Productive and Unproductive Friction in Game Design

by

Isaac Sung

A dissertation submitted in partial fulfillment of
the requirements for the

degree of

Doctor of Philosophy
(Computer Sciences and Education)

at the

UNIVERSITY OF WISCONSIN–MADISON

2021

Date of final oral examination: 3/16/2021

The dissertation is approved by the following members of the Final Oral Committee:
Matthew Berland, Associate Professor, Curriculum & Instruction; Computer Sciences
Erica Halverson, Professor, Curriculum & Instruction
Yea-Seul Kim, Assistant Professor, Computer Sciences
Krista-Lee Malone, Faculty Associate, Curriculum & Instruction

To all the future dreamers.

Acknowledgments

First, I would like to thank all of my current and former committee members who have supported me along the way. They have all been incredibly encouraging, insightful, and helpful throughout this process. Without them, this work would not have been possible. Specifically, I would like to thank Dr. Matthew Berland for his unending optimism and belief in my abilities even when I doubted myself. Without his persistence, I would not have been able to finish this project as I did. I count myself lucky and blessed that he replied to my out-of-the-blue email when I was looking for a new direction and advisor. He also deserves thanks and credit for the origins of this project.

I am indebted to all the great faculty at UW–Madison that I had the pleasure to learn from. Through their guidance and expertise, I gained many skills that will be invaluable for my future as a researcher, scholar, and educator. There are too many names to name all of the faculty who influenced me, but it would be remiss of me not to mention Dr. Martina Rau by name. Through serendipity, she taught a large number of the courses I took in graduate school. Her careful attention to detail and constant encouragement pushed me to become a better researcher and writer, and for that I will be forever grateful.

I owe many thanks to my friends and colleagues who have supported me during this project. Zach Buller deserves special thanks for being a constant friend willing to provide valuable insights and unending encouragement.

My professional colleagues at Ripon College have also been immensely understanding while I tried to complete my dissertation. I am grateful for my departmental colleagues who have been flexible and supportive with whatever I needed to finish

my degree.

I would also like to thank my parents who have been the biggest supporters of my education since the beginning. They instilled in me a curiosity and thirst for knowledge that has stayed with me my whole life. It is no accident that I have followed in the footsteps of my father by entering academia.

Finally, and most importantly, my wife Charis and her ever constant love and support in everything I do has given me strength over the past several years. This document surely would never have been completed had you not been with me, and I would not be who I am today without you.

Contents

List of Figures

List of Tables

Abstract

Video game design is an important aspect of how games can impact players in terms of behavior, player experience, and learning. *How* a game is designed and not just *what* is in a game affects how users are influenced. This dissertation introduces productive frictional designs as a new category of designs that may be useful for researchers and designers searching for creative and empirically studied methods to affect players in their games. Productive friction is hypothesized to create points of friction in the gameplay but result in positive outcomes for the player (e.g. learning, increased game performance, improved player experience). The framework for productive friction draws on theories from various fields such as the learning sciences, human-computer interaction, and player-computer interaction. For example, research on productive failure [73] and microboundaries [30] provide a theoretical basis for expecting positive outcomes from introducing friction into a game.

Two types of productive frictional designs are proposed and examined: obstructions on game actions and juicy design. This study aimed to evaluate the effectiveness of productive friction in the form of obstructions and juice empirically. In order to do so, two primary research questions were pursued:

1. How do productive and unproductive design frictions in a casual word game influence players' behaviors and experiences?

2. What effects might productive and unproductive design frictions in a casual word game have on game outcomes?

A mixed-methods approach was conducted to investigate the effects and mechanisms of obstructive and juicy designs in a custom designed word game called *Brave New Word*. One study gathered anonymous game log data from 200+ participants who downloaded and played the game. Another study consisted of one-on-one video conferencing sessions with 7 participants that included think-alouds and interviews. The studies found that productive frictions in the form of obstructions have the potential to change player behavior by introducing microboundaries that encourage mindful interactions. Another general finding was that productive frictions in the form of juicy design have the potential to improve player experiences by creating engaging feedback but do not seem to impact player performances in games. These findings indicated that certain design frictions can be included in games to benefit players. Current user experience and player experience literature mostly align with these findings and support the conclusions made in this document.

Chapter 1: Introduction

1.1 Design Patterns and Design Frictions in Games

The design of games is an important field of study because design can influence behavior and attitudes in players. Many of the studies surrounding games investigate what effects games have on players, but fewer describe deeply how games achieve these effects. Though most of this research involves educational or "serious" games [26] for understandable reasons—the belief among many people being that games are purely for entertainment or wasting time [109], this idea of changing players through games can apply to non-educational games as well. Most people are already familiar with the notion that video games can affect the people who play them. For example, a common issue that video game researchers must deal with is the topic of violence and how violent games may negatively influence the people who play them. Though findings on this are heavily debated and sometimes contradictory [18,83], most video game researchers agree about the potential value of games to effect positive change in players [26, 47]. When attempting to explain how a game might impact players, the first things to be investigated are most likely the content and the design. This work will approach the subject from the design perspective and remain agnostic towards the content within games. This is not to suggest that content is unimportant; it most certainly is. However, the main focus will be on game design and the findings presented in this study should be interpreted with this in mind.

Design is a broad term; to narrow the scope, this work will specifically examine game design from the perspective of design patterns and design frictions. Design patterns are formal descriptions of practical solutions to common design problems. The

origin of design patterns is attributed to Alexander [2] in his work on architecture. His goal was to create a pattern language that could be used to communicate about solutions and ideas with other designers. This concept was later adopted successfully by other fields, most notably in software engineering [45]. Gamma [45] defines design patterns as something that names, motivates, and explains a general design solution that addresses a common design problem. Soon, the concept of game design patterns was proposed by Björk and Holopainen [12] and taken up by others [37,58,84,87,122], which led to the development of educational game design patterns [6, 8, 21, 79]. In order to support educational game design, researchers and designers proposed and compiled libraries of educational game design patterns as a framework for aligning learning mechanics with game design elements.

A similar concept to game design patterns is the idea of lenses [120]. Lenses provide various non-exclusive perspectives one can take when approaching game design. Each lens varies in the level of its abstraction and can be used in conjunction with other lenses to complement each other. Design patterns can also vary in their abstraction, but tend to be more granular in detail and describe specific implementation ideas for concrete problems, such as the pattern of *collections* (of items, badges, etc.) or *player level system* (where a player's character levels up and gains strength over time) [87].

Some have resisted the idea of game design patterns, imagining that they will result in dull, uncreative games if all designers start to draw from the same standard pattern library. However, when examining other creative disciplines, design patterns could be compared to tropes and genre elements. Though overuse of tropes and standard genre elements in creative writing, music, or film is generally seen as uninspired and banal, a professional creator can only access a toolbox of solutions and creatively subvert expectations if they are well-versed in the standard elements

of their craft (e.g. tropes). In the same way, if game designers wish to be creative in their solutions to creating effective and entertaining games (educational or not), familiarity with design patterns and lenses is essential.

One area of research that needs to be explored more thoroughly is the theoretical connection between game design components and their effects on players' experience and behaviors. Dormann, Whitson, and Neuvians' study [37] investigating game design patterns and affective learning sought to fill in some of these gaps. Although the patterns they listed were not comprehensive, the ones that were described contained rich descriptions of not only the game mechanics but also the emotional, social, and cultural theories connecting them to affective change and learning. They also carefully mapped the relationships among the patterns, including patterns that were previously identified not in direct connection with affective learning, thus establishing a pattern language as envisioned by Alexander [2]. Carvalho et al. [21] also established a respectable model for serious games analysis that joined activity theory and common game mechanics. For this study, I will examine a particular category of game design patterns called frictional design patterns. I aim to extend this category by introducing the theory of productive friction and identifying productive and unproductive frictional design patterns.

Design frictions are anything that creates difficulties for a user while interacting with a technology [30]. For example, font that is too small to read or pop up windows on a website are frictions that create a negative experience for the user. There are design patterns and philosophies that aim to find benefits in friction. Frictional design patterns are those that impede the player in some way (provide a point of friction) in order to advance some other goal of the designer [90]. Productive friction builds on this idea and incorporates the idea of intentionally adding design friction to have a positive effect. In general, user experience design research and guidelines

promote the removal of all friction from an interaction to increase ease of use and user engagement [82, 116]. However, there has been a surge of interest in user experience literature in the potential positive benefits of certain kinds of intentional friction, such as *microboundaries* to prompt more mindful interaction [30]. In addition, game designers have always known that player experience is not about removing friction in every part of the experience; indeed, adding just the right amount of friction to the appropriate elements of a game can be key to improving player engagement and experience because challenge is an essential element of most game experiences [32].

Productive friction also draws inspiration from principles of productive failure [73] and subversive game design [93] in how they transform elements normally seen as negative, such as failure and common game tropes, into benefits. The hope is that productive friction may provide one explanation for how game design can positively influence players and become another tool in the toolbox for game designers and researchers. In order to study productive friction, two promising examples of what could be labeled productive friction in games were chosen: *obstructions* (as a form of microboundaries [30]) and *juice* [66]. Obstructions in games include opponents and any other obstacles or limitations on the game actions that can be taken [9]. Juiciness is typically defined as "excessive positive feedback" [35, 66]. Chapter 2 will explore in further detail the connections between productive friction, obstructions, and juice.

Though the concepts of juicy design and obstructions in games are considered to be important by many game designers and researchers [43, 120], there is a dearth of empirical research showing their potential benefits and disadvantages. Researchers have identified potential theories and explanations for how juice could impact players' experience and performance such as the effects of color, audio, and feedback, but these theories require further testing. Similarly, there is relatively little understand-

ing of how design frictions in the form of obstructions influence players in games. This study aims to make progress in those gaps.

1.2 Research Questions

There is a need for knowledge that reveals the connections between specific design choices and players' behaviors and experiences. Current theory combined with empirical evidence will be required to show the relationships between design frictions and player outcomes. In order to address these needs, this study investigates the following research questions in the context of a casual word game developed specifically for this project:

1. How do productive and unproductive design frictions in a casual word game influence players' behaviors and experiences?
    a. How do productive and unproductive obstructions impact players' behaviors and experiences?
    b. How does productive and unproductive juice impact players' behaviors and experiences?
2. What effects might productive and unproductive design frictions in a casual word game have on game outcomes?
    a. What effects might productive and unproductive obstructions have on game outcomes?
    b. What effects might productive and unproductive juice have on game outcomes?

These research questions aim to determine whether productive and unproductive frictions in games have any measurable impacts on players and if so, how they had those effects and through what mechanisms. Analyzing game outcomes in addition to players' behaviors and experiences aligns with current approaches to player experience research. Specifically, it has been argued that differences in objective game metrics, such as game scores or other measurable indicators (e.g. number of clicks), can demonstrate changes in player behavior or attitudes [38]. As such, the second research question naturally follows from the first.

The findings to these questions would provide evidence for the usefulness of productive friction as a design concept. If significant effects are discovered with regards to productive friction, then further questions investigating the theory could be pursued. Existing literature suggests that productive frictional designs might result in minor boosts to players' game performance abilities and experiences while unproductive frictional designs could correlate with negative player experiences and game outcomes. Therefore, I formulate the following hypotheses:

- *H1a:* Productive obstructions will lead to more reflective gameplay behaviors.
- *H1b:* Unproductive obstructions will worsen the overall player experience.
- *H1c:* Productive juice will improve the overall player experience and aesthetic appeal.
- *H1d:* Unproductive juice will be distracting and worsen the overall player experience.
- *H2a:* Productive obstructions will result in higher game performance metrics.
- *H2b:* Unproductive obstructions will result in lower game performance metrics.
- *H2c:* Productive juice will result in higher game performance metrics.
- *H2d:* Unproductive juice will result in lower game performance metrics.

This study will address these questions through the analysis of game log data, think-aloud studies, interviews, and surveys. If productive friction is shown to be an effective theory, then the next question would be to explore the mechanisms by which productive frictional designs support change in players. This is motivated by a desire to understand how game designs can influence people. When we know why certain designs work the way they do, we will be able to make better choices when designing future games. My hypothesis is that the frequency and severity of the frictional elements will determine the effectiveness (negative or positive) of the design. By formalizing the definition of productive friction and the principles of productive frictional design, I will be able to better examine these questions. The use of treatment and control group testing, think-aloud studies, interviews, and surveys facilitated the investigation of this question.

1.3 Summary

1.3.1 Intended Audience

This dissertation is primarily targeted at researchers who are interested in video game design—both educational and non-educational in nature. The findings may be especially relevant when researchers are concerned with creating more mindful gameplay experiences. Some results may also be beneficial to game designers who are interested in the impacts of user interface (UI) elements and juicy game design on player experiences.

1.3.2 Document Outline

Chapter 2 discusses relevant literature and provides the background for productive friction, obstructions, and juice in game design. Since measuring player experience is key to answering the research questions, the chapter describes several common methods for measuring it, including interviews and think-aloud sessions. Productive friction is defined within such frameworks as microboundaries, slow design, and reflective design. Obstructions and juicy design are then linked to productive friction. The material in this chapter sets up much of the terminology and design decisions used in the rest of the document.

Chapter 3 describes and justifies the design of the game used in this study, *Brave New Word*. It explains the design choices for adding appropriate obstructions and juice to the game. The features of each version, including the control design, are thoroughly explained.

Chapter 4 provides a description of the methods used to gather and analyze data for this dissertation. Two separate but complementary studies are described—one is quantitative in nature, while the other is essentially qualitative. Since the design of the game is described in Chapter 3, this chapter only contains details of the participants, forms of data collected, experimental procedures, and methods of data analysis. It includes how logistic regression was used in the quantitative study and the qualitative coding process for the qualitative study.

Chapter 5 reports the results of the quantitative study, including the findings from logistic regression. The chapter summarizes the descriptive statistics of various gameplay metrics collected from a large sample, and it details the findings of multiple logistic regression models across different variables.

Chapter 6 details the findings of the qualitative study. It includes a summary

of the qualitative coding scheme and detailed summaries of each qualitative study session with the participants.

Chapter 7 is an analysis of the main findings from the two studies. It summarizes the key results and makes connections between the quantitative and qualitative results to support the analyses. The main findings are concerned with the effects and mechanisms of productive obstruction and juice as well as the benefits and disadvantages of the chosen methodology.

Lastly, Chapter 8 provides a summary of the theoretical and practical implications of this study, the limitations, as well as potential future directions for related work.

Chapter 2: Obstructions and Juice as Productive Friction

2.1 Overview

To help address both **RQ1** and **RQ2**, this chapter will provide a description of the nascent theory of productive friction drawn from relevant knowledge in user experience (UX) and player experience (PX) research, such as microboundaries [30], slow design [127], and juicy design [120]. These pertinent theories about productive friction in game design will motivate and guide the work proposed in this study. Next, the chapter describes obstructions and juice as possible design frictions and how they relate to productive friction. These frameworks for productive friction, obstructions, and juice will lead to the design of the game used for this study, which is explained in Chapter 3. Finally, the chapter will conclude with a brief discussion on how productive friction could impact game designers and researchers.

2.2 Productive Friction

Productive friction is the idea that intentionally introducing frictions into a game—in the form of obstructive user interface elements, distracting juicy feedback, or other features that have seemingly negative consequences—can lead to more productive outcomes (e.g. mindful interactions [30], immersion [17], or flow [31]). If designed well, the obstacles and impasses in a game can lead to productivity and not frustration. Carefully balancing challenge and players' skill-levels can lead to greater immersion [59]. This concept draws its inspiration from a variety of frameworks,

including microboundaries [30], slow design [50], and productive failure [73], which are discussed below.

The hope is that productive friction will provide a useful framework for designers and researchers to create and analyze games. Though some of the framing of productive friction will be presented from an educational perspective, the concepts are not limited to games for learning per se. Commercial game designers may also be interested in these theories for effecting change in player behavior and thinking in order to create productive and satisfying gaming experiences.

2.2.1 Intentional Design Friction

Design friction is a concept discussed by popular media articles about UX [19,51,140] and peer-reviewed research [30] which refer to "points of difficulty encountered during users' interaction with a technology" [30]. In user experience research, the goal is typically to minimize or eliminate design frictions so that the user can engage with the product fully. Many usability studies are focused on optimizing ease-of-use of the application by eliminating unnecessary friction points. For example, a study by Wiseman et al. [138] found that the interfaces of number entry systems on medical devices could be improved to increase accuracy and efficiency.

Cox et al. [30] describe the potential for creating mindful interactions by introducing intentional design frictions in the form of *microboundaries*. They define a microboundary as any design element that restricts the user from "rushing from one context to another" [30]. A microboundary creates an opportunity for the person to reflect on their actions before acting. For example, a warning prompt window might ask whether or not the user really wants to quit the application: the user can still easily close the application, but the pause provides a moment to make sure they

have saved their progress or to change their mind. The goal of microboundaries is to move people from System 1 thinking to System 2 thinking [70]. System 1 thinking is described as "quick thinking," or automatic thinking with little-to-no conscious thought, while System 2 thinking is defined as "slow thinking," or calculated, intentional thinking that requires conscious effort. A microboundary offers a chance for people to use the more mindful System 2 thinking.

Cox et al. [30] define "mindful interactions" as being those interactions that require deliberate and intentional System 2 thought. In the context of games, one might ask why mindful interactions would be beneficial. Cox et al. provide three possible benefits, all of which could apply to games: (1) increasing accuracy by avoiding speed accuracy trade-offs in memory related actions, (2) avoiding actions that do not align with personal values, and (3) directing the user towards a specific action.

Productive friction generalizes the idea of microboundaries to other forms of intentional design friction that can improve the user experience. Mejtoft, Hale, & Söderström [91] similarly cite other examples of design friction that are intended for good: slow design [50] and uncomfortable interaction [11]. Mejtoft et al. [91] describe these simply as "design frictions," but since that term can also have a negative connotation in UX literature, I am proposing that the term *productive friction* be used instead to clearly identify intentional design frictions meant to have a positive impact.

2.2.2 Designing for Productive Failure in Games

Productive failure is a learning theory first proposed by Kapur [73] that uses short term failures to allow students to explore representations and solution methods before arriving at long term success. Past research has shown the potential value of

designing for productive failure in the context of games [5, 46]. Games provide an excellent vehicle for productive failure because of their ability to engage and motivate players through challenges and difficult gameplay. Productive failure in games requires players to follow a failure or negative experience with a successful action or result in the game [5, 46]. In Anderson et al.'s study of the educational game Virulent [5], they found that the number of failures in level completion before the first success was predictive of learning outcome gains. This can be compared to Kapur and Bielaczyc's two-phase design of a productive failure experience [78]: the problem-solving phase and the consolidation (or instruction) phase. The failures occurring in the initial problem-solving phase correspond to the failures players experience in a game, while the consolidation phase relates to the follow up success after a failure in a game. The concept of productive failure closely relates with productive friction in that they both design for short term obstacles that eventually lead to a positive outcome.

2.2.3 Other Related Work

Dark game design patterns describe a category of design patterns that take advantage of players and negatively impact them against their consent [141]. Dark game design patterns are becoming increasingly common in the form of so-called "free-to-play" (F2P) mobile games that dominate the market. Despite their name, the business strategy of some F2P games involves manipulating players into spending exorbitant amounts of money in the form of "microtransactions" through dark design patterns [3].

