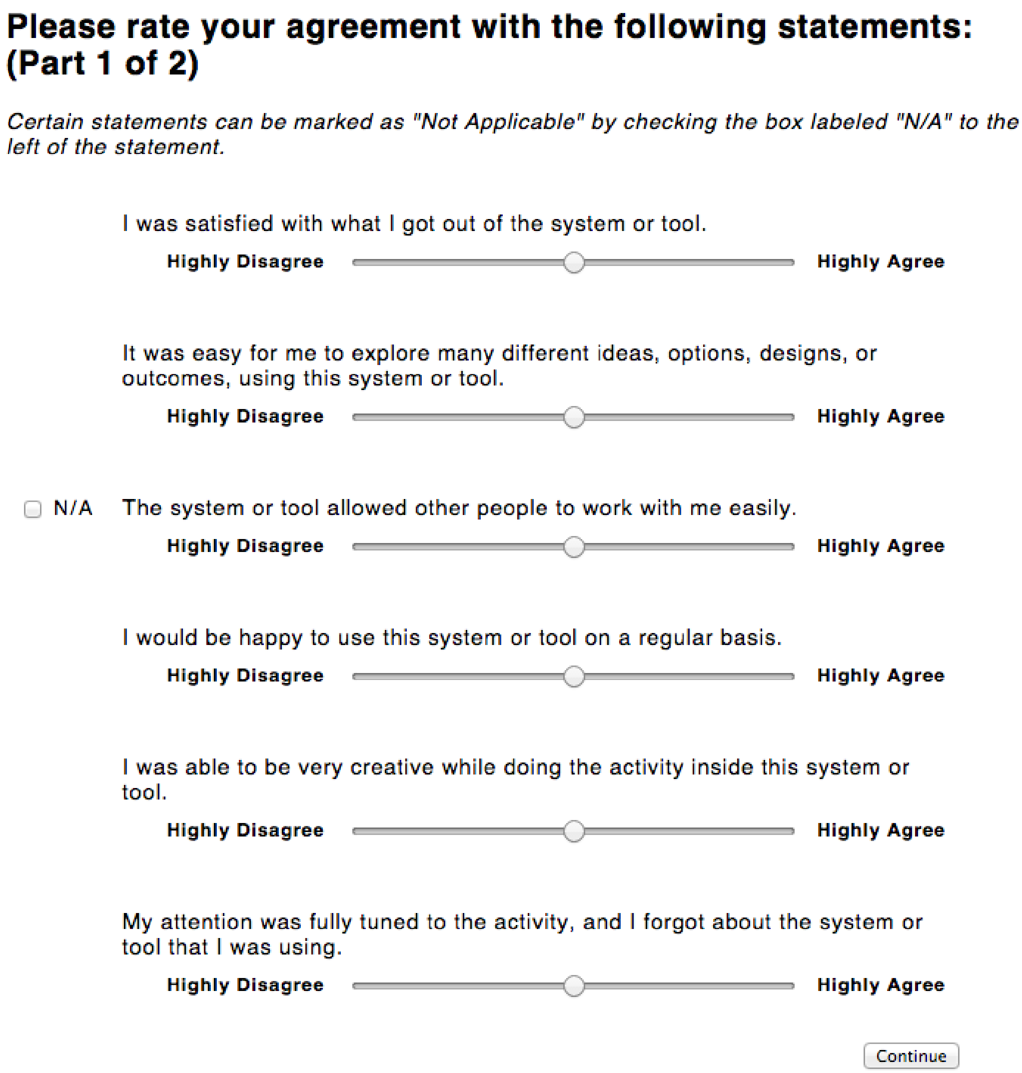
2.2 Why? *Please answer as honestly as possible*

Enter your response here...

Submit


SLIDERS
Crunchy
Please Select ONE Thumbnail below to submit
Clicking the image will submit it and move to the next page.
Last Working Image

APPENDIX C: CREATIVITY SUPPORT INDEX FOR BOTH STUDIES



Please rate your agreement with the following statements: (Part 1 of 2)

Certain statements can be marked as "Not Applicable" by checking the box labeled "N/A" to the left of the statement.

I was satisfied with what I got out of the system or tool.

Highly Disagree **Highly Agree**

It was easy for me to explore many different ideas, options, designs, or outcomes, using this system or tool.

Highly Disagree **Highly Agree**

N/A The system or tool allowed other people to work with me easily.

Highly Disagree **Highly Agree**

I would be happy to use this system or tool on a regular basis.