Frictional design patterns also involve game design elements that may go against players' interests in the short-term, but they are not intended to harm the players.

Rather, they are design patterns that are used by game designers who need to sacrifice usability to achieve other necessary goals [90], typically in the realm of education. The description of frictional design patterns is highly relevant to the definition of productive friction. The key part of the description is the temporary nature of the impact on the player which does not impede their free will or progress, and the objective of productive friction is to improve players' experiences or outcomes. This is in stark contrast with dark design patterns that actively attempt to coerce the player into negative or unproductive states. Mechtley & Berland [90] provide the example of obfuscating game state as an example of a frictional design pattern. In their study, the design requirement needing to be fulfilled was to measure changes in players' attention without the use of specialized hardware. By constraining players to interact with the game interface in a way that would typically go against acceptable usability standards, they were able to get accurate measures of when and how long players accessed a specific window in the game. By measuring when users accessed the game window, they were able to infer when players paid attention to the target content. This is a good example of a frictional pattern that impedes the user slightly but provides value to the researcher/designer.

Research in subversive game design shows that upending common game design patterns can provoke a shift in thought, mental models, and strategies for players [93]. By designing games that subvert common game patterns, and thus the usual expectations of players, designers are able to prompt users to rethink their approach to playing the game and adjust their assumptions and mental schemas which can result in System 2 thinking or behavior change.

2.2.4 Definition and Principles

Defining Productive Friction

I will attempt to iteratively develop a definition for productive friction by comparing and contrasting it with related work. I will then condense the important characteristics into a proto-definition for productive friction. The four related frameworks I will use as guideposts are dark design patterns, frictional design patterns, productive failure, and subversive game design.

Dark design patterns (1) negatively impact the player, (2) go against the player's best interest, and (3) happen without the player's consent [141]. Productive friction, on the other hand, (1) should only have temporary or minor setbacks, (2) should take into consideration the user's best interests, and (3) should have the player consent to the opportunity for growth. From this, we can create our first proto-definition for productive friction:

> Proto-definition 1: Productive friction is a game element that has a temporary negative impact on the player for which the designer has the user's best interests in mind.

At this stage, the proto-definition is overly general and not very useful. In the following iterations, I will try to further define what friction is beyond "temporary negative impact on the player" and clarify what the "user's best interests" are. The definition does not explicitly include anything about the user's consent to keep the definition concise and because the phrase, "user's best interests," should implicitly cover consent. In order to better define friction, frictional design patterns will be examined next.

Frictional design patterns (1) sacrifice usability to (2) advance the designer's objectives [90]. Productive friction can also (1) sacrifice usability and have temporary negative impacts on the player, but it needs to (2) have a productive outcome for the player, not just the designer/stakeholder. This leads to our second proto-definition:

> Proto-definition 2: Productive friction is a game element that sacrifices usability and has a temporary negative impact on the player, but it also has a productive outcome for the player's best interests.

This second definition goes further in specifying what friction is by defining it as something that "sacrifices usability" along with the temporary nature of the effect on the player. However, that limits what can count as friction, as the term usability has strong connotations in the design field. The scope of friction can probably be broadened without making the definition completely impractical, and a better scope needs to be defined for what a "productive" outcome is. This definition also calls into question how productive friction will result in a productive outcome for the player if the other parts of the definition are essentially the same as that of frictional design patterns. For more insights into how friction can be defined and what mechanisms could generate productive outcomes, productive failure will be explored in relation to productive friction.

Productive failure consists of (1) failures in attempting to solve problems and (2) exploring multiple representations and solution methods before (3) discussing canonical representations and methods to solve a problem. This leads to (4) better learning outcomes and transfer [73]. Productive friction is (1) not necessarily interested in the success or failure of a high-level "problem," but is more focused on a granular level of detail (e.g. friction for a small task or action). However, (2) exploring multiple options and potential solutions is a useful mechanism, and (3) though discussion

could be valuable, productive friction is not situated in a classroom context. As players overcome impasses (i.e. friction), (4) better learning outcomes and transfer could be considered productive. Combining these comparisons brings us to our next proto-definition:

> Proto-definition 3: Productive friction is an obstacle or impasse that impedes a player's ability to accomplish a task or action in a game and challenges the player to overcome it which leads to productive learning outcomes.

This definition widens the scope of friction to include any obstacles or things that impede players' actions, but it contains the clause, "challenges the player to overcome it" to describe a mechanism that might lead to productive outcomes. It also more strictly defines the scope of productive outcomes to those in the educational sphere. In a practical sense, a game designer might find it more useful if there were more guidance in what kinds of obstacles could be used to prompt productive friction, and designers might be interested in outcomes besides learning. For this, the last model to investigate is subversive game design.

Subversive game design (1) subverts common game patterns to upend players' expectations to (2) promote shifts in strategy, mental models, or paradigmatic thinking [93]. Productive friction (1) could also introduce friction through subverting game patterns, but that is not the only useful category of friction. (2) Promoting shifts in strategy, mental models, or paradigmatic thinking is another good candidate for what constitutes a "productive" outcome. And so, that leads us to our final proto-definition of productive friction:

> Proto-definition 4: Productive friction is an obstacle or impasse that impedes players' ability to accomplish a task or action in a game by sac-

> rificing usability or by subverting expectations and promotes productive learning outcomes or shifts in thinking by challenging them to overcome the obstacle.

This definition essentially defines two categories of productive friction in games and establishes learning outcomes and shifts in thinking as the two types of productive outcomes. Although this definition is much more practical than the first proto-definition, it may require more refining as time goes on as more categories for friction and productive outcomes are identified. Now that a serviceable definition for productive friction has been developed, the next question to consider is, how can one design for productive friction in games?

Design Principles for Productive Friction

As with the process for coming up with a definition, analyzing the design principles used for other frameworks is helpful for theorizing principles for productive friction. For productive failure, Kapur synthesized a set of design principles to promote productive failure based on multiple studies:

1. The initial problem-solving task should be challenging enough to engage the learner in the exploration, but not so challenging that the learner gives up.
2. It must admit multiple strategies, and representations—that is, afford sufficient problem and solution spaces for exploration.
3. The problem should activate learners' prior knowledge—formal as well as intuitive—to solve the problem.

4. A teacher or an expert should build upon the student-generated solutions by comparing and contrasting them with the correct solution, thereby directing attention to and aiding encoding of the critical features of the targeted concept. [77]

Similarly, Mitgutsch & Weise suggest the following principles for designing for learning through subversive game design:

1. Players must develop an understanding of a situation, challenge, or obstacle.
2. Players must create an expectation of how to overcome it only to have that expectation subverted by an unexpected result.
3. Players must then restructure their understanding, their expectations, or their frame of reference.
4. Players then develop a new approach to overcome the obstacle.
5. The previous steps are repeated until players develop a pattern for how to overcome the problem. [93]

Drawing inspiration from the related works and the proto-definition for productive friction developed above, I propose the following design principles for productive friction:

1. The frictional element must subvert the user's expectations or be challenging/disruptive to the gameplay in some way, but not so much that it frustrates the player enough to quit.
2. Overcoming the frictional element must be possible and result in a sense of satisfaction (e.g. accomplishing an objective, mastering a skill).

3. The challenge of the frictional element should be proportional with the prior skill and understanding of the player.

Principle 1 is based on the first principle of designing for productive failure and the model of affect dynamics in complex learning [36]. It is intuitive that if an obstacle causes too much distress on a player, it will have a negative impact. D'Mello and Graesser's study of affect dynamics in learning also showed that persistent failure can lead to frustration and disengagement from the task [36]. Thus, it is important that the frictional element in the game not be overly frustrating for the player.

Principle 2 is inspired by the design principles from subversive game design and the model of affect dynamics. The recursive learning theory from subversive game design applies to productive friction as players will, in principle, learn as they refine their strategies and overcome challenges [93]. The model of affect dynamics is relevant as well because of the importance of impasse resolution in achieving a productive affective cycle of flow and engagement [36].

Principle 3 is guided by the principles of productive failure and the deeper roots of Vygotsky's theory of the zone of proximal development (ZPD) [136]. Players do not exist in a context-free bubble; rather, their prior knowledge and sociocultural context shape their frames of reference and ability to achieve certain tasks. Educational game designers would be remiss not to consider the developmental level and local contexts of the target audience, and non-educational game designers should also consider their players' backgrounds and skill levels.

These principles are derived from the manifold theories and principles discussed above. They form the initial groundwork for designing games to promote productive friction from which the principles can later be refined. This work aims to empirically study the efficacy of productive friction as a design pattern.

To examine how game design patterns and lenses can lead to learning and shifts in attitude or thinking, the following sections will present two examples of productive friction that show the viability of influencing through design. These were chosen specifically because there exist explicit theoretical connections between their designs and player change. The two lenses that will be explored are *obstructions* and *juice*.

2.3 Obstructions as Friction

Obstructions in games are a fundamental method for creating challenge and meaningful play. Many games researchers agree that challenge is an integral part of games. For instance, Salen & Zimmerman [119] discuss the importance of "artificial conflict" in games. In particular, obstructions to game actions are essential to providing meaning to the gameplay. Myers defines gameplay from a semiotic perspective and discusses the significance of opposition in games [95]. Obstructions in games are often realized in the form of enemies or opponents to overcome [9], such as Goombas in Mario games. However, Barr & Khaled [9] also note that obstructions can be more generalized and be anything that applies to players' actions. Essentially, all obstructions are, in one form or another, obstructions on game *actions*. Thus, once a game has been broken down into all of the possible actions that can be taken, it is easy to analyze potential obstructions or design new ones. This specification of obstructions on game actions is an important lens through which researchers can analyze or design games. It is natural to move from the perspective of *actions* to *obstructions* when analyzing or designing games.

What can obstructions result in, besides building challenge in a game? In their qualitative study on game obstructions, Barr & Khaled also found that obstructions can prompt learning [9]. That is, they found that obstructions "provide critical

prompts to players in terms of learning how to play" [9]. They found instances of player learning due to obstructions in four games of vastly different styles: *Civilization III* [42], *Half-Life 2* [133], *Grand Theft Auto: San Andreas* [114], and *The Sims 2* [41]. This illustrated that obstructions on game actions can be a useful pattern for not only bringing challenge and excitement to a game but also to encourage instances of learning.

The framework of microboundaries [30] aligns well with that of obstructions. Obstructions in games could be designed to act as microboundaries to promote System 2 thinking and reflection, which could lead to positive behavior change or PX. By creating micropauses or explicit prompts for the player, the obstructions could have an objective besides creating challenge. Microboundaries introduce a way for designers to reason about how to create productive obstructions. Examples of microboundaries that could be used as obstructions in games include reflective prompt windows, confirmation screens, etc.

In this study, instances of obstructions on game actions will be examined to attempt to identify the characteristics that distinguish productive obstructions that prompt learning or System 2 thinking and unproductive obstructions that merely frustrate or negatively impact the player.

2.4 Juice as Friction

The concept of "juice" in video games evolved from game designers striving to achieve a good game "feel" as defined by Swink [129]. Juice is a term for "constant and bountiful user feedback" [43]. Juiciness has become a popular topic at games conventions, and popular talks have been given showcasing the importance of juice in games [64]. Juice is typically implemented as an abundance of animation, particle, and audio

effects [57]. For example, in *Super Mario Brothers* [99], jumping as Mario creates a distinct sound effect as does collecting coins.

Game developers typically talk about juice as a method of improving the player experience and increasing player engagement. Swink [129] theorized that juice can boost players' perceived competence and result in better engagement. However, there are few empirical studies showing definitive effects of juiciness. Juul & Begy found no statistically significant effects when comparing a juicy version of a game to a non-juicy version [67], while Hicks et al. [57] found that juiciness did not improve performance in a VR simulation, though there was evidence that it could impact player behavior. Most recently at the time of this writing, Kao [72] found that moderate amounts of juiciness significantly improved player experience and game performance.

Could there be other potential outcomes to juice besides player experience or performance? In a Master's thesis investigating the effects of juiciness on a prototype game, Buckthal showed that it was possible that juice could lead to less confusion while undergoing tutorials and learning the game [16]. This result is similar to the potential effects that obstructions in games can achieve. More research needs to be conducted on the effects that juiciness can have, and what characteristics of juiciness contribute to its effectiveness.

One question that will be considered in this dissertation is if juiciness can be a predictably useful design pattern to improve player experience and game performance despite the potentially distracting visual and audio effects. Since the additional cognitive load introduced by juiciness could be considered a friction for the player, this work views juicy design through the lens of productive friction. Juiciness may hinder players in some ways, but there are also potential benefits such as improved player experience and performance.

2.5 Potential Impacts on Game Design

The theory behind productive friction and the results of this study may be of interest to researchers in HCI or PCI, especially those concerned with influencing users' behaviors and experiences through design. Obstructions and juice as productive friction offer the hope that subtle changes in game design can have significant effects on players. Game design researchers may also find the individual topics of obstruction and juiciness relevant to their interests. Juiciness is an especially juicy topic to investigate due to its disproportionate mainstream popularity compared to the amount of empirical studies on the subject.

Using productive friction to influence players' behaviors and cognitive processes could be considered a game design pattern. Game design patterns are important to establish a common language among game designers and researchers, including educational experts. Effective communication is key to creating effective games that can truly impact players and advance game design towards the future. A case study in how game design patterns can be useful when designing a new educational game is presented in Mitgutsch & Weise's study on *Afterland* [93]. In the case of *Afterland*, through the identification of common game design patterns and the theory of subversive game design, they were able to create a unique game that overturned players' expectations and challenged their modes of thinking. With proper expansion, productive frictional design patterns and principles could have similar success for different audiences. For educational researchers, this provides an opportunity to communicate their needs with professional game designers in a practical manner. Being grounded in some educational theories, productive friction presents more opportunities for researchers to explore video games and learning with a theory-

informed approach.

The framework of productive friction may be of interest to commercial game designers as well. Productive friction could provide game designers a novel lens (à la Schell [120]) with which to view their games. Commercial game and UX designers often desire their players to experience the game in a certain way. Viewing obstruction and juicy design as intentional design friction could unlock creative ideas for how to influence players' behaviors and experiences without negatively impacting them.

Chapter 3: Designing a Game with Productive Friction

3.1 Overview

A mobile game called *Brave New Word* was designed and developed specifically for this study using the Unity game engine [132] to facilitate easy modification of the UI and game mechanics. The game is essentially a word search game where players seek to form high scoring words by connecting letters that are adjacent to each other. The concept is similar to Boggle [131] and other word games like SpellTower [44]. A puzzle word game was chosen because of its appeal to a wide variety of audiences, especially casual and mobile gamers. The mobile game market has rapidly overtaken the PC and console game markets as the largest share in terms of revenue. According to a report by Newzoo, mobile game revenues in 2020 accounted for roughly 48% of the global market [137]. In addition, the report also estimates that this share is equivalent to $77.2 billion, which is a 13.3% year-over-year increase. These statistics indicate mobile games are a significant part of many people's lives, and so their designs and effects should be seriously investigated.

To this end, two frictional designs were implemented in *Brave New Word*, *obstructions* and *juice*. Each frictional design was also implemented as a productive friction and an unproductive counterpart. In addition to these 4 versions, a control version with no frictional elements, a combined productive obstruction & juice version, and a combined unproductive obstruction & juice version were created for a total of 7 versions. This chapter will present the design objectives for each of the frictional

designs. Then, it will describe the process of developing the prototype. Finally, it will conclude with an overview of the final design of *Brave New Word*.

3.2 Design Objectives

The main design objectives of *Brave New Word* were to influence players to change their gameplay strategies and behaviors and/or alter their overall PX—either negatively or positively. These goals were pursued through the utilization of productive and unproductive frictions in the form of obstructions and juiciness.

3.2.1 Productive Obstruction

The first frictional design is *productive obstruction*. This pattern includes any designs that intentionally decrease the usability of taking a game action with the goal of increasing player attentiveness, mindful interactions, or otherwise improve the gameplay experience. This draws on the concept of microboundaries [30]. Obstructions on game actions were chosen as a design friction because it is one of the most obvious points in which friction can be introduced in a game. As discussed by Barr et al. [9], obstructions in video games are integral to defining the player experience. As such, adding various obstructions to actions such as submitting a word or selecting letters was a natural place to start in the design.

The intention for adding productive obstructions is to introduce subtle obstacles that encourage System 2 thinking or mindful interactions. By adding obstructions on game controls, the goal is not to simply slow down the player; rather, the main objectives are to promote reflection and guide the user towards a strategy of finding rare words. These can be accomplished through the introduction of "micropauses"

[30] and nudging the player to consciously think about what they are doing before they act. For instance, a game might ask the player to confirm their action before allowing them to make a big decision.

3.2.2 Unproductive Obstruction

The next frictional design is *unproductive obstruction*. The primary objectives for including unproductive frictional designs in this study are to assess the potential negative effects they can have and to help analyze what characteristics of design frictions cause them to be unproductive. Many common types of undesirable design frictions could be used as unproductive obstructions in a game, such as extra clicks to perform an action.

These kinds of unwanted design frictions typically decrease efficiency, accuracy, and engagement of the user [116]. Although games typically include intentional obstructions and obstacles to create entertaining challenges, frictions in the wrong areas can merely impede the player without improving the player experience. In theory, unproductive obstructions in the game controls or user interface would frustrate players and inhibit them from reaching optimal game performance. For example, having to sit through an extended dialogue sequence before being able to purchase an item in a store could be seen as being an unproductive obstruction if it did not provide any valuable information to the player.

3.2.3 Productive Juice

The third frictional design is *productive juice*. As discussed in Chapter 2, "juice" or "juiciness" refers to "the emotional imprint of a responsive game through abundant feedback" [7]. Juicy design is considered by many game developers to be an essential

element for creating engaging games [43, 64]. However, juiciness can sometimes lead to an overabundance of feedback that can be distracting or disruptive to gameplay. Proponents of juice claim it improves the overall player experience, and some theories suggest it may increase player competence levels as well [67]. This leads to the idea that juiciness could improve player performance or game outcomes.

Productive juice attempts to find the careful balance between engaging feedback and distracting / disruptive feedback. Too little juice may not have enough of an impact to make a difference in players' experiences, while too much juice could have a negative effect [72]. Common visual effects used for creating juicy design include particle animations (e.g. sparkles, explosions), energetic animations, and screen shaking. Audio is also a common component of juice. The choice of background music and sound effects when certain actions are taken can impact the game experience. For example, the signature "bling" noise when Mario collects a coin can be immensely satisfying for players.

3.2.4 Unproductive Juice

The final frictional design is *unproductive juice*. As mentioned in the previous section, juiciness theoretically can have either positive or negative effects. The unproductive juice design aims to create overwhelming amounts of feedback that leads to distraction, disrupted gameplay, worse player experience, and/or worse game performance. By testing unproductive juice alongside productive juice in this game, the hope is to discover the mechanisms and conditions in which juicy design becomes negative and unproductive for the player. There is UX and PX literature to suggest that too much juice could cause cognitive overload and result in worse player performance [67,71,72].

As with many elements of player experience, juiciness can also be quite subjective.

For example, *Dead Cells* [94] is an indie game that won many accolades including Best Action Game at The Game Awards 2018 [49] for its "perfect 'game feel"' [40] and visual mechanics. At the same time, some people considered its over-the-top visual effects and juicy design to be too distracting, which prompted one YouTube reviewer to coin the derogatory phrase, "diarrhea Christmas lights," in reference to all of the particle effects and distracting juice [48]. This subjectivity shows that juiciness is a complex and nuanced subject and that caution should be taken before making generalizations about its effects.

3.3 Development of Prototypes

3.3.1 Base Game Design

The base design (or control version) of *Brave New Word* contains the basic gameplay mechanics and user interface elements required for a functional game. Each of the other six "treatment" versions feature modifications that introduce frictions.

A round in *Brave New Word* lasts 2 minutes and 30 seconds. The game is set up with a board of randomized letters lined up in 9 rows by 6 columns. See Figure 3-1(a) for a screenshot of what a typical board in the game looks like. Throughout the game, the player selects a word from the board to play and receives points based on the rarity and length of that word. A word must be at least three letters long for it to be accepted by the game. Once a word is submitted, the letter tiles forming that word disappear, and the remaining tiles "fall" toward the bottom of the screen. New letter tiles replenish the board, and the player selects a new word to play. The game ends when the timer is finished.

To form a word, letters can be selected if they are adjacent to each other in

any direction, including diagonals. Letter tiles can only be used once in a single word (e.g. the word cannot use the same letter tile twice). The basic formula for scoring a word depends on the length and relative rarity of the word in common usage. For example, the word "THE" would have a low rarity and score, while a word like "ZANY" would score much higher, not just because it has one more letter but also because it is a less common word. It is important to note that unlike many other popular word games, the frequency of the word itself has a more significant impact on the score in *Brave New Word* than the rarity of the individual letters. For example, the word "ETUI" would have a very low score in a game like *Scrabble*, while it scores fairly high in *Brave New Word* because of the rarity of the word despite each letter being common. The relative rarity of the words was determined using a frequency dictionary datamined from Reddit posts and comments. The Reddit corpus was selected because it represented common modern vocabulary usage on the Internet and because it was easily accessible. The word list collected from Reddit was cross-referenced with an open-source word game dictionary to ensure that all words were valid English words.

In the control design, users are able to select multiple letters by simply swiping their finger (or their mouse in the case of the web version) across the screen. When lifting their finger off the screen, the game submits the currently selected word for play. This design is as close to frictionless as possible in that it minimizes the number of actions the player has to take to submit a word. Entries for friction are available at multiple points in this control that can increase the number of necessary actions, but it is difficult to imagine a more efficient method for selecting and playing a word in this context. Therefore, it is the logical choice for the base/control design of *Brave New Word*.

The base design also attempts to have a minimal amount of juice for a game of this

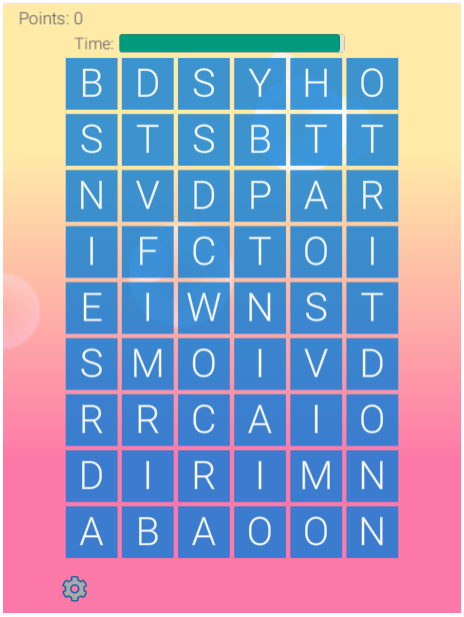
style while still making it feel like a polished, professional game. Calming, peaceful background music and nonintrusive sound effects combined with minor particle visual effects are the main juicy features in this design. Some extra juicy feedback and aesthetically pleasing features in a game may make it more exciting to engage with, but are unnecessary for the basic gameplay in this design.

3.3.2 Initial Frictional Features

The initial idea for a productive obstruction design was to add a button that must be pressed to submit the word. Although this introduced friction while selecting and submitting a word, it did not seem like an overwhelming burden for users considering that it was only one extra step. This is in line with the principles for designing for productive friction; the obstacle should not be too difficult for players to overcome. The productive goal of introducing friction in this case was two-fold: prompt reflection and direct the player's goal towards searching for rare words. The thought was that adding a button would create the microboundary necessary for players to use System 2 thinking and consider their options before submitting a word.

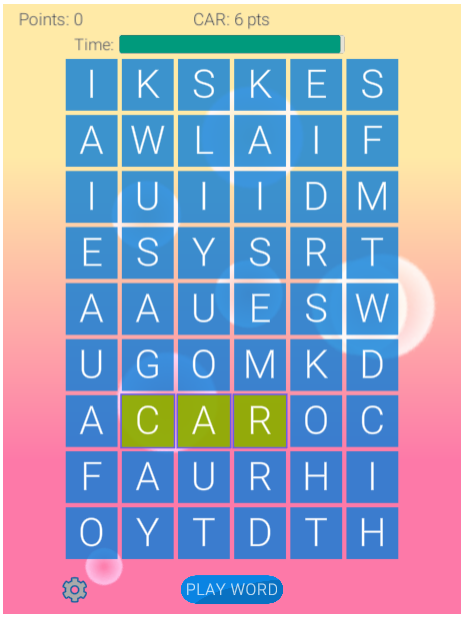
The first idea for a productive juice feature was to change the highlighted color of a selected word to reflect the rarity of the current selection. In this way, the juicy feedback (i.e. changing colors) would provide valuable information to the player who could make decisions based on that feedback. Since the objective of productive juice in this case was to improve player competency, the hypothesis was that highlighting the rarity of words would cause players to focus on that aspect, and thus play higher scoring words and improve their game performance.

No unproductive obstruction or unproductive juice idea was designed or implemented in the first prototype.

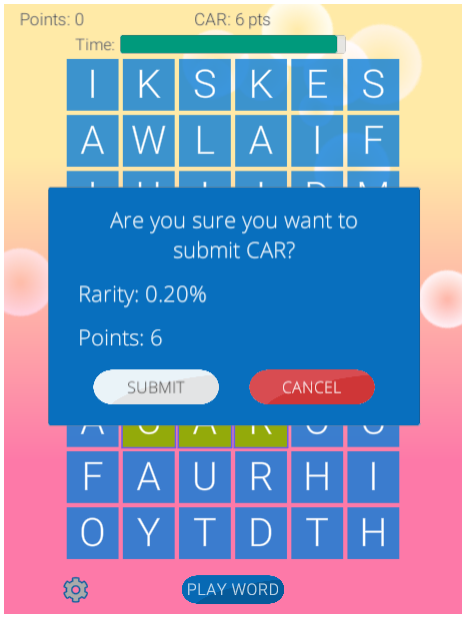


(a) The main game screen of the control version of *Brave New Word*.

(b) The word TAR selected in the control version.



(c) Button in the obstruction versions.



(d) Prompt window of productive obstruction version.

Figure 3.1: Control and obstruction versions of *Brave New Word*.

3.4 Final Design

3.4.1 Summary

In the final design of the game, several features were created for the productive and unproductive design frictions. The productive obstruction version included the original button design and added a prompt window that asked to confirm their choice before submitting. The prompt window also displayed the rarity of the word and the number of points the word was worth. See Figure 3.1 for screenshots of the control and obstruction versions of *Brave New Word*. In the vein of microboundaries [30], the prompt window was included to encourage the user to reflect on their choice. Also, the hope was that by displaying the rarity of the word, the player would align their gameplay goals to searching for rarer words. This introduced two extra steps for the action to be taken compared to the frictionless design, but this was considered simple enough to overcome that it had the potential to be "productive."

The new unproductive obstruction design idea was to take away the swiping motion as an action. To add friction, the player had to select each letter individually with a separate tap of their finger or click on the mouse. This design also included the button to submit the word, but it did not contain the reflective prompt window. Thus, this design objectively contained more steps to take the action of submitting a word than the productive frictional design. Since the prompt window was not included, none of the obstructions seemed to guide the player towards a productive outcome. The hypothesis was that a design with friction that is too frustrating to overcome would result in negative game outcomes and frustrate the user. In this case, friction was introduced by impeding the basic action of selecting a letter and by also adding a step before submitting a word. The key for this proposed unproductive

frictional design was that it did not incorporate any implicit productive goals for the user.

The design for productive juice included over-the-top visual and sound effects that matched the state of the game. While the control design had enough feedback to create a basic gaming experience, none of it was extreme enough to attract too much attention. The productive juicy design attempted to impede the user by being more distracting (e.g. large, colorful particle effects and screen shaking). However, the goal was to remain productive by not making it too difficult for the player to understand what was happening in the game. The semantics of the juice (e.g. the visual and sound effects) aligned with the game state, and the feedback was designed to make sense to the player without overwhelming them. For example, when the user played a valid word, the screen would shake, an explosion sound effect would trigger, and a colorful display of particle effects would flash across the screen. This may have been distracting or unnecessary during other parts of gameplay, but it served the purpose of providing positive feedback about the game state (i.e. successful word submitted). For additional details on the specific features included in each frictional version of the game, see Table 3.1.

The unproductive juicy design, on the other hand, consisted of distracting sound effects and visual effects that were too extreme or incongruous to the actual state of the game. Some effects provided too much feedback, such as sparkling particle effects and screen shaking whenever a letter was selected. In addition, selected letter tiles wiggled around on the screen, creating constant motion on the screen. Also, some of the juice was simply irrelevant to the game. By creating an experience where some of the visual and audio cues did not align with what was happening in the game, the game could confuse players and cause unproductive outcomes. This design included similar types of juicy feedback as the productive juicy design, but

Table 3.1: Frictional Features in Game Versions

Feature	P.O.	U.O.	P.J.	U.J.
Prompt window that displays score and rarity	X			
"Play Word" button	X	X		
Click to select each letter individually		X		
Upbeat background music			X	
Explosion and screen shaking when submitting a word			X	X
Screen shaking when submitting an invalid word			X	X
Sparkling particle effects in background			X	X
Selected letter tiles animate and move			X	X
Intense background music				X
More background particle effects				X
Particle effects when selecting letters				X
Screen shaking when selecting letters				X
Extra sparkling audio effects				X
Random foreground particle effects and laser sounds				X

P.O. = Productive Obstruction
U.O. = Unproductive Obstruction
P.J. = Productive Juice
U.J. = Unproductive Juice

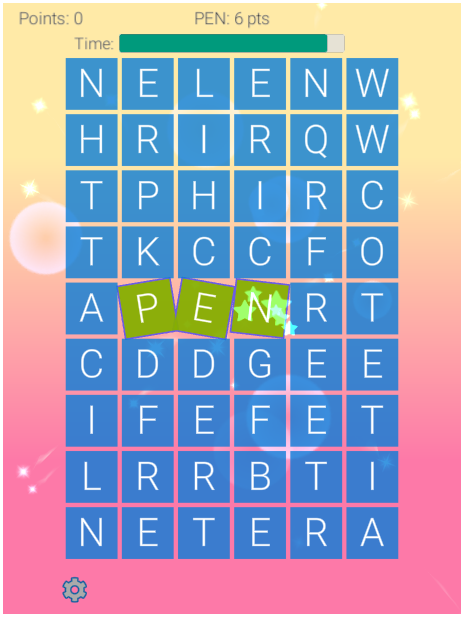
they were seemingly random or were extreme without providing much value to the user in terms of feedback. For example, at random points during the round, a flash of particles would stream across the screen and blast laser sound effects. The prediction was that the double layers of extreme distraction and confusing feedback would lead to an unproductive frictional design. See Figure 3.2 for screenshots of the juicy designs.

3.4.2 Platforms and Marketplaces

The game was developed for Android, iOS, and web systems and published publicly on Google Play and Apple's App Store platforms. The web version was published privately on the itch.io marketplace so that interview participants could play the game on their computers and easily share their screen while playing the game.

(a) Selected tiles moving in the juicy versions.

(b) Explosion particles when submitting a word in the juicy versions.



(c) Unproductive juicy version with extra particle effects.

Figure 3.2: Juicy versions of *Brave New Word*.

Chapter 4: Methodology

4.1 Overview

The main goal of this research was to increase the understanding of productive and unproductive friction in the designs of games by examining the effects of obstructions and juice in the context of a casual word game. This dissertation aimed to answer the following research questions about productive friction:

RQ 1: How do productive and unproductive design frictions in a casual word game influence players' behaviors and experiences?

RQ 2: What effects might productive and unproductive design frictions in a casual word game have on game outcomes?

This work investigated these questions by examining data collected from a game designed with productive and unproductive frictions. The game, *Brave New Word*, is a casual word search game developed for mobile and web browser platforms (e.g. tablets, smartphones, itch.io) specifically for this research study. Two main friction patterns were considered, obstructions and juice:

1. *Obstructions* on game actions, in which the user interface requires more steps than necessary to perform a given action.

2. *Juice*, which is potentially distracting visual and audio effects in the game that provide abundant feedback.

These designs were chosen primarily through studying the literature on UX and PX [96, 100] and finding characteristics of the game that could be modified in order to insert friction into the player experience. Through a combination of productive

and unproductive variations on obstructions and juice, a total of seven versions were created and studied.

This project was conducted in two parts: Study 1 collected and analyzed quantitative data from game logs, while Study 2 processed game log data as well as think-aloud sessions, interviews, and surveys from one-on-one sessions with participants. First, this chapter will review the methods required to measure player experience. Next, participant recruitment and the materials required for both studies will be discussed. Finally, a description of each study's design and procedures will be presented, along with an explanation of the data analysis methodology.

4.2 Measuring Player Experience

Being able to accurately measure player experience and the usability of video games has become increasingly valuable in both the commercial and academic worlds. As video game sales have continued to become more profitable in both the casual/mobile markets and the more traditional console/PC markets, developers have realized the value of gathering data on their players' game behaviors and experiences. Profitability and competition have driven industry practices for usability testing and accumulating accurate player data [39]. Separately, games (and specifically video games) have become a widely acknowledged tool for education in classrooms and informal spaces. This has led to much research on not only measuring the effectiveness of games on learning outcomes but also accurately measuring PX so that researchers can describe how and why learning occurs in games. In addition, the field of games studies, or ludology, has become increasingly popular and accepted as a serious scholarly pursuit. An important aspect of studying games is analyzing and measuring PX. Therefore, accurately defining and measuring PX has become

the goal of many people.

The term player experience draws its roots from the realm of user experience, which itself evolved from usability research in the fields of human-computer interaction (HCI) and human factors. Usability is defined as the experience of the user as they interact with the user interface of a product, whether it be software or something else [98]. Usability research is typically concerned with optimizing the interface for ease-of-use and clarity. UX encompasses usability in that it describes the holistic experience of the user with the entire software or product and not just the interface [100]. Research in the field of UX can be quite broad and requires a range of skills such as gathering qualitative and quantitative data, forming analyses, and designing creative solutions to user problems. PX is a specific subset of UX that is concerned with the user experience of games [96]. One might question why PX needs to be a separate term and why UX cannot be used for both games and non-games. The main reason is that video games typically have different standards for what is a satisfying experience than other software. While most software aims to make the user's experience as painless and accessible as possible, games usually incorporate challenges and problem-solving in order to provide a meaningful experience [29]. This contrast results in the area of PX which innately has distinct goals from UX research. These differences resulted in some researchers describing themselves with the more specific label of "player-computer interaction" (PCI) researchers.

This distinction of PX is especially relevant for this research on productive friction in games. Design friction implies that the usability of the game is not the smoothest experience possible, which goes against the goals of typical UX research [138]. However, this fits well within the bounds of PX research, where the objective is focused on the overall experience of gaming which includes intentional challenges and obstacles. Some commonly agreed upon constructs important to PX in the literature are

flow/immersion, enjoyment, ease of control, and challenge [60, 134].

The next sections will describe the following typical methods for measuring PX: (1) interviews, (2) surveys and questionnaires, (3) think-aloud studies, and (4) game analytics. For each method, there will be a brief description of the methodology and its place in the canon of UX/PX research. Then, there will be a summary of the strengths and weaknesses of each method.

4.2.1 Interviews

Interviews are a common qualitative approach for gathering insights from users. They lie on a spectrum of rigidity, from completely unstructured to strictly structured. Unstructured interviews are completely open-ended, and the interviewer will ask questions that follow the flow of conversation and not have any prepared questions to guide the discussion. Structured interviews, on the other hand, follow a script of questions and tend not to stray from the pre-written questions. In practice, many interviews are semi-structured and lie somewhere between the extreme ends of the spectrum with the researcher being free to go with the flow of the conversation as needed while having a set of prepared questions to guide the general direction of the interview [89, 124]. Directly asking participants questions and being able to follow up with additional questions provide interviewers with a greater degree of flexibility than more structured methods such as surveys or focus groups.

Interviews require participants to self-report their thoughts, opinions, and reflections on what happened during gameplay, and so they can introduce personal bias. For example, players might tell the interviewer what they think they want to hear, or players may misinterpret events that happened in the past due to faulty memory or otherwise clouded judgment. However, interviews can provide deep insight into

the thoughts of the players as they describe how and why they behaved in certain ways.

4.2.2 Surveys and Questionnaires

Surveys and questionnaires are widespread and commonly used to measure constructs by having participants respond to a number of items that relate to that construct. Surveys and questionnaires can be open- or closed-ended (or a combination of both); however, in our case, we will focus on closed-ended questionnaires that can be easily quantified. Questionnaires are common in user testing because of their ease of distribution and the minimal direct involvement required by a researcher. As with many things in life, this strength is also one of its weaknesses: the self-reported nature of surveys can lead to issues of questionable validity and reliability. If a questionnaire is too long or tedious, for instance, participants may begin to answer without much thought or care.

Survey design is a difficult process. There are many potential biases and pitfalls that must be avoided when designing questions. For a sample of such biases, see section 9.4.2 of Brühlmann and Mekler's chapter in *Games User Research* [15]. The subjective nature of the commonly used Likert scale is also much debated when researchers discuss its quantification of attitudes [1,65]. Considering the complexities of designing a survey from the ground up, and given that it is easier to compare across studies when they all use a standard questionnaire, it is expedient to always first consider existing surveys/questionnaires that measure the construct researchers are interested in.

Fortunately, there are many existing questionnaires and surveys related to PX that are widely used, such as the Game Experience Questionnaire (GEQ) [60], the

Player Experience of Need Satisfaction (PENS) [113], the Immersive Experience Questionnaire (IEQ) [62], and the Player Experience Inventory (PXI) [134]. The prevalence of questionnaires for measuring PX indicates that PX is multifaceted and that it is difficult for one questionnaire to successfully capture all aspects of PX that researchers may need [63]. Denisova et al. [33] and Johnson et al. [63] assessed the validity and accuracy of the most commonly used questionnaires—GEQ, PENS, among others—and concluded that there is still room for improvement to create a questionnaire that can consistently and accurately measure aspects of player experience. Vanden Abeele et al. [134] concurred and developed the PXI to address this need. PXI is still being validated, but many of the constructs they identified and developed are relevant to this study.

4.2.3 Think-Aloud Protocol

The think-aloud protocol is one of the most notable methods used for gathering insights from participants' thoughts as they use a product [97]. It was originally developed for researching the usability of products [86] and was used to great effect for identifying usability issues and getting insights into why users behave the way they do. Think-aloud studies are aptly named because the method requires the participant to use the product while simultaneously vocalizing what they are thinking. This is generally done in a laboratory setting as opposed to a more naturalistic or field environment since the researcher will want to record the audio and/or video of the session for future analysis. Think-alouds are useful because they make visible what is typically invisible and impossible to directly observe: the thought processes of a person.