Highly Disagree **Highly Agree**

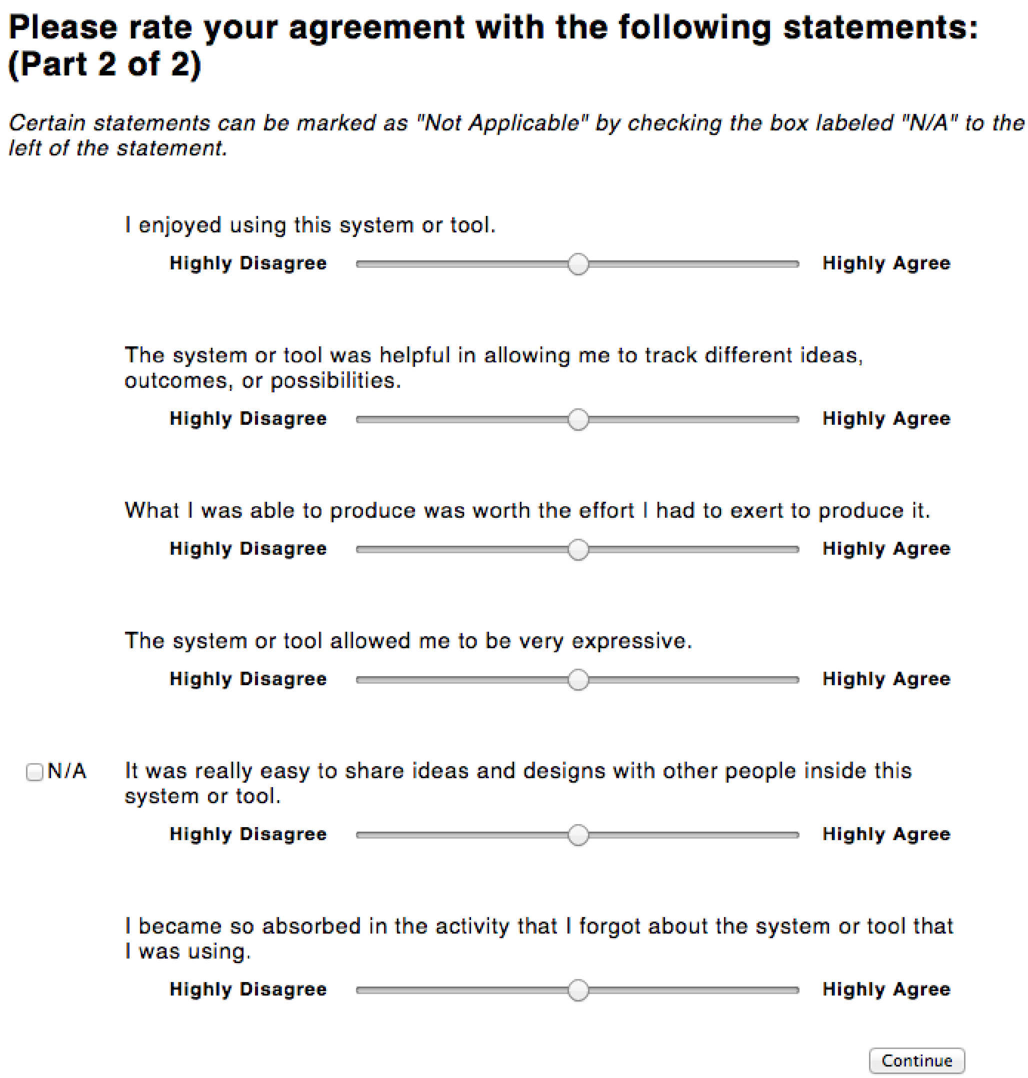
I was able to be very creative while doing the activity inside this system or tool.

Highly Disagree **Highly Agree**

My attention was fully tuned to the activity, and I forgot about the system or tool that I was using.

Highly Disagree **Highly Agree**

Continue



Please rate your agreement with the following statements: (Part 2 of 2)

Certain statements can be marked as "Not Applicable" by checking the box labeled "N/A" to the left of the statement.

I enjoyed using this system or tool.

Highly Disagree **Highly Agree**

The system or tool was helpful in allowing me to track different ideas, outcomes, or possibilities.

Highly Disagree **Highly Agree**

What I was able to produce was worth the effort I had to exert to produce it.

Highly Disagree **Highly Agree**

The system or tool allowed me to be very expressive.

Highly Disagree **Highly Agree**

N/A It was really easy to share ideas and designs with other people inside this system or tool.

Highly Disagree **Highly Agree**

I became so absorbed in the activity that I forgot about the system or tool that I was using.

Highly Disagree **Highly Agree**

Continue

When doing this task, it's most important that I'm able to...

Explore many different ideas, outcomes, or possibilities ○ ○ Work with other people

1/15 Continue