Think-aloud protocols are typically cognitively demanding on the user as they

must "think out loud" while also playing the game or using the product. Because of this, researchers will sometimes make neutral statements while observing to encourage the participant to continue to think-aloud [81]. The difficulty of accurately thinking-aloud while using the product efficiently is a common critique of this protocol; the context of a think-aloud study is not representative of "typical" use of the product. However, the observation of a user struggling to think-aloud can also be a useful data point for the researcher. This generally implies that the action being done by the participant at that time required additional cognitive attention and made it difficult to think-aloud, which provides clues as to what parts of the task are more demanding than others [81].

Another common critique of think-aloud studies that also applies to interviews is that there is a risk of people misrepresenting their thoughts or feelings when it is self-reported. However, think-alouds are generally considered to contain less risk for misrepresentation than interviews because participants do not have as much time to reflect before giving their answers, and the data are more immediate and just-in-time.

4.2.4 Game Analytics

Game analytics, or game log data analysis, is another method of measuring player experience. This methodology is quantitative in nature and results in detailed information about what player behaviors actually took place. Game analytics refers to the gathering, processing, and analysis of telemetry data collected from players as they play the game [38]. Telemetry is defined as measuring something over a remote distance, so telemetry data technically refer to any data that are collected from players from a distance (e.g. not in a lab). Game log data can typically be gathered through online servers and databases and does not require another person

to be present with the player. While telemetry data generally refer to the raw data collected from players (e.g. clickstream data, log events for each action the player takes), gameplay metrics refer to a variety of interpretable measures calculated by summarizing the raw data [38]. Metrics can be simple statistics, like a basic aggregate or an average, but they can also be incredibly complex and nuanced formulas.

Game analytics involves more than simply gathering quantitative data and running statistical models. Rather, game analytics can refer to the entire process of setting goals and objectives for the analysis, cleaning and processing the data after collection, evaluating the results of statistical analyses, and communicating the results to other stakeholders (e.g. through creating visualizations or written reports) [38]. Researchers need to accomplish all of these tasks in order to successfully harness the power of game log data.

Unlike the methods of interviews and think-aloud studies examined above, game analytics requires no self-reporting from the player, and so it limits potential misrepresentation by the participant. However, the quantitative nature of the method does not shield it from bias; researchers must still decide what kinds of telemetry data to record and which to ignore. Those decisions can impact what kinds of analyses are performed and can be potential sources for bias as well.

4.2.5 Summary

All four methods mentioned above have their own strengths and weaknesses. In an ideal world, a game developer or researcher will not be limited to one method, and triangulating the data using multiple methods in conjunction would be the best approach. However, a researcher must always be prepared to know which methods will be most useful in answering the types of questions they are investigating. In

general, the more qualitative methods, such as interviews and think-aloud studies, are well-suited to answer "why" and "how" questions, while quantitative methods, such as game analytics, are best for answering "what" questions. For these reasons, a combination of methods including interviews, think-alouds, questionnaires, and game analytics were used in this project.

Interviews can answer why players did what they did and require reflection after gameplay sessions. The major downside to this approach is that players can, intentionally or unintentionally, misrepresent their thoughts or opinions in an interview for a variety of reasons, e.g. faulty memory, answering with what they think the interviewer wants to hear. Surveys are another method for procuring self-reported data on attitudes and opinions. The limitations to this method are similar to that of interviews; validity and accuracy are major concerns with self-reported data. Additionally, finding or creating accurate and reliable survey items is no easy feat. Think-aloud studies can mitigate this potential weakness by having participants verbalize their thoughts as they are playing the game. This minimizes the chances that players will edit their thoughts and be dishonest about what they are thinking and feeling at the time. The main critique of the think-aloud protocol is the high cognitive demand it places on subjects as they play the game. Attempting to verbalize thoughts out loud while also playing a game can result in suboptimal performance of both tasks. If the researcher is interested in gathering data on participants playing the game in an authentic manner, think-alouds may not be the best option. Game analytics, on the other hand, are best suited to answer questions about what actions players are actually engaging in. The challenge with game log data analysis is finding the right statistical model and the best way to process and interpret the potentially massive amounts of collected data. This is a challenge that is currently being addressed in the fields of data science and big data, and games researchers have much

to learn from those communities.

Combining qualitative and quantitative solutions to identify both what is happening and why players are doing what they do is a preferred strategy by many researchers. If a specific research question can be answered using only one method, then that would be appropriate; in most cases, however, triangulating a variety of data streams leads to more robust answers.

4.3 Study 1: Quantitative Study

4.3.1 Participants

For this study, players were recruited to download the game onto their mobile device through the Google Play Store or Apple App Store. No compensation or other form of incentive was provided for these participants.

Two hundred and thirty-eight players downloaded the game and played at least two rounds of the game ($n = 238$). However, when users who had not submitted a single word during their gameplay were filtered out, the number of participants shrank to $n = 216$. Players were randomly assigned[1] to one of seven condition groups (one control group and six frictional treatments): control ($n = 42$), productive obstruction ($n = 15$), unproductive obstruction ($n = 38$), productive juice ($n = 28$), unproductive juice ($n = 41$), productive obstruction & juice ($n = 29$), and unproductive obstruction & juice ($n = 23$).

The game was advertised on social media not through ad purchases but by general appeal to my personal networks and through public posts. The social media outlets

[1]Due to a bug in one of the updates to the game, the random distribution was not completely uniform throughout the data gathering process as can be seen in the resulting spread of participants. However, there were enough participants in each condition to continue with the study.

included Reddit, Facebook, Twitter, and Instagram. The advertisement posts were made between the dates of February 13, 2020 and March 27, 2020. The posts were worded to encourage people to participate in science and help progress research in video game design and learning. This populist framing of research has been successful in other science games such as Foldit [27] and EteRNA [85] that used crowdsourcing to solve challenging puzzles.

Public posts were shared on Facebook, Twitter, Instagram, and Reddit. Reddit is a social networking platform that allows users to "upvote" and "downvote" shared posts and comments, resulting in the most popular content rising to the top of the feed. Reddit was chosen specifically because of its easy access to niche communities through its many "subreddits," which are sub-communities generated by users on Reddit dedicated to specific topics. The advertising posts on Reddit were made in appropriate subreddits, such as /r/ludology (the study of games) and /r/playmygame (where people are encouraged to share their games). Reddit as a whole was the sixth most popular social media app in the United States [126] with over 52 million daily active users worldwide [111]. The social media posts and game descriptions on the app stores made it clear that the game was part of a research study in an attempt to encourage more downloads and participation. Participation in the study was voluntary, and all game logs collected were anonymous in nature. Players were not required to create an account or sign-in to play the game, so no personally identifiable data were collected in the logs.

Given the open nature of recruitment for this group of participants, it was essentially impossible to accurately predict or measure the demographic information of the players. No pre-surveys asked them to provide personal information in order to protect their anonymity and to provide as little friction into gameplay as possible. If most of the participants were gathered through my immediate personal social

media connections (e.g. Facebook friends, followers on Instagram), then some inferences and generalizations may be possible by analyzing the demographics of my social network. However, it may be the case that many of the participants were not in my immediate social network (e.g. friends of friends, strangers from Reddit or others seeing the public posts). Since how participants discovered the game was not tracked, it was not possible to record demographic information. In this case, the goal was to recruit as many participants as possible, and the demographics were not a concern.

4.3.2 Materials

This study required a video game with multiple frictional versions, game logs, and metrics derived from game log data. The game design along with frictional versions are described in Chapter 3. The following sections will describe these in more detail.

Game Logs and Metrics

Detailed game log data were gathered from the players who installed and played *Brave New Word* from the mobile markets. Game metrics were derived from the raw game log data.

Log Structure The game collected detailed logs about each player's gameplay. The logs contained telemetry (also known as clickstream) data about the actions taken by each player. Each log element consisted of two parts: (1) the metadata relevant to all logged data, such as the timestamp, game design version (i.e. the treatment group), etc. and (2) the actual meaningful data to be collected, e.g. the word submitted, word frequency, or the game board state. Descriptions of the log

Table 4.1: Summary of data elements collected in logs.

Element	Description
Design Version	The version of the game used, i.e. which type of frictional design was used in this game.
User ID	Identifies each unique user. Enables linking each log entry to a specific user.
Game ID	Identifies unique game sessions. Each game play session generates a new game ID.
Timestamps	The time at which specific actions occurred.
Word	The word that was submitted.
Word Score	The score of the word that was submitted.
Word Rarity	The rarity of the word that was submitted. The higher the number, the rarer it is. The rarity is expressed as a percentile.
Board State	The board state is a snapshot of all the letters currently on the board. The game logs the board state both pre- and post-word submission.

elements are summarized in Table 4.1.

Each log entry reflected an "event" in the game, such as selecting a letter, submitting a word, or pausing the game. Every log element contained metadata about the data they represent, and these metadata are important to help answer certain research questions. All of the log events were stored in a NoSQL cloud database using Google Firebase's Realtime Database. All players' data, regardless of operating system, were stored in the same database for easy processing and analysis. Afterwards, the data were retrieved and stored on a secure local server to be analyzed using various methods.

Metrics Note: the term "game time" refers to the amount of time the user has spent within one round of the game, and one "round" refers to each $2\frac{1}{2}$ minute long session of play before the game ends. The following metrics were used to explore the data and help answer this study's research questions:

1. **High_Score**: $\max_{all\ rounds} Game\ score$
   This metric measures the highest score ever achieved in a single round by a user.
2. **Avg_Score**: $\frac{\sum_{all\ rounds} Game\ score}{Total\ num.\ of\ rounds\ played}$
   This metric measures the average score achieved by the player in a round (sum of all scores divided by the number of rounds played).
3. **Wds_Per_Round**: $\frac{Total\ num.\ of\ words\ played}{Total\ num.\ of\ rounds\ played}$
   This metric measures the average number of words (successfully) submitted in a single round of the game.
4. **Avg_Word_Len**: $\frac{\sum_{all\ words} Length\ of\ word}{Total\ num.\ of\ words\ played}$
   This metric measures the average word length that is successfully submitted by the player.
5. **Max_Word_Len**: $\max_{all\ words} Length\ of\ word$
   This metric is a measure of the longest word that was ever played by the user.
6. **Avg_Rarity**: $\frac{\sum_{all\ words} Word\ rarity}{Total\ num.\ of\ words\ played}$
   This metric measures the average rarity of all the words successfully submitted by the player.
7. **High_Rarity**: $\max_{all\ words} Word\ rarity$

This metric measures the highest rarity of a submitted word achieved by the player.

8. **Avg_Single_Wd_Score**: $\frac{\sum_{all\ words} Word\ score}{Total\ num.\ of\ words\ played}$
   This metric is a measure of the average score of all the words submitted by the player.

9. **High_Single_Wd_Score**: $\max_{all\ words} Word\ score$
   This metric is a measure of the highest scoring word submitted by the player.

10. **Rounds_Played**: This metric is simply the total number of rounds played by the user.

4.3.3 Procedure

Telemetry data were collected from participants who downloaded the game from the two major mobile markets, the Google Play Store and the Apple App Store. These two markets were chosen because Google Android and Apple iOS together represent roughly 99% of the world's smartphone market share [125]. There was a description on the download pages of the game that explained the purpose of the research study and that anonymous game data would be collected for research purposes. Because no personally identifiable information was collected through the game logs, there was no screening process. Anonymous data were gathered because having the extra barrier of filling out a demographic survey before playing the game may have overburdened some users and resulted in fewer willing participants. It was decided that demographic variables such as age and gender were outside the scope of this study and prioritizing a larger sample size was preferable. The easy entry

into downloading and participating in the study likely contributed to the success of finding 200+ participants.

Users were allowed to play the game for as long and as often as they liked since log data were collected remotely. Game log data were collected for over 3 months from February 13, 2020 until May 28, 2020. Each user was randomly assigned to 1 of 7 treatment groups when they first installed and played the game.

There were 7 treatment groups for this study: 6 frictional design conditions of the game and 1 control design. Four of the frictional design conditions were the productive and unproductive versions of the *obstructions* design pattern and the productive and unproductive versions of the *juice* design pattern. The fifth condition was a combination of the two productive frictional designs, and the sixth frictional design treatment was a combination of the two unproductive frictional designs. The control design was a version of the game that has none of the frictional design patterns.

To ensure that the study was a between-subjects experiment, participants were not allowed to switch condition groups easily after they installed the game. There was a vulnerability in that users could work around the system by playing the game on multiple devices or uninstalling and re-installing the game on an Apple device (the Google Play Store API made it possible to prevent this on Android devices). However, if this vulnerability had been exploited by many players, the logs would have most likely been filled with many instances of users who installed the game but did not play through an entire round of the game in order to cheat the system until they found a version of the game they preferred. This was not the case as fewer than two dozen users out of over 320 logged users did not finish a round before quitting.

The independent variables in this study were the condition group of the game (i.e. the version of the game the participants played) and *Rounds_Played*. The dependent

variables were *Avg_Rarity*, *High_Rarity*, *Avg_Word_Len*, *Max_Word_Len*, and *Wds_Per_Round*.

4.3.4 Data Analysis

To gain insights from the game logs, the raw game log data were first processed to generate the game metrics described in the *Metrics* subsection above. This was done in Python with the help of several modules including *pandas* [130] and *NumPy* [104]. Then, I explored the data using data exploration and visualization techniques before creating logistic regression models to determine the impact of the frictional design versions on players. The logistic regression models revealed which versions correlated significantly with the dependent variables. This helped address important aspects of RQ2 such as what effects frictional designs may have had on players.

4.4 Study 2: Qualitative Study

4.4.1 Participants

The second group was recruited to participate in a one-on-one think-aloud protocol along with an interview and survey.

For the think-aloud and interview sessions, 7 students from two courses in the Curriculum and Instruction department at the University of Wisconsin–Madison participated. Out of the 7 participants, 4 identified as men and 3 as women. Each participant participated in a think-aloud and interview session and completed a short survey. The students were recruited by an advertisement posted online by the instructor of each course. Every student who wanted to sign up was given a chance to participate. No students were denied the opportunity to participate for any reason.

Each participant was given a gift card worth $10 as an incentive to participate, but no other incentives were offered.

4.4.2 Materials

This study required an implementation of *Brave New Word* on a web platform, a think-aloud protocol, an interview protocol, and a post-task questionnaire. The following sections will describe these features in more detail.

Think-aloud Protocol

Before the participant began playing the game, I explained that they needed to verbalize their thoughts while playing the game (i.e. "thinking aloud"). If a participant became silent for too long as they considered their next move, they were prompted gently with phrases to motivate them to talk such as, "What are you thinking right now?", "What is your current strategy?", or other similar questions. As the players talked, follow up questions were asked to probe for deeper insight. In this way, the think-aloud protocol often morphed into a semi-structured interview.

Interview Protocol

One-on-one, semi-structured interviews were conducted with the 7 participants in order to gather insights about players' interactions with the game. The interview questions were interspersed throughout the session between the think-aloud gameplay sessions. The semi-structured approach was chosen so that all potentially interesting avenues of questioning could be explored [89, 124]. These interview questions were blended in with the overall think-aloud protocol with questions being asked as interesting elements were brought up during the think-aloud. There was not a

strict boundary between a “think-aloud session” and the “interview session”. In most cases, these two protocols were merged into a single seamless session. Some prepared questions on the list of interview questions included the following:

1. Describe an experience from the game you found particularly engaging or frustrating.
2. How did you react to this situation?
3. How did X feature make you feel?
4. Why did X feature make you feel that way?

Since the interviews were semi-structured and intermingled with the think-aloud protocol, their lengths varied from person to person.

Post-Task Questionnaire

Basic demographic information including gender and previous gaming experience was collected from players through a post-task questionnaire distributed online during the one-on-one study sessions. The aim of the questionnaire was to gauge the participant’s experience with video games and what their typical goals were when playing a game like *Brave New Word*. Gender was included because some early hypotheses wondered if it may be a significant factor in terms of goals or behavior. The following items were included in the questionnaire (see Appendix A for full details):

1. What best describes your gender?
2. How often do you play video games?
3. When I play puzzle/word games similar to "Brave New Word", I like to (choose as many options as you want):

4. How often do you play similar puzzle/word games (i.e. Candy Crush, Boggle, Wordscapes, Words with Friends, etc.)?

4.4.3 Procedure

This study used the same 7 condition groups for the game as in Study 1. This section describes the procedures for the think-aloud sessions, interviews, and survey collection. For the one-on-one qualitative sessions, participants played through all versions of the game in a semi-random order so that the number of participants in each condition were distributed evenly.

Seven participants were recruited to participate in this qualitative study using video conferencing technology. Originally, the study sessions were supposed to take place in-person in a lab room. However, due to the onset of the COVID-19 pandemic in the region on March 2020, these face-to-face meetings were canceled in favor of remote video study sessions. After the study was advertised in two classrooms, potential participants signed up by submitting their email addresses through an online form. The potential participants were sent a follow up email asking to schedule a time to video chat for the study. Some participants did not respond after the initial email. These participants were sent one additional follow up email if they did not respond to the initial communication. The study sessions with the seven participants were conducted between April 7, 2020 and April 21, 2020.

Each session lasted about 30-45 minutes. At the start of the video meeting, the participant was sent an electronic consent form and given a chance to hear a summary of the study and walked through the informed consent process. The participant at that point could decide to continue with the study by signing the electronic consent form and returning it or decline to participate. They were then

given a brief description of the think-aloud protocol [86] and asked to verbalize their thoughts as they played the game for 7 rounds, switching versions of the game between each round.

There are different approaches to the think-aloud protocol, each with its own set of advantages and disadvantages. Two common think-aloud methods are the concurrent think-aloud and retrospective think-aloud [4]. Concurrent think-alouds require users to verbalize their thoughts as they interact with the game, leading to real-time results. Conversely, retrospective think-alouds allow participants to complete their tasks in silence first before asking them to speak their thoughts out loud. Concurrent think-alouds can distract the user while they interact with the system, which may make their gameplay nonrepresentative of the general population. However, waiting until the end of the interaction to verbalize their thoughts can result in users forgetting about issues and details that they may have observed if done concurrently. The concurrent think-aloud protocol was chosen for this study because of the existing evidence of its effectiveness and cost efficiency compared to the retrospective think-aloud method [4].

Semi-structured interview questions [89, 124] designed to elicit responses about how the design of the game influenced their behaviors and attitudes were intermingled as they played through the rounds. The order in which the participant played the 7 different versions of the game was randomly determined before the session began.

Recordings were made of the audio and the game screen as the participant played rounds of the game while thinking aloud and answering questions. The participant shared their screen as they played the game through a web browser. The participants were given a link to a short survey afterwards to collect information about their demographics and experience with gaming and word games (see Appendix A). The participants filled out the surveys immediately following the gaming and interview

sessions.

4.4.4 Data Analysis

This study generally follows the qualitative data analysis methods outlined by Miles, Huberman, & Saldaña [92]. First Cycle qualitative coding (FCQC) and Second Cycle qualitative coding (SCQC) stages were used to label and find patterns among the verbal data from the interviews and think-alouds. Each qualitative study session with the participants was summarized so that they could be analyzed for in-depth insights for each particular player and general patterns within the group.

Qualitative Coding Overview

Before qualitative coding began, the data from the interview / think-aloud sessions were transcribed through the automatic transcription technology built-in to the video conferencing software. Because the auto-transcription service was not 100% accurate, the transcriptions were manually reviewed and corrected. The transcription data included labels of the speakers and video timestamps marking when the interviewer and participant spoke.

A combination of qualitative coding methods described by Miles et al. [92] were used during FCQC, including descriptive coding, in vivo coding, and emotion coding. As recommended in their handbook, an initial deductive code list (see Table 4.2) was created using the theories and hypotheses on hand. While analyzing the data, inductive codes were added to the list. Several qualitative coding passes of the entire dataset were made while revising the code list before settling on the final list of selected codes. Some codes were removed due to their infrequent occurrence while others were merged. Any codes with fewer than 3 occurrences were removed from the

Table 4.2: Initial Deductive Code List

Code	Definition	Rationale
OBS:BUTTON	Mentions of the *Play Word* button obstruction.	Obstruction feature being tested. This frictional feature was included in both the productive and unproductive versions.
OBS:PROMPT	Mentions of the submit word prompt window.	Obstruction feature being tested. This friction was only in the productive version.
OBS:SELECTION	Mentions of the frictional selection controls where players have to click on letters one-by-one.	Obstruction feature being tested. This friction was only in the unproductive version.
JUICE:AUDIO	Mentions of the sound effects or background music.	Juicy feature being tested. Music and sound effects varied between productive and unproductive juice versions.
JUICE:SHAKE	Mentions of when the screen shakes when submitting a word.	Juicy feature being tested. This feature was present in both productive and unproductive versions.
JUICE:SHAKE UN	Mentions of when screen shakes when selecting each letter.	Juicy feature being tested. This feature was only in unproductive juice.
JUICE:BG SPARKLES	Mentions of the background sparkle effects that emanate from the center of the screen.	Juicy feature being tested. This feature was present in both productive and unproductive versions.
JUICE:TILES MOVING	Mentions of the effect where tiles jiggle when selected	Juicy feature being tested. This feature is present only in the unproductive version.
JUICE:FG SPARKLES	Mentions of the occasional effect where streams of sparkly particles streak across the screen.	Juicy feature being tested. This was only in the unproductive version.
DISTRACTION	Mentions of being distracted while playing the game.	Perception statement. Distraction can be a negative reaction to juicy design elements.
FRUSTRATION	Mentions of frustration with the game or its features.	Emotion statement. Frustration can be a negative indicator for player experience.
ENJOYMENT	Stating that they are having fun, enjoying the game, liking the feeling of a feature, etc.	Emotion statement. Enjoyment can be a positive marker for "fun" and "flow."
REFLECTION	Mentions of reflecting or pausing before taking an action.	Action statement. Player action of interest related to obstructions.
PREFERENCE (+/-)	Mentions of preferring a version / feature over another version / feature. + is a positive preference, - is not preferring it.	Value statement. The user's preference implies an overall positive player experience.

list, and these included FCQC codes such as REFLECTION and SATISFACTION. Afterwards, the FCQC codes were quantified by participant and by game version in order to highlight any potential patterns.

As part of the SCQC stage as recommended by Miles et al., groups of FCQC codes were organized into categories that emerged from the data. These categories were used to help analyze the relationships between frictions and various outcomes.

4.4.5 Ethical Considerations

All participants recruited for this qualitative study were adults (18+) and were required to sign electronic informed consent forms (see Appendix B). Identifiable data were removed from transcripts and were replaced with pseudonyms which are used throughout this document. All data collected for this study, including transcriptions, were stored on IRB-approved secure servers.

4.5 Data Map

A data map was created to map how all of the data collected could help to answer the relevant research questions. See Table 4.3 for details.

Table 4.3: Mapping data to research questions

	Research Question	**Data Collected**	**Sample Analysis Question**	**Metrics Used**
1	How do productive and unproductive design frictions in a casual word game influence players' behaviors and experiences?	• Think-alouds • Interviews • Surveys	Which attitudes and feelings are most reported by players after playing a game with a productive frictional design element?	• Number of pauses • Counts of codes • Overall performance • Overall player experience
1a	How do productive and unproductive obstructions impact players' behaviors and experiences?	• Think-alouds • Interviews • Surveys	How many times does the participant mention pausing or reflecting on their actions?	• Number of pauses • Overall performance • Overall player experience
1b	How does productive and unproductive juice impact players' behaviors and experiences?	• Gameplay logs • Think-alouds • Interviews • Surveys	What frictional points do players refer to in their interviews as being memorable or impactful?	• Avg word length • Max word length • Avg rarity • High rarity • Word per round • Overall player experience
2	What effects might productive and unproductive design frictions in a casual word game have on game outcomes?	• Gameplay logs • Think-alouds • Interviews	Do productive frictions result in higher game scores?	• Word score • Avg rarity • Avg word length • Overall performance
2a	What effects might productive and unproductive obstructions have on game outcomes?	• Gameplay logs • Think-alouds • Interviews	Does the prompt window affect the average rarity of the words?	• Word score • Avg rarity • Avg word length • Overall performance
2b	What effects might productive and unproductive juice have on game outcomes?	• Gameplay logs • Think-alouds • Interviews	Does productive juice increase the overall game performance?	• Word score • Avg rarity • Avg word length • Overall performance

Chapter 5: Study 1 Results

5.1 Overview

Research question 2 asked, "What effects might productive and unproductive design frictions in a casual word game have on game outcomes?" The hypothesis was that productive frictional designs would result in increases in game outcomes while unproductive frictional designs would result in negative player experiences and game outcomes. The results from the game log data provided insight into how frictional designs impacted players' behaviors and game outcomes.

Table 5.1 shows the descriptive statistics of the independent and dependent variables chosen for the logistic regression models. The independent variables chosen were *Rounds_Played* and *Version*. *Rounds_Played* was selected because it was an intuitive variable that could potentially serve as a proxy for the player's skill or expertise level (assuming that players grew in their skill level the more they played the game).

The dependent variables chosen for analysis were *Avg_Rarity*, *High_Rarity*, *Avg_Word_Len*, *Max_Word_Len*, and *Wds_Per_Round*. *Avg_Rarity* is the average rarity value of all the words successfully submitted by the player. The rarity values can range from 0.0–0.79 and if multiplied by 100 are percentiles representing the relative rarity of that word. *High_Rarity* is simply the highest rarity value ever scored from a single word by the player. *Avg_Word_Len* is the average number of letters in each word successfully submitted. Similarly, *Max_Word_Len* is the number of letters in the longest word ever submitted by the player. *Wds_Per_Round* is the average number of words that were successfully submitted by the player in each

Table 5.1: Descriptive Statistics for Selected Variables

Variable	Control	P.O.	U.O.	P.J.	U.J.	P.O.J.	U.O.J.	Overall
Avg_Rarity								
M	0.078	0.096	0.080	0.077	0.068	0.085	0.080	0.079
SD	0.025	0.031	0.027	0.025	0.031	0.026	0.057	0.032
Median	0.072	0.095	0.081	0.071	0.065	0.083	0.071	0.076
High_Rarity								
M	0.559	0.586	0.469	0.509	0.436	0.486	0.481	0.497
SD	0.227	0.243	0.223	0.240	0.252	0.227	0.258	0.239
Median	0.541	0.790	0.422	0.511	0.403	0.422	0.422	0.446
Avg_Word_Len								
M	3.504	3.826	3.688	3.540	3.519	3.702	3.672	3.611
SD	0.315	0.401	0.389	0.315	0.314	0.394	0.317	0.356
Median	3.473	3.838	3.646	3.509	3.495	3.685	3.662	3.552
Max_Word_Len								
M	5.619	6.267	5.763	5.500	5.512	5.862	6.000	5.727
SD	1.103	1.668	1.283	1.000	1.121	1.356	1.243	1.225
Median	6.000	6.000	5.000	6.000	6.000	6.000	6.000	6.000
Wds_Per_Round								
M	15.654	10.346	12.499	12.877	12.706	9.280	11.931	12.559
SD	6.458	5.217	7.554	4.310	6.027	3.956	6.218	6.193
Median	16.125	10.500	11.159	11.875	13.875	8.000	11.333	11.542
Rounds_Played								
M	82.500	48.267	39.816	23.214	45.463	119.621	52.696	59.708
SD	264.812	102.877	135.118	31.426	160.220	539.850	117.624	249.712
Median	10.500	7.000	5.500	8.000	7.000	4.000	6.000	7.000

Notes: M = mean. SD = standard deviation.
P.O. = Productive Obstruction
U.O. = Unproductive Obstruction
P.J. = Productive Juice
U.J. = Unproductive Juice
P.O.J. = Productive Obstruction & Juice
U.O.J. = Unproductive Obstruction & Juice

Table 5.2: Logistic Regression Model Differences

Variable	Log-likelihood of model without *version*	Log-likelihood of model with *version*	LLR χ^2 p-value
Avg_Rarity	−146.595	−140.075	< 0.001
High_Rarity	−99.951	−92.638	< 0.001
Avg_Word_Len	−148.205	−140.700	< 0.001
Max_Word_Len	−124.184	−118.760	< 0.001
Wds_Per_Round	−139.393	−126.073	< 0.001

round of the game. These variables are described in the "Metrics" section in Chapter 4.3.2 as well.

Logistic regression models were built with both *Rounds_Played* and *Version* included and with only *Rounds_Played* included, and there were significant differences between the models, indicating that *Version* was a significant factor. This suggested that the various frictional designs did have a significant impact on game outcomes. For instance, productive obstruction correlated with a higher chance of high *Avg_Word_Len* (OR=4.857, p=0.018). See Table 5.2 for detailed results of the differences in the models and Table 5.8 for detailed results from the logistic regression model that used *Version*.

5.2 Logistic Regression Aggregate Results

Next, the frictional design conditions that impacted game outcomes were investigated. Using the *statsmodels* [121] and *scikit-learn* [108] libraries, several binomial logistic regression models for the data were built. A separate regression model was created for each of the 5 dependent variables: *Avg_Rarity*, *High_Rarity*, *Avg_Word_Len*, *Max_Word_Len*, and *Wds_Per_Round*. Each dependent variable was converted to a binary variable, where 0 indicated a "low" value and 1 indicated

Table 5.3: Variable Thresholds

Variable	Threshold / Median	Number Below Threshold (Low)	Number Above Threshold (High)
Avg_Rarity	0.076	108	108
High_Rarity	0.446	114*	102*
Avg_Word_Len	3.552	108	108
Max_Word_Len	6.000	158*	58*
Wds_Per_Round	11.542	108	108

*These variables contained many duplicates, thus the split by median did not result in equal splits.

a "high" value. The low-high bin cutoffs were determined by splitting the values at the median. The thresholds and number of users in each bin for each variable are shown in Table 5.3.

For the first set of logistic regression models, the independent variables consisted of *Rounds_Played* and a binary conversion of *Version* that separated the versions based on various categories. Table 5.4 shows the results of the regression model where the versions were separated by those that had any *obstruction* designs (productive obstruction, unproductive obstruction, productive obstruction & juice, and unproductive obstruction & juice). The variable *Obs_Agg* was set to 1 whenever *Version* was one of the four versions that contained the *obstruction* design and 0 otherwise. In the same way, Table 5.5 shows the results for the aggregate variable *Juice_Agg* that represents all of the versions that contained elements of *extra juice*. The third aggregate variable, *Productive_Agg*, included all of the versions that had productive frictional design elements (i.e. productive obstruction, productive juice, and productive obstruction & juice). Table 5.6 displays the results for the productive versions. The fourth and final aggregate variable, *Unproductive_Agg* similarly contained all of the versions that had unproductive frictional design elements (unproductive obstruc-

Table 5.4: Obstruction Versions Logistic Regression Results

Variable	Avg_Rarity	High_Rarity	Avg_Word_Len	Max_Word_Len	Wds_Per_Round
Intercept	−0.350	−1.539	−0.436	−1.469	0.237
odds ratio	0.705	0.215	0.647	0.230	1.267
z-statistic	−1.754	−4.972	−2.181	−5.981	1.102
p-value	0.079*	< 0.001***	0.029**	< 0.001***	0.270
Obs_Agg†	0.494	−0.255	1.039	0.749	−1.241
odds ratio	1.639	0.775	2.827	2.116	0.289
z-statistic	1.779	−0.730	3.667	2.362	−4.148
p-value	0.075*	0.465	< 0.001***	0.018**	< 0.001***
Rounds_Played	0.003	0.128	−0.001	0.001	0.014
odds ratio	1.003	1.137	0.999	1.001	1.015
z-statistic	1.788	5.447	−1.361	1.480	2.776
p-value	0.074*	< 0.001***	0.174	0.139	0.006***
LLR χ^2 statistic	9.438	99.404	16.940	8.685	38.793
LLR p-value	0.009***	< 0.001***	< 0.001***	0.013**	< 0.001***

† Versions with obstruction ($n = 105$). Versions without obstruction ($n = 111$).

*$p < .1$. **$p < .05$. ***$p < .01$.

tion, unproductive juice, and unproductive obstruction & juice). This variable was necessary because *Unproductive_Agg* was not simply the inverse of *Productive_Agg*; the control group was considered neither a productive nor unproductive frictional design. The results for this logistic regression is shown in Table 5.7.

Compared to versions of the game that did not contain any frictional obstruction designs, versions that did contain obstructions (including those with both obstruction and juice combined) appeared to result in higher *Avg_Rarity* (OR=1.639, $p = 0.075$), *Avg_Word_Len* (OR=2.827, $p < 0.001$), and *Max_Word_Len* (OR=2.116, $p = 0.018$). At the same time, they were correlated with a decrease in *Wds_Per_Round* (OR=0.289, $p < 0.001$). In other words, users playing the game with obstructions were correlated with submitting rarer and longer words, but they also played fewer words per round.

Game versions that included juice were negatively correlated with *Avg_Rarity* (OR=0.618, $p = 0.085$) and *High_Rarity* (OR=0.469, $p = 0.032$). This suggested that players in conditions with juice struggled to play words with higher rarity values.

Table 5.5: Juice Versions Logistic Regression Results

Variable	Avg_Rarity	High_Rarity	Avg_Word_Len	Max_Word_Len	Wds_Per_Round
Intercept	0.163	−1.279	0.138	−0.889	−0.131
odds ratio	1.177	0.278	1.148	0.411	0.877
z-statistic	0.759	−4.172	0.649	−3.891	−0.580
p-value	0.448	$< 0.001^{***}$	0.516	$< 0.001^{***}$	0.562
Juice_Agg†	−0.481	−0.757	−0.126	−0.333	−0.364
odds ratio	0.618	0.469	0.882	0.717	0.695
z-statistic	−1.720	−2.147	−0.456	−1.072	−1.266
p-value	0.085^{*}	0.032^{**}	0.648	0.284	0.205
Rounds_Played	0.002	0.129	−0.001	0.001	0.014
odds ratio	1.002	1.138	0.999	1.001	1.014
z-statistic	1.673	5.531	−1.384	1.534	2.572
p-value	0.094^{*}	$< 0.001^{***}$	0.166	0.125	0.010^{**}
LLR χ^2 statistic	9.229	103.561	3.238	4.109	22.265
LLR p-value	0.010^{**}	$< 0.001^{***}$	0.198	0.128	$< 0.001^{***}$

† Versions with extra juice ($n = 121$). Versions without extra juice ($n = 95$).

$^{*}p < .1$. $^{**}p < .05$. $^{***}p < .01$.

A caveat is that the p-value for *Avg_Rarity* was relatively high, and so there remains a good chance that the differences were insignificant for this particular variable. Nevertheless, the results for *High_Rarity* indicated that there was some relationship between juice and lower rarity compared to game versions without any juice. In plain terms, participants playing versions with juice submitted words with lower rarity values.

Versions that contained productive frictional designs showed slightly higher instances of high *Avg_Rarity* (OR=1.707, $p = 0.070$) and a high chance of low *Wds_Per_Round* (OR=0.386, $p = 0.003$). The relationship between productive friction and high *Avg_Rarity* should not be overstated because the p-value was relatively high. The evidence supporting a negative correlation between productive friction and *Wds_Per_Round* was more significant. This correlation appears to be mostly due to the relationship between *Wds_Per_Round* and obstruction versions of the game.

The aggregate variable for unproductive friction versions was significantly corre-

Table 5.6: Productive Versions Logistic Regression Results

Variable	Avg_Rarity	High_Rarity	Avg_Word_Len	Max_Word_Len	Wds_Per_Round
Intercept	−0.291	−1.856	−0.018	−0.984	−0.049
odds ratio	0.747	0.156	0.982	0.374	0.952
z-statistic	−1.645	−6.015	−0.103	−5.170	−0.260
p-value	0.100	$< 0.001^{***}$	0.918	$< 0.001^{***}$	0.795
Productive_Agg†	0.535	0.473	0.248	−0.267	−0.952
odds ratio	1.707	1.604	1.282	0.766	0.386
z-statistic	1.813	1.272	0.852	−0.792	−2.996
p-value	0.070^{*}	0.203	0.394	0.428	0.003^{***}
Rounds_Played	0.003	0.133	−0.001	0.001	0.015
odds ratio	1.003	1.142	0.999	1.001	1.015
z-statistic	1.837	5.474	−1.352	1.552	2.751
p-value	0.066^{*}	$< 0.001^{***}$	0.176	0.121	0.006^{***}
LLR χ^2 statistic	9.572	100.487	3.757	3.598	30.096
LLR p-value	0.008^{***}	$< 0.001^{***}$	0.153	0.165	$< 0.001^{***}$

† Versions with productive friction ($n = 72$). Versions without productive friction ($n = 144$).

$^{*}p < .1$. $^{**}p < .05$. $^{***}p < .01$.

Table 5.7: Unproductive Versions Logistic Regression Results

Variable	Avg_Rarity	High_Rarity	Avg_Word_Len	Max_Word_Len	Wds_Per_Round
Intercept	0.063	−1.222	−0.032	−1.311	−0.377
odds ratio	1.065	0.295	0.968	0.269	0.686
z-statistic	0.323	−4.229	−0.167	−5.643	−1.751
p-value	0.747	$< 0.001^{***}$	0.868	$< 0.001^{***}$	0.080^{*}
Unproductive_Agg†	−0.365	−1.159	0.207	0.477	0.100
odds ratio	0.694	0.314	1.230	1.610	1.105
z-statistic	−1.318	−3.105	0.755	1.526	0.352
p-value	0.188	0.002^{***}	0.450	0.127	0.725
Rounds_Played	0.003	0.137	−0.001	0.001	0.013
odds ratio	1.003	1.146	0.999	1.001	1.014
z-statistic	1.749	5.442	−1.354	1.615	2.547
p-value	0.080^{*}	$< 0.001^{***}$	0.176	0.106	0.011^{**}
LLR χ^2 statistic	7.994	109.245	3.600	5.309	20.778
LLR p-value	0.018^{**}	$< 0.001^{***}$	0.165	0.070^{*}	$< 0.001^{***}$

† Versions with unproductive friction ($n = 102$). Versions without unproductive friction ($n = 114$).

$^{*}p < .1$. $^{**}p < .05$. $^{***}p < .01$.

lated with a lower high score for rarity (OR=0.314, $p = 0.002$). The *High_ Rarity* variable is the highest rarity value ever achieved by the player, and so this finding indicated that players in the unproductive friction conditions were more likely to achieve lower *High_ Rarity* values than their peers in the productive or control versions of the game. *Unproductive_ Agg* was not found to be a significant factor in any of the other variables.

5.3 Logistic Regression Individual Results

The dependent variables remained the same as those chosen for the aggregated set of logistic regression models. The independent variables were *Rounds_ Played* and *Version*. In this set up, *Version* represented the 7 different treatment groups (i.e. productive juice, productive obstruction, etc.) as a categorical variable. For the purposes of implementation, binary dummy variables were created to represent the different versions. Only 6 of the 7 dummy variables were used in the regression models; the control version was left out to create a baseline and to avoid the "dummy variable trap" [128]. In Table 5.8 which shows the full results, the suffix $\_P$ indicates the productive version of a design while $\_U$ indicates the unproductive version. *Juice* is the variable for any version with *juice*, and *Obs* is an abbreviation for obstruction. *Obs_ Juice* is the variable for the versions that combined both *obstructions* and *juice* frictional elements.

Obs_ P was positively correlated with *Avg_ Word_ Len* (OR=4.857, $p = 0.018$) and negatively correlated with *Wds_ Per_ Round* (OR=0.142, $p = 0.007$). There were also correlations with *Avg_ Rarity* (OR=3.187, $p = 0.083$) and *Max_ Word_ Len* (OR=3.395, $p = 0.058$) that were not quite statistically significant but worth noting. Two variables, *High_ Rarity* (OR=0.286, $p = 0.029$) and *Wds_ Per_ Round*

(OR=0.319, $p = 0.019$), were negatively correlated with *Obs_ U*. *Avg_ Word_ Len* was positively correlated to *Obs_ U* with an odds ratio of 2.669 and a p-value of 0.035.

Juice_ P had no clear correlations with any of the 5 dependent variables; none of the models were statistically significant. *Juice_ U* was negatively correlated with *High_ Rarity*, and it was statistically significant (OR=0.204, $p = 0.007$).

Obs_ Juice_ P was negatively correlated with *Wds_ Per_ Round* (OR=0.079, $p < 0.001$). There were no other statistically significant relationships, but the positive correlation with *Avg_ Word_ Len* was worth noting with an OR=2.604 and $p = 0.058$. Based on the findings from the other individual variables, this was most likely the result from the obstruction design and not the juiciness. *Obs_ Juice_ U* was negatively correlated with *High_ Rarity* (OR=0.202, $p = 0.036$) and *Wds_ Per_ Round* (OR=0.250, $p = 0.016$). It was positively correlated with *Avg_ Word_ Len* and statistically significant (OR=3.328, $p = 0.028$). This also matched the patterns from the individual results and was unsurprising given the significant correlations found in *Juice_ U* and *Obs_ U*.

There are several interesting results to highlight. Although *Obs_ P* was the only *obstruction* version that was a significant factor in the models for *Avg_ Rarity* (OR=3.187, $p < .1$) and *Max_ Word_ Len*, the model that used the aggregate *obstruction* versions also showed that it was significant for those variables. The individual results clearly show that productive obstruction was the biggest factor in increasing the odds of achieving high *Avg_ Rarity* and *Max_ Word_ Len*. These findings help answer research question 2 by showing that there is a significant difference between productive obstruction and the other versions.

The individual results also supported the aggregate *obstruction* results for the variables *Avg_ Word_ Len* and *Wds_ Per_ Round*. All of the versions with *obstruction*

were a significant factor in increasing *Avg_Word_Len* and lowering the number of *Wds_Per_Round*. These results signaled that some game outcomes were more reliant on the existence of obstructions than on if the designs were meant to be productive or unproductive.

Juice_P was not found to be significantly correlated to any of the 5 dependent variables. This is not a surprising finding since the prevalent hypothesis regarding juiciness is that effective juice in a game will improve the overall quality of the player experience and make the game controls more intuitive [67, 101], which were not easily measurable with game logs. Other studies have shown that there is a possibility for juicy designs to lead to improved game performance [56, 72]. However, another hypothesis states that the redundant feedback in juice can increase cognitive load and lead to worse performance in games [67, 71]. With productive juice, there was no statistically significant evidence supporting the latter hypotheses, but with unproductive juice, there was a significant negative correlation with *High_Rarity*, suggesting that unproductive juice can lead to lower performance in terms of high rarity scores.

Table 5.8: Individual Versions Logistic Regression Results

Variable	Avg_Rarity	High_Rarity	Avg_Word_Len	Max_Word_Len	Wds_Per_Round
Intercept	−0.264	−0.913	−0.500	−1.408	0.598
odds ratio	0.768	0.401	0.607	0.245	1.818
z-statistic	−0.819	−2.164	−1.522	−3.629	1.682
p-value	0.413	0.030**	0.128	< 0.001***	0.093*
Obs_P†	1.159	0.402	1.580	1.222	−1.952
odds ratio	3.187	1.495	4.857	3.395	0.142
z-statistic	1.736	0.558	2.362	1.897	−2.686
p-value	0.083*	0.577	0.018**	0.058*	0.007***
Obs_U†	0.603	−1.250	0.982	0.709	−1.142
odds ratio	1.828	0.286	2.669	2.032	0.319
z-statistic	1.307	−2.185	2.110	1.379	−2.340
p-value	0.191	0.029**	0.035**	0.168	0.019**
Juice_P†	0.056	−0.601	0.095	−0.738	−0.804
odds ratio	1.057	0.548	1.100	0.478	0.448
z-statistic	0.112	−0.976	0.189	−1.022	−1.536
p-value	0.911	0.329	0.850	0.307	0.125
Juice_U†	−0.299	−1.587	0.107	0.221	−0.452
odds ratio	0.742	0.204	1.113	1.248	0.636
z-statistic	−0.656	−2.685	0.235	0.419	−0.944
p-value	0.512	0.007***	0.814	0.675	0.345
Obs_Juice_P†	0.646	−0.903	0.957	0.116	−2.535
odds ratio	1.907	0.405	2.604	1.123	0.079
z-statistic	1.287	−1.441	1.897	0.196	−4.015
p-value	0.198	0.150	0.058*	0.845	< 0.001***
Obs_Juice_U†	−0.724	−1.599	1.202	0.908	−1.386
odds ratio	0.485	0.202	3.328	2.479	0.250
z-statistic	−1.293	−2.095	2.202	1.582	−2.405
p-value	0.196	0.036**	0.028**	0.114	0.016**
Rounds_Played	0.003	0.135	−0.001	0.001	0.016
odds ratio	1.003	1.144	0.999	1.001	1.016
z-statistic	1.826	5.413	−1.328	1.588	2.913
p-value	0.068*	< 0.001***	0.184	0.112	0.004***
LLR χ^2 statistic	19.289	113.497	18.039	13.810	47.294
LLR p-value	0.007***	< 0.001***	0.012**	0.055*	< 0.001***

† Control ($n = 42$), Juice_P ($n = 28$), Juice_U ($n = 41$), Obs_P ($n = 15$), Obs_U ($n = 38$), Juice_Obs_P ($n = 29$), and Juice_Obs_U ($n = 23$)

*$p < .1$. **$p < .05$. ***$p < .01$.

Chapter 6: Study 2 Results

6.1 Qualitative Coding

To investigate how productive frictional designs could potentially support positive outcomes in players, think-alouds, interviews, and questionnaires were analyzed. To test whether *obstructions* elicited reactions such as reflection or if *juice* distracted players, participants' feedback related to design factors were collected and qualitatively coded.

First Cycle and Second Cycle qualitative coding stages [118] were conducted on the transcribed verbal data collected from the interviews and think-alouds. For a list of the generated codes and their respective definitions and rationales, see Tables 6.3, 6.4, 6.5, 6.6, and 6.7.

During the First Cycle qualitative coding (FCQC) phase, a combination of deductive and inductive codes were generated. The deductive codes were compiled from the list of frictional features that were of interest in this study, along with the conjectured player reactions to the frictional designs. The inductive codes were assembled through several passes through the verbal data. This coding process also included counting the occurrences of each code (see Tables 6.1 and 6.2).

Following Miles et al.'s recommendations on Second Cycle qualitative coding (SCQC), five categories based on the patterns from the FCQC codes emerged. These five categories related to the players' reactions and experiences with frictional design elements in the game. The categories concerned the friction designs of *obstruction* and *juice*, *actions*, *values*, *emotions*, and *perceptions*. Tables 6.3, 6.4, 6.5, 6.6 and 6.7 list all of these categories, subcodes, and their definitions.

Table 6.1: Code Counts by Participant

Code	Claudia	Toby	Sam	Josh	Donna	Charlie	Zoey	TOTAL
OBS:BUTTON	3	0	5	2	2	7	3	**22**
OBS:PROMPT	0	1	5	1	2	3	2	**14**
OBS:SELECTION	0	4	3	1	3	2	0	**13**
JUICE:AUDIO	0	10	0	8	2	3	3	**26**
JUICE:SHAKE	1	0	0	0	1	2	0	**4**
JUICE:SHAKE UN	3	0	6	1	1	0	0	**11**
JUICE:BG SPARKLES	3	1	1	0	0	2	1	**8**
JUICE:TILES MOVING	1	2	0	0	0	0	0	**3**
JUICE:FG SPARKLES	1	0	1	0	1	3	0	**6**
ACT:STRATEGY	12	10	4	0	3	3	6	**38**
ACT:DIAG SELECT	1	1	2	1	0	2	0	**7**
ACT:IGNORE JUICE	0	0	1	0	0	2	2	**5**
VAL:PREFERENCE+	4	4	5	0	4	4	2	**23**
VAL:PREFERENCE-	1	4	2	0	5	10	1	**23**
VAL:RARITY	1	1	1	0	0	1	0	**4**
VAL:LONGER WORDS	1	2	3	0	1	1	2	**10**
VAL:SHORTER WORDS	0	1	3	0	0	1	0	**5**
EMO:FRUSTRATION	0	2	0	0	1	0	0	**3**
EMO:ENJOYMENT	0	2	0	1	2	0	0	**5**
EMO:ANXIOUS	2	1	1	1	0	1	1	**7**
EMO:CALM	2	4	0	0	2	4	0	**12**
EMO:COMFORT	1	0	1	0	0	0	4	**6**
EMO:CONFUSION	0	1	0	3	0	0	2	**6**
EMO:RELAXATION	0	1	0	0	2	0	1	**4**
PER:DISTRACTION	6	2	1	7	1	7	1	**25**
PER:EASY	1	1	2	1	0	1	1	**7**
PER:HARD	0	0	4	16	5	2	6	**33**
PER:FAST	2	2	4	0	6	0	2	**16**
PER:SLOW	2	3	10	8	0	0	4	**27**
PER:MISMATCH	0	2	0	1	0	3	0	**6**
PER:MATCH	0	4	0	1	0	3	0	**8**

Table 6.2: Code Counts by Version

Code	Control	P.O.	U.O.	P.J.	U.J.	P.O.J.	U.O.J.	TOTAL
OBS:BUTTON	3	5	4	4	1	3	0	**20**
OBS:PROMPT	2	5	1	0	0	6	0	**14**
OBS:SELECTION	0	1	5	0	1	0	5	**12**
JUICE:AUDIO	4	1	1	3	6	6	2	**23**
JUICE:SHAKE	0	1	0	2	0	0	0	**3**
JUICE:SHAKE UN	1	0	0	0	7	0	3	**11**
JUICE:BG SPARKLES	1	1	0	1	2	0	2	**7**
JUICE:TILES MOVING	0	0	0	2	0	1	0	**3**
JUICE:FG SPARKLES	0	1	0	0	2	0	2	**5**
ACT:STRATEGY	4	8	4	9	5	2	6	**38**
ACT:DIAG SELECT	1	0	0	3	2	1	0	**7**
ACT:IGNORE JUICE	0	1	0	0	2	0	1	**4**
VAL:PREFERENCE+	7	0	3	3	3	2	2	**20**
VAL:PREFERENCE-	1	5	5	2	4	2	2	**21**
VAL:RARITY	0	0	1	0	0	2	1	**4**
VAL:LONGER WORDS	1	2	4	2	0	1	0	**10**
VAL:SHORTER WORDS	1	1	3	0	0	0	0	**5**
EMO:FRUSTRATION	0	0	2	0	0	0	1	**3**
EMO:ENJOYMENT	0	0	1	0	4	0	0	**5**
EMO:ANXIOUS	1	0	1	0	2	1	1	**6**
EMO:CALM	8	0	2	0	0	0	0	**10**
EMO:COMFORT	2	3	0	0	0	0	1	**6**
EMO:CONFUSION	1	0	1	0	0	2	2	**6**
EMO:RELAXATION	1	1	0	0	1	0	0	**3**
PER:DISTRACTION	1	4	1	4	10	0	4	**24**
PER:EASY	1	2	1	2	0	0	1	**7**
PER:HARD	2	0	8	7	6	2	8	**33**
PER:FAST	2	2	4	4	2	2	0	**16**
PER:SLOW	0	6	7	0	2	4	8	**27**
PER:MISMATCH	0	1	1	0	1	0	2	**5**
PER:MATCH	1	1	1	0	1	0	2	**6**

P.O. = Productive Obstruction
U.O. = Unproductive Obstruction
P.J. = Productive Juice
U.J. = Unproductive Juice
P.O.J. = Productive Obstruction & Juice
U.O.J. = Unproductive Obstruction & Juice

Table 6.3: Friction Code Definitions

Code	Definition	Rationale
OBS:BUTTON	Mentions of the *Play Word* button obstruction.	Obstruction feature being tested. This frictional feature was included in both the productive and unproductive versions.
OBS:PROMPT	Mentions of the submit word prompt window.	Obstruction feature being tested. This friction was only in the productive version.
OBS:SELECTION	Mentions of the frictional selection controls where players have to click on letters one-by-one.	Obstruction feature being tested. This friction was only in the unproductive version.
JUICE:AUDIO	Mentions of the sound effects or background music.	Juicy feature being tested. Music and sound effects varied between productive and unproductive juice versions.
JUICE:SHAKE	Mentions of when the screen shakes when submitting a word.	Juicy feature being tested. This feature was present in both productive and unproductive versions.
JUICE:SHAKE UN	Mentions of when screen shakes when selecting each letter.	Juicy feature being tested. This feature was only in unproductive juice.
JUICE:BG SPARKLES	Mentions of the background sparkle effects that emanate from the center of the screen.	Juicy feature being tested. This feature was present in both productive and unproductive versions.
JUICE:TILES MOVING	Mentions of the effect where tiles jiggle when selected	Juicy feature being tested. This feature is present only in the unproductive version.
JUICE:FG SPARKLES	Mentions of the occasional effect where streams of sparkly particles streak across the screen.	Juicy feature being tested. This was only in the unproductive version.

Table 6.4: Action Code Definitions

Code	Definition	Rationale
ACT:STRATEGY	How they search for words on the board, mentions of their strategy in playing the game, etc.	Action statement. Player behavior or strategy changes are important to note.
ACT:DIAG SELECT	Mentions of the diagonal selection control system.	Action statement. Selecting letters diagonally was an important factor in how difficult or frustrating the controls could be.
ACT:IGNORE JUICE	Mentions of ignoring the juicy effects to play the game.	Action statement. How players responded to juice was an item of interest.

Friction Qualitative Codes

In terms of the friction codes (Table 6.3), each of the descriptors were derived from the game design features discussed in Chapter 3 (see Table 3.1). Each participant was asked about the frictional features in the various versions, so an obstruction (OBS) or juice (JUICE) code was assigned in the first coding pass. The specific feature was further labeled through subcodes such as BUTTON or PROMPT. These features were of primary concern for this study's research questions and any mentions of the feature were labeled with the appropriate code. The friction codes helped identify which features were related to the game outcomes being measured.

Action Qualitative Codes

Action codes (Table 6.4) labeled any references to specific actions that the player made during gameplay. For instance, a deductive action code that was created in the initial list was REFLECTION, since one of the predicted outcomes of productive obstruction was an increase in mindful interactions. The other action codes were created inductively during successive coding passes. For example, the code DIAG SELECT refers to when participants mentioned the act of diagonally selecting let-

Table 6.5: Value Code Definitions

Code	Definition	Rationale
VAL:PREFERENCE (+/-)	Mentions of preferring a version / feature over another version / feature. + is a positive preference, - is not preferring it.	Value statement. The user's preference implies an overall positive player experience.
VAL:RARITY	Mentions of looking for rare words or more uncommon words.	Value statement. Looking for rare words is one strategy to achieve success in the game.
VAL:LONGER WORDS	Mentions of searching for longer words.	Value statement. Playing longer words is another strategy to achieve success in the game.
VAL:SHORTER WORDS	Mentions of making smaller words.	Value statement. Another strategy was to play shorter words so that the player could play faster.

ters. Players considered the act of diagonal selection to be more frustrating in some versions and would bring up the topic when asked about their feelings about the ease-of-control of the game. These action codes provided insight into what types of behaviors occurred in relation to the game versions.

Value Qualitative Codes

Value codes represented any statements made by the participant expressing an attitude or belief about their game preferences or strategy. Players were prompted to speak about their preferred versions or features and their favored strategies, and their responses generally fit into this category. For example, the RARITY code tagged utterances that referenced the desire to form rarer words or recognized that doing so would result in higher scoring words. Similarly, LONGER WORDS was the code for any mentions of looking for longer words. Generally, the term "longer" applied to any word longer than the minimal length for a word, 3 letters. These values helped identify whether players changed their attitudes or behaviors throughout the study.

Table 6.6: Emotion Code Definitions

Code	Definition	Rationale
EMO:FRUSTRATION	Mentions of frustration with the game or its features.	Emotion statement. Frustration can be a negative indicator for player experience.
EMO:ENJOYMENT	Stating that they are having fun, enjoying the game, liking the feeling of a feature, etc.	Emotion statement. Enjoyment can be a positive marker for "fun" and "flow."
EMO:ANXIOUS	Verbalizing their anxious feelings arising from the game, usually in relation to time pressure.	Emotion statement. Anxiety can be both a positive and negative marker for "fun" and "flow."
EMO:CALM	The game experience feels calm to the player.	Emotion statement. Calmness could be a positive indication of "immersion" / "flow."
EMO:COMFORT	In relation to controls feeling nice, smooth, comfortable. Or comfort level with the game??	Emotion statement. Comfort could be a positive marker for "ease-of-use" and "flow."
EMO:CONFUSION	The player stated that they felt confused while playing the game.	Emotion statement. Confusion is a negative indicator for "fun" and "flow."
EMO:RELAXATION	Mentions of relaxation in relation to the game.	Emotion statement. Relaxation is a positive marker for "flow" and "ease of control."

Emotion Qualitative Codes

The emotion category marked codes concerned with players' emotional responses to the game. In particular, ANXIOUS was an emotion code that occurred frequently around the issue of time, or rather the lack thereof. Because players only had two minutes and thirty seconds to find words in each round, some participants expressed anxiety given the time pressure. Another instance of an emotion code was CALM. When a player communicated that they were feeling calm due to the music, controls, or other aspects of the game, this was labeled as CALM. These emotions were key to understanding the participants' PX of each version of the game.

Table 6.7: Perception Code Definitions

Code	Definition	Rationale
PER:DISTRACTION	Mentions of being distracted while playing the game.	Perception statement. Distraction can be a negative reaction to juicy design elements.
PER:EASY	Stating that the game is easier to play concerning the controls or mechanics.	Perception statement. Easiness of the game can indicate positive "ease of control" and "flow."
PER:HARD	The game felt difficult to play for them, either in terms of mechanics or controls.	Perception statement. Difficulty can be negative marker for "ease of control" and "flow" but could be either a positive or negative marker for "fun."
PER:FAST	The player feels like they can play the game faster.	Perception statement. The speed could both be a positive or negative marker for "fun" or "flow."
PER:SLOW	The player feels like the game has been slowed down.	Perception statement. The speed could both be a positive or negative marker for "fun" or "flow."
PER:MISMATCH	The feedback did not match the mood or feelings of the game.	Perception statement. The mismatch in the feedback could cause confusion or distraction for the player.
PER:MATCH	In contrast to MISMATCH, the feedback matches the mood or feelings of the game.	Perception statement. The matching feedback could contribute to positive "engagement" and "flow."

Perception Qualitative Codes

The codes for perception were concerned with players' aesthetic experiences of the game that were not emotions. For example, SLOW was a code that labeled when players perceived the game to be slower in feel because of the controls or aesthetics. Another case is EASY, which is when participants considered the game to be easier to play, due to the controls, aesthetic effects on the screen, or other aspects of the game. The players' perceptions helped clarify their PX and why they might feel the way do about certain features.

6.2 Overview of Study Sessions

The following sections present the summaries of the study session for each participant. The summaries were written originally as interim documents to assist in the data analysis process [92]. The summaries provided a method to reflect on potential themes and patterns that seemed to be emerging from the data.

For the most part, participants improved their gameplay as the rounds progressed, as can be seen in Figure 6.1 (word length and rarities across rounds). Most of the players played more words and scored higher in the later rounds compared to the first two to three rounds. This was not universally the case, however, with Josh and Donna being the main counterexamples to this. Zoey's play also remained relatively stagnant throughout the entire session, though she improved in the last two rounds.

The chart comparing participants across versions in Figure 6.2 showed that there were no clear patterns that emerged. This was not entirely surprising since the first chart showed a tendency for players to improve their gameplay as they played more rounds. The order of the versions for each round was randomized, so the effect of playing more rounds seemed to overshadow any noticeable effects from the differing versions. Despite this, there were significant qualitative differences in how the participants played and reacted to each of the frictional versions.

6.3 Session 1: Claudia

The first participant, who will be referred to by the pseudonym Claudia, identified herself as a woman and indicated on the questionnaire that she played video games "less than once a month." She also answered on the questionnaire that she played word/puzzle games similar to *Brave New Word* "a few times per month." She was

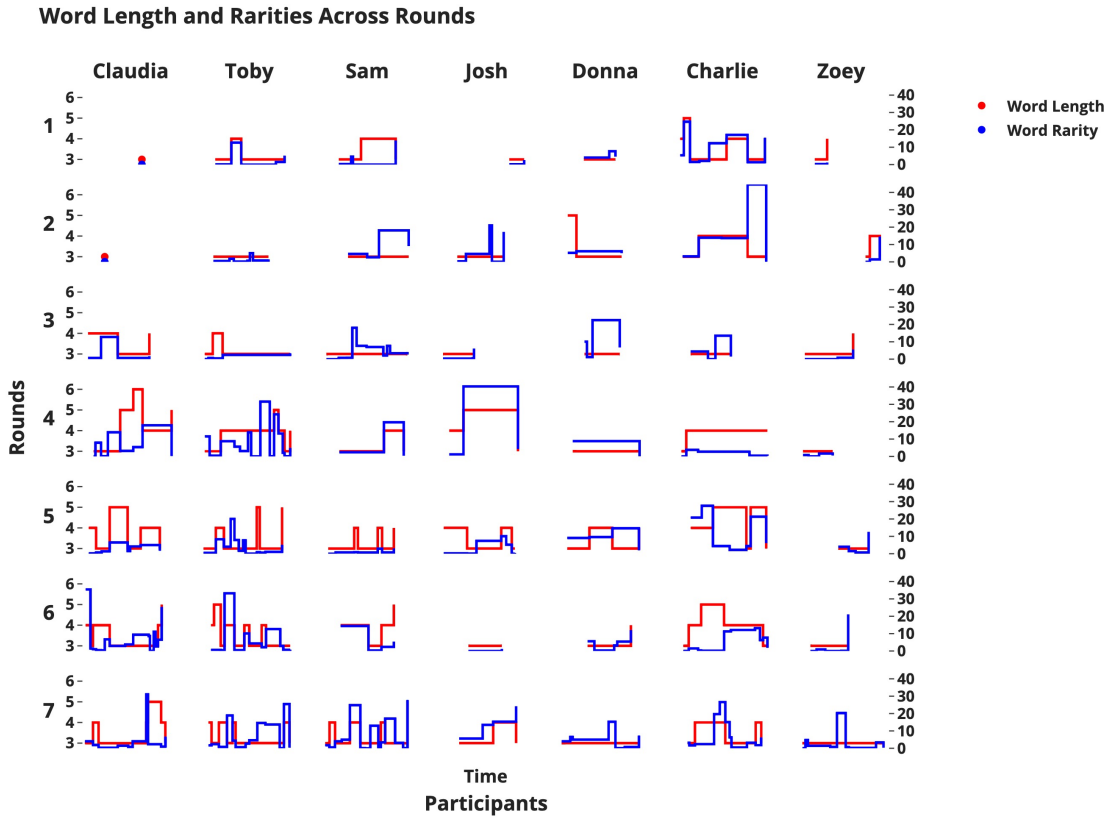


Figure 6.1: A tiled facet plot showing the word lengths and rarities played in each round, chronologically.

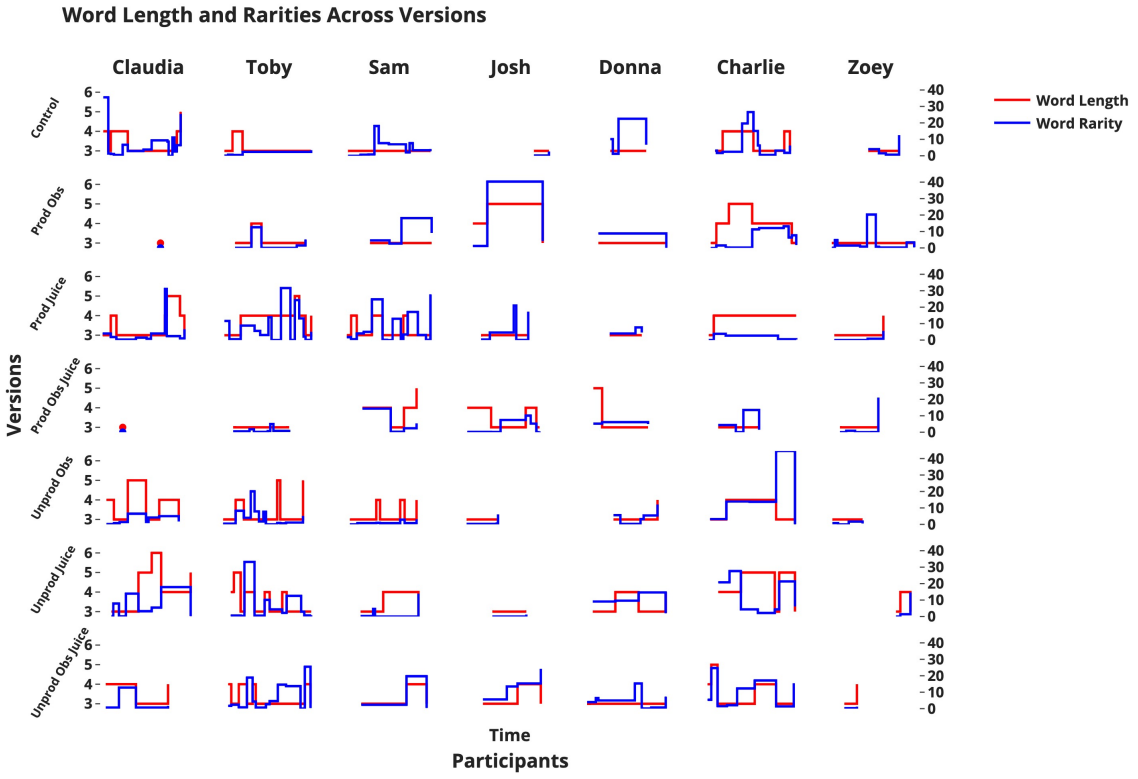


Figure 6.2: A tiled facet plot showing the word lengths and rarities played in each round by version.

familiar with games like *Scrabble* and *Boggle* and said that she used to play *Scrabble* with her mom frequently when she was younger. When asked what her goals were when she played games similar to *Brave New Word*, Claudia responded by choosing the following options: maximize my point total, relax, waste time and/or distract myself from other activities, mess around, and gather achievements. Interestingly, the questionnaire results indicated that Claudia did not consider mobile word/puzzle games to be video games, which is a common perspective due to misogynistic gatekeeping in video game culture. Though she may not have considered herself a frequent gamer, her responses revealed that she played games—especially mobile word games—regularly. Claudia was a native English speaker.

The first round that Claudia played was productive obstruction. In this first round, she seemed overwhelmed and was only able to submit one word, NEW for 6 points. Claudia was almost able to submit a second word, SEE, but time ran out before she could play it. Notably, she implied that searching for longer words was part of her strategy when she found the word NEW: "And so, okay, there's NEW. That's not a super big word." This showed that even though the game never explicitly stated that longer words could be a goal, she made that part of her objective.

The second round played was productive obstruction and juice. In this round, Claudia was still only able to submit one word for 6 points (the word TOO). This could partially be due to the fact that she was still familiarizing herself with the game, and it could also be a byproduct of having to think-aloud and answer interview questions while playing. A highlight from this round was that Claudia commented that the button added extra time to play which was not her preference. She would have preferred to "just highlight it" and have it "disappear automatically."

The third round played was unproductive obstruction and juice. This time, Claudia was able to play 4 words before the round ended. She submitted three 4-letter

words and one 3-letter word for a total of 46 points. The rarest word that she played was SEAM with a rarity of 12.8. A reminder that rarity was expressed as a percentile, thus a rarity of 12.8 meant that 12.8% of all words in the corpus were as common or more common than the word SEAM. Her average rarity was 3.975. Claudia most likely became more comfortable with the game by this point, because she had more success in this round than in the previous two rounds despite saying that the unproductive juice effects were distracting. She also stated that the distractions made her feel more anxious while playing the game. It should be noted that she mentioned that the anxiety in part also stemmed from the fact that each round was timed.

The fourth round played was unproductive obstruction. In this round, Claudia played 8 words and scored 116 points with an average rarity of 6.1. Her longest word was CARROT with 6 letters, and her rarest word was KALE at 17.9. She clearly played her best round so far in this fourth round. When prompted to compare this version with the previous round, she mentioned that the game was "calmer" and less distracting. Notably, she said that this version did not make her feel as "rushed" as in the version with unproductive juice. According to her, this was due to the lack of distractions relative to the juicier version.

The fifth round played was unproductive juice. Claudia also played 8 words in this round, but scored 71 points with an average rarity of 2.65. Her longest word, CRUST, was also her rarest with 5 letters and a rarity of 6.5. Interestingly, despite the return of the distractions that contributed to her anxiety in round 3, she preferred the unproductive juice version over the unproductive obstruction. She said that because it was faster to select letters and submit words, she liked the unproductive juice version "way more." She recognized the distracting effects and the anxiety from those effects in this version once more, but still seemed to prefer this over the obstructions. Ironically, she quantifiably played slightly worse in this

version, but it was difficult to compare any of these rounds because of the interference from the think-aloud protocol. She played worse in this unproductive juice version by playing shorter words and less rare words overall. There were several potential reasons why she might have played worse, (1) the controls did not afford pausing to consider options while selecting a word, thus leading to shorter words overall; (2) the controls made it more difficult to select longer words; (3) because she played shorter words, it is more likely that she played less rare words. In this particular case, it appeared that there was a higher number of actions where she deselected letters (to modify her word selection) and pausing while selecting a word in the unproductive obstruction version. This suggested that the controls in unproductive juice were not conducive to her being able to create longer words or reflect on her options while she played.

The sixth round played was the control version. Claudia played 16 words in this round for a total of 240 points with an average rarity of 7.84. Her longest word was ROOKS with 5 letters, and the rarest word was LADE with a rarity of 35.5. When asked to compare this version to the others she had played up to this point, she declared this one her favorite. When asked why, she mentioned the things that she had pointed out before. The lack of distracting features (especially the screen shaking) and the ability to select and submit words quickly without a button were at the top of her list of reasons. She also mentioned that she felt less "overwhelmed" and more comfortable with the game. This comfort came partly from having played the game multiple rounds, but it also had a lot to do with the lack of extra juicy features and obstructions.

The final round played was productive juice. Claudia played 10 words for a total score of 125 with an average rarity of 5.37. Her longest word was DRAFT at 5 letters, and the rarest word HOAR at 31.1. She confirmed that the control version

was still her favorite version to play. She stated that there were more distractions than the control version, but that it was "not the most distract[ing] thing." When asked about the distractions, she made a contradictory statement about her strategy. At first, she claimed that she only clicked on letters when she knew what word she was going to attempt. However, when asked to expand on her strategy, she discussed how there were no negative effects for trying incorrect words and that she sometimes discovered words on accident or by randomly exploring. That was how she explained her discovery and submission of the word HOAR.

6.4 Session 2: Toby

Participant 2, who will be referred to as Toby, self-identified as a man and indicated on the questionnaire that he played video games "almost always daily." He did not indicate that he played many casual or mobile games. Instead, he said that he mostly played console or PC games. He also answered that he does not usually play puzzle or word games like *Scrabble*, *Boggle*, etc. He stated that he was not good at word games. His response to what his goals are when playing games similar to *Brave New Word* were to maximize my point total, relax, waste time and/or distract myself from other activities, be better than other players, and gather achievements. Toby was also a native English speaker.

The first version that Toby played was productive obstruction. He played 5 words for 45 points with an average rarity of 3.88. His longest word, WART, was also his rarest at 12.6. After he played the first word, TOO, he commented that his goal was to find longer words than TOO. Despite his initial objective, he said halfway through the round that he was struggling to find words. This was a fairly common issue for participants in the first couple of rounds while getting familiar with the game. At

the end of the round, Toby stated that he found the background music calming, even though the timer aspect of the game generated anxious feelings.

The second round was productive juice and obstruction. Toby played 7 words worth 43 points with an average rarity of 1.5. He did not play any words longer than 3 letters, and WAX was his rarest word with a rarity of 5. He noticed the inclusion of the extra juice immediately and said that it made the game feel more "actiony." Though he did not say that he preferred this experience over the version without productive juice, he did say that the "explosion" sound whenever a word was submitted was "kind of satisfying." He also observed that the game felt faster. He attributed this feeling mostly to the change in the soundtrack (the music for the juicy versions were more energetic, synth-heavy tracks compared to the quieter piano track for the non-juicy versions).

Toby's third round was the control version. In this round, he played 5 words worth 34 points with an average rarity of 1.42. His longest word was STEP, and MUD was the rarest with a rarity of 3.4. When asked about the controls of the game, he said that he preferred this control version. He liked that he did not have to press a button and go through an additional prompt window which he said "disrupted the flow." He also preferred the "calmer" music (in comparison to the juicy version's music) saying that it felt "better for this sort of game." I interpreted that to mean that he associated the gameplay with a more relaxing atmosphere as opposed to an energetic, adrenaline-pumping type of game.

It should be noted that up to this point, Toby only played words in a linear fashion where the letters lined up straight horizontally, vertically, or diagonally. It was not until the fourth round that he discovered he could play words in such a way that the letters were adjacent to each other in any direction without being completely linear.

The fourth round played by Toby was productive juice. Due to his breakthrough in realizing how he could form words, he was able to play significantly more words. He played 14 words for a total of 249 points with an average rarity of 8.56. His longest word was MOTES with 5 letters, and the rarest word was SETT at 31.6 rarity. Interestingly, he commented that selecting letters diagonally was challenging on the trackpad. This corroborates one of the findings from Claudia's case where non-obstructed controls potentially made it more difficult to select longer words, at least in the case of using a trackpad. It is not a stretch to imagine that there might be a similar effect even when using a different controller such as a mouse or a touchscreen. Another interesting aspect was that he found the more energetic music in the juicy version more fitting when he was successfully playing more words. When he was playing multiple words in rapid succession, he commented that he found the soundtrack more appealing. It is an interesting idea that music might not necessarily elicit the desired affective response in the player but achieves maximum effect when the player's behaviors or outcomes match the mood of the music. Instead of music being a unidirectional effect, player behaviors and performance levels could influence what types of music are appealing. In the previous round, he said that he preferred the calmer music while he was still playing relatively few words, but his preferences changed as soon as he was able to play many words quickly.

The fifth version played was unproductive juice. Continuing his strong performance, Toby played 12 words for a total of 121 points with an average rarity of 4.83. His longest word was once again 5 letters long, and the rarest word was RIN with a rarity of 20.1. After learning that he could select letters in any direction in the fourth round, he began to find words more quickly and more consistently. Toby did not enjoy the unproductive juicy version of the game as much as the other settings he had played so far. He called the screen shaking when selecting tiles "obnoxious"

and "disruptive" and found other features, such as the laser sound effects, distracting. Another unproductive juicy effect was that the letters would "flash" briefly and blink in and out while the screen shook. An unforeseen consequence of this was that Toby had to plan out his actions more intentionally when selecting words. He explained it as having to "visualize the entire word" because of the flashing effect. It is uncertain if the unproductive juicy effects directly led to the lower word count and average rarity or if other factors were involved. Regardless, Toby found this version to be obnoxious, less enjoyable, and more distracting than the other versions. These negative player experience outcomes along with the lack of measurable positive outcomes supported the view that this version was unproductive.

The penultimate round played was unproductive obstruction. In this round, he also played 12 words. He scored slightly higher than the previous round with a score of 158 and an average rarity of 5.75. His longest word was once again 5 letters long, and his rarest word had a rarity of 33.2. This was Toby's first interaction with the unproductive obstruction features. He had immediate strong negative reactions to the obstructed selection process. As he commented on how slow the controls were, his voice conveyed dejection and strong emotions in a way that he had not up to this point. For the most part his voice had remained calm and emotionless until this round. He found the controls "frustrating" and "super inconvenient." Interestingly, he stated that he was "more inclined to choose smaller words just to do less work." Despite enjoying the game much less and even stating that he would have quit playing the game if it were not a study, he performed similarly to the previous two versions.

The last round played by Toby was unproductive juice and obstruction. He continued his improved gameplay by playing 14 words with a total score of 167 and an average rarity of 6.47. His longest word was 4 letters long this time, and his rarest word was GOOS with a rarity of 25.5. When starting this round, he noted that this

version was a combination of "the two bad settings" (i.e. unproductive juice and obstruction). To him, the gameplay and the sounds did not match in this version. This was an additional observation on top of all of the other features he commented on in the previous two rounds. A new insight that Toby gained during this round was that sometimes shorter words can have higher points due to their rarity. He was confused at first and thought the rules did not make much sense. When asked if this revelation would change his strategy, he claimed that it would not because his strategy was just to play as many words as possible to gain points. His reasoning was that he would not be able to look at a word and determine its rarity, so playing as many words as possible was still the best strategy to him. He apparently did not consider the prompt window in the productive obstruction versions of the game which presented the user with the rarity of each word.

6.5 Session 3: Sam

Sam was the third participant in the study. Sam identified as a man. He indicated on the questionnaire that he played video games a few times per week, but played word or puzzle games less than once a month. Although he was familiar with games like *Scrabble*, he did not play them regularly. Maximize my point total, be better than other players, and solve complex challenges were the options he chose when asked about his goals while playing games similar to *Brave New Word*. Sam was also a native English speaker.

The first version that Sam played was unproductive obstruction. He played 5 words for 43 points with an average rarity of 3.76. His longest word was 4 letters long (THEM), and his rarest word was LOO at 13.9. He did not enjoy the obstructive selection controls. He wanted to be able to drag to select words instead of having to

click each letter. Although he stated that he "kind of liked" the button because it gave feedback on whether or not the words were valid, he also stated that the game would go faster if he could submit words without the button. At first, he stated that his goal was to find familiar and bigger words. However, he found that it was much easier to find shorter ones, and so began to settle for shorter words.

The second round Sam played was productive obstruction. He played 4 words for 41 total points with an average rarity of 8.6. His longest word was only 3 letters long, and his rarest word was ANI with a rarity of 18.1. In this version he enjoyed that there was a drag-to-select mechanism. However, he was annoyed by the submit prompt window. He felt that the window was "repetitive" and unnecessary. He also thought that the extra window and button slowed down his gameplay. Overall, he spent slightly more time in this round experimenting with the controls than the first round.

The third round was the control version. In this round, Sam was able to play significantly more words than in the first two rounds. He played 10 words for a total of 78 points and an average rarity of 5.17. His longest word was 3 letters long and his rarest word was once again ANI at 18.1. Sam instantly expressed his preference for this version over the previous versions. He liked that he could let go to automatically submit the word instead of having to press a button. He also noted that the controls felt "very clean cut." He also said that "this version seems to have everything just the way I want it." When asked about his strategy, he said that he had shifted to finding shorter words because it was easier. This seemed to have paid off for him in this round because he was able to play significantly more words than the previous two rounds.

The fourth version Sam played was unproductive obstruction and juice. He was only able to play 3 words for a total of 39 points with an average rarity of 7.5. His

longest word was LEER with four letters, and it was also the rarest word at 19.7. He instantly reacted negatively towards the juicy feedback. He found that the excessive screen shaking made him feel like he "did something wrong." He was unsure how to interpret the screen shaking initially and that slowed him down. He also found a lot of the visual feedback, such as the background sparkle animations, distracting. He once again commented that the unproductive obstruction selection controls made the game feel slower. He was only able to play a few words in this round, which corroborated the statements he made about the slowness of the game.

The fifth version was unproductive juice. In this round, Sam played 8 words with a total score of 62 and an average rarity of 1.25. His longest words were 4 letters long and the rarest word he played was LEAN at 3.1. He continued to dislike the excessive screen shaking in this version. Specifically, he was annoyed that the shaking could disrupt his selection of words when trying to drag diagonally. This slowed down his playing. He thought that this feature made the game "extra challenging."

Sam played the productive obstruction and juice version next. He was only able to play 4 words for a total score of 54 and an average rarity of 5.63. His longest word was HEELS with 5 letters, and his rarest word was WILT with a rarity of 14.4. In this version, Sam discovered that he could drag-to-select and let go and modify his selection afterwards. This opened up the door for him to explore and compare words before submitting. He only played 4 words in this round mostly because he spent much of the round talking and reflecting on how the scoring system for the game worked. For example, he used the submit prompt window to compare the rarities of different words before submitting them and used those rarities to compare the scores of the words. He also realized in this round that the highlighted color of the letters reflected the rarity of the selected word. He unfortunately came to the wrong conclusion that rarer words do not give extra points because he compared two words

that had different rarities but were worth the same number of points. The two words were close enough in rarity that it did not cross the threshold to change the score of the words. He did demonstrate that productive obstruction might promote more reflection when choosing words. He stated that he could "scan around [the word] before submitting it," which he did multiple times in this round. He also came to prefer this selection method (being able to drag, let go, and still modify the word) over the automatic submission method (where letting go of the word submitted the word).

The final round Sam played was productive juice. In this version he was able to play 14 words for a total score of 185 (a personal best) with an average rarity of 7.21. His longest words were 4 letters long, and his rarest word was MOG with a rarity of 27.9. Even though he had just come to the conclusion that he preferred the obstructive selection method over the non-obstructive controls, he found this version to be "more fast-paced." He thought that he might have preferred this version if he was "better at the game." He was able to play significantly more words in this round compared to his previous attempts. His play was similar to the control version where he was able to rapidly find and submit short words.

6.6 Session 4: Josh

The fourth participant, Josh, self-identified as a man. He also indicated on the questionnaire that he did not usually play video games, though he made the caveat that this was mainly because he was busy with schoolwork. He also responded that he did not usually play puzzle/word games. When asked what his goals were when playing games like Brave New Word, he chose: mess around, find the best strategy, gather achievements, and solve complex challenges. Josh was a non-native English

speaker, and he indicated multiple times during gameplay that the language barrier made it more difficult to play the game.

The first round Josh played was the control version. He was only able to play 2 words for a total of 12 points with an average rarity of 1.4. Both words were 3 letters long, and the rarest word was RAT with a rarity of 2.6. His immediate reaction to the game was confusion, mainly due to the language barrier. He found the all uppercase letters difficult to understand. After struggling through reading the letters, he did start experimenting with selection and eventually found the two words.

The second round was the productive juice version. Josh was able to improve his play and submitted 5 words for a total of 65 points. The average rarity of the words was 8.66. All of the words were 3 letters long, and the rarest word was TAM with a rarity of 17.29. When asked about the more energetic music in the productive juice version, Josh commented that the music was distracting and made it harder to focus compared to the control version. When asked about the animations and visual juice in this version, he responded that the explosion effects made successful submissions of words feel more satisfying. He seemed to have settled into the gameplay more in this round.

The third version Josh played was unproductive juice. He played 3 words for a total score of 19 points. The average rarity was 2.3, and the longest word was only 3 letters long. The rarest word was TOW with a rarity of 6.1. When starting the round, he immediately commented that the music was "so powerful" and even "more distracting" than the productive juice version. He also said that the game made him feel "not so good." He found things like the extra screen shaking distracting, but his main complaint was with the music. He called it "terrible" and found it difficult to focus his attention while playing. The unproductive juice elements made him

"nervous."

Productive obstruction was the version of the fourth round. Josh again was only able to play 3 words. The total score was 93 points, and the average rarity was 15.23. The longest and rarest word was DISHY with a length of 5 letters and a rarity of 40.3. He correctly noted that more operations are needed to submit words in this version. Although it slowed down his gameplay, he also recognized that it provided "guidance for help" about each word to submit. He thought the controls were "much more friendly this time" while also acknowledging that they took more time. He also thought that this version was more informative because it showed the rarity of each word in the prompt window. It was notable that he played his longest and rarest word in this round.

Next, Josh played the productive obstruction and juice version. In this round he played 7 words for a total score of 66 points and an average rarity of 3.36. His longest word was 4 letters long, and the rarest word was KANE with a rarity of 10.1. Having played both the productive obstruction and productive juice versions previously, he understandably felt more comfortable and familiar with the game. He was able to play the most words in this round, perhaps because he was more comfortable. He noted in this round once more that he struggled with the language barrier.

The sixth round was unproductive obstruction. Josh played only 2 words for a total score of 12. The average rarity was 0.6. The longest word was 3 letters long, and the rarest word was HAT with a rarity of 1.2. His immediate response was that it was "slower to select words." He also noted that the background music and sound effects did not seem to match the overall mood of the game. His overall feeling was that this version was much more difficult and confusing to play.

The last round was unproductive obstruction and juice. Josh played 4 words for a total score of 83 points and an average rarity of 14.65. The longest word was 4

letters long, and his rarest word was REE with a rarity of 24.2. His observations from the previous unproductive obstruction and unproductive juice versions remained unchanged. He thought the animations and extra screen shaking was distracting and the controls harder to play. He seemed to struggle on focusing on playing the game while talking because of all the distractions in this version.

6.7 Session 5: Donna

The fifth participant, Donna, self-identified as a woman. She indicated that she played video games "a few times per month" and word games "a few times per month." When asked what her goals were when playing games like Brave New Word, she chose the following: relax, waste time and/or distract myself from other activities, and mess around. Donna was a native English speaker.

The first version she played was productive juice. She played a total of 3 words for a total score of 20 points and an average rarity of 5.43. The longest word was 3 letters long, and the rarest word was TAT with a rarity of 7.7. She spent much of this round getting acquainted with the game, like many of the other participants. When asked how the controls felt, she said that she liked the natural feel of dragging to select letters and submitting words by letting go. She also enjoyed the juicy effect of the screen shaking when submitting a word.

The second round was productive obstruction and juice. Donna again played a total of 3 words. The total score was 20 points and the average rarity was 5.43. The longest word was 5 letters long, MOTTO, and the rarest word was COD with a rarity of 6.1. Her response to the added button and prompt was mostly ambivalent. She thought that it could be useful if the game was not timed, but since the game mode was timed she thought it was a useless feature. She also noticed certain audio

feedback in this round that she either did not notice or did not comment on in the first round. She enjoyed the background music and thought the audio feedback made it feel "like you did something good."

The control version was the third version Donna played. In this round, she played 4 words for a total of 57 points and an average rarity of 10.13. Her longest word was 3 letters long, and the rarest word was HAP at 22.5. When asked about her strategy, she said that she was looking for longer words to play. Unfortunately, that strategy must not have worked for her since all of the words she played in this round were only 3 letters long. She commented that she preferred the faster, upbeat music in the previous versions compared to the calmer piano music in the control. She thought that it was easier for her to focus with the more energetic music.

The next version Donna played was productive obstruction. She only played 2 words for a total of 14 points and an average rarity of 4.45. Her longest word was 3 letters long, and the rarest word was TAO at 8.8. When asked to compare this version to the previous versions, she said that she preferred the controls when letting go automatically submitted the words. She found that adding the Play Word button and prompt slowed down the game and made it more difficult to get more points.

Unproductive obstruction was the fifth version played by Donna. She played 4 words for a total of 45 points and an average rarity of 8.8. Her longest word was 4 letters long, and her rarest word had a rarity of 14.6. Her immediate reaction to the new controls for selecting letters was that she "definitely like[d] the dragging more." She explained that she could play words more quickly and "glide" instead of having to click on each letter. She also preferred being able to just click on the Play Word button in this version rather than having to also click through a submit prompt window like in the productive obstruction versions.

Unproductive juice was the next version Donna played. She played 5 words for

a total of 45 points and an average rarity of 5.32. Her longest word was 4 letters long and the rarest word had a rarity of 12.1. She enjoyed the even more energetic music in this version more than the previous versions. She said that she felt like she could "get into...the game mode" because of the music. She also thought that some of the juicy elements made the game feel more "alive" and reactive to her playing. For example, she thought it was fun to see the letters moving and shaking when she selected them. She was the only participant who found the unproductive juicy elements to not be a negative distraction. She thought they were good distractions and made the game more "lively," "energetic," and "engaging." She preferred this version over the previous versions she had played.

The last version Donna played was unproductive obstruction and juice. She played 7 words for a total of 58 points and an average rarity of 5.57. Her longest word was 3 letters long, and her rarest word had a rarity of 15.29. A notable comment during this round was that she thought it was more difficult to "guess" and "find random words" with the selection system in unproductive obstruction. She theorized that it was harder to explore and try new words because of this hindrance. Because of this, she found it a little more frustrating to play the game. In spite of this, she was able to play the most words in this round, perhaps because she had more experience playing the game.

6.8 Session 6: Charlie

The next participant, Charlie, self-identified as a man. He indicated on the questionnaire that he played video games almost daily but that he did not usually play puzzle/word games. When asked what his goals were when playing games like Brave New Word, he chose maximize my point total, relax, and waste time and/or distract

myself from other activities. Charlie was also a native English speaker.

The first round Charlie played was unproductive obstruction and juice. He played a total of 8 words for 142 points with an average rarity of 10.0. His longest word was 5 letters long, and the rarest word had a rarity of 24.7. Interestingly, when asked about the controls of the game, he said they felt "pretty normal." He also said he liked the click-to-select controls. This contrast with the other participants who generally disliked this version can probably be attributed to the fact that this was Charlie's first round. He had not played any other versions to compare. He did compare it to a hypothetical drag-and-let-go version, saying that it would be potentially "annoying" to have to go back and redo actions if he had accidentally let go. However, he found the juicy effects distracting, and he thought the energetic music did not match the style of game.

Unproductive juice was the second version Charlie played. He played 5 words for a total of 105 points with an average rarity of 15.08. His longest word was 4 letters long, and the rarest word had a rarity of 44.6. Now that he was playing a drag-to-select version, he had a chance to actually compare the two controls. He continued in saying that he preferred the obstructed version and that he was not "as big of a fan of this dragging version." He also continued to find the juicy features like the visual sparkles and extra effects distracting.

The third round he played was productive obstruction and juice. Charlie played 4 words for a total of 33 points with an average rarity of 4.85. His longest word was 3 letters long, and the rarest word had a rarity of 13.5. When asked about the differences in this version, he felt that the submit prompt window was an "unnecessary step" because a player has already dragged to select and hit the Play Word button. Like Donna, he pointed out that if the game was turn-based as opposed to timed, the prompt window might make sense. However, he thought it was "not very helpful"

because it was timed.

Productive juice was the version of the next round Charlie played. He played 5 words for a total of 48 points with an average rarity of 1.46. The longest word was 4 letters long, and his rarest word was POUR with a rarity of 3.8. He commented during this round that his goal was to find longer words that were at least 4 or 5 letters long, but he admitted that it was difficult to find longer words. He did not make any new observations about the productive juice features.

The fifth version Charlie played was unproductive obstruction. He played a total of 8 words for a score of 184 with an average rarity of 11.12. He played 4 words that were 5 letters long, and the rarest word had a rarity of 27.6. An intriguing strategy that he employed in this round that no other participant did was to pause and analyze the board before starting the game. Part of the board is visible to the player before they click to start the game. This way he could start the round by immediately playing a word. In fact, he found a decently rare word to start the game, BUOY, with a rarity of 20.8. He found this word by analyzing his options before starting the game. When he realized that rarity had a significant impact on the scores of the words, he changed his strategy to look not only for longer words but also for more uncommon ones. He felt more comfortable in this round and also attributed his relative success in this round to the randomness of the letter distribution. He stated that he still preferred this version including the selection controls and the Play Word button.

Productive obstruction was the sixth round Charlie played. In this version he played a total of 10 words for a score of 124 with an average rarity of 5.44. His longest word was 5 letters long, and the rarest word had a rarity of 13.2. He once again employed the strategy he discovered in the previous round. He searched the board for a good word to play before starting the round. When faced with the Play

Word button and submit prompt window, he once again stated that he was "not a fan" of having to press multiple buttons to play a word. When asked about the lack of juicy effects, he said that it was less distracting to not have the extra juicy visual effects. However, he also stated that he had come to "tune out" the effects for the most part, so he "barely noticed that they weren't in." He ignored the juice, so was it having an effect one way or the other? When asked if he preferred to have the juice in the game or not, he said that it would be "better to not have them in at all." He justified this by saying that he thought it made the game more distracting and that this type of game was not one where he was "trying to be on the edge of [his] seat."

The final version Charlie played was the control version. In this round he played a total of 11 words for a score of 149 with an average rarity of 7.89. The longest word he played was 4 letters long, and the rarest word had a rarity of 26.6. His immediate reaction to this version was that it was much more "peaceful." He actually found the letter selection controls in this version more challenging. He found that it was more difficult to select diagonal letters in this drag-and-let-go method. He found the versions with the button much easier to play because he could modify his selection after he let go.

6.9 Session 7: Zoey

The last participant, Zoey, self-identified as a woman. She responded on the questionnaire that she played video games a few times per week. She mentioned that she liked to play multiplayer online battle arena (MOBA) games like League of Legends and Dota 2. She also responded that she played puzzle/word games a few times per month. When asked what her goals were when playing games like Brave New Word,

she chose the options: maximize my point total and solve complex challenges. Zoey was a non-native English speaker.

The first round Zoey played was unproductive obstruction and juice. She only played 2 words for a total of 16 points with an average rarity of 0.75. Her longest word was 4 letters long, and the rarest word had a rarity of 1.2. Just like many of the other participants, Zoey had to get acquainted with the game at first. For instance, she had to select a few words before realizing that she had to press the Play Word button in order to submit them. When asked about the controls, she said that they felt fluent, which I interpreted to mean something along the lines of fluid or smooth. I thought this was interesting as most of the other participants did not like the obstructed controls in this version. Potentially, she felt this way because this was the first version she played, just like Charlie.

The second version Zoey played was unproductive obstruction. She played a total of 3 words for a score of 30 with an average rarity of 5.33. Her longest word was 4 letters long, and the rarest word had a rarity of 14.6. She was still struggling with the controls in this round. Once she realized that she could select letters in all directions including diagonals, she was able to more confidently search for words. It is also important to note the language barrier may have made it more difficult for her to play the game. After finishing this round, she stated that she thought the game could be improved if there was no button. Although she originally stated that she liked the controls, she changed her mind after playing another round with unproductive obstruction. Becoming more familiar with the game probably changed her opinion. She thought that the button made it slower to play. However, she also observed that there was a potential benefit to having the obstructed controls. She thought that the obstructed controls could help players form longer words because they could pause before submitting shorter words and potentially find larger words

and modify their selection.

The third round she played was productive juice. Zoey played 3 words for a total of 24 points with an average rarity of 2.13. Her longest word was 4 letters long, and the rarest word had a rarity of 5.6. Just like Charlie, she also found the drag and slide controls to be more challenging than the obstruction controls. She said it was harder to keep moving in a straight line to select the correct letters. However, she found the controls for submitting words to be faster, so in the end she preferred this version to the previous versions. Even though at first she found the controls to be harder, by the end of the round she said that it seemed to be easier. There was a quick acclimation to the new controls, and it seemed to be her preferred style because it was faster to submit words. She admitted that it would be more difficult to select bigger words with this version, but made the observation that most of the words she played were short anyway and that long words tended to be rare in this game.

The next version Zoey played was unproductive juice. She played 4 words for a total of 24 points with an average rarity of 0.75. Her longest word was 3 letters long, and the rarest word she played had a rarity of 1.7. Although she recognized the extra juicy effects in this version, she said that it did not "influence [her] a lot" because she just focuses on playing the game for most games. I interpreted this as her also tending to ignore juice like Charlie did.

The fifth round Zoey played was the control version. She played 4 words for a total of 32 points with an average rarity of 4.7. Her longest word was 3 letters long, and the rarest word she played had a rarity of 12.6. She felt that this version was the most comfortable out of all the versions she had played so far. She preferred the control version because of the lack of distractions and the calmer audio.

Productive obstruction and juice was the sixth version Zoey played. She played 5

words for a score of 50 with an average rarity of 4.42. Her longest word was 3 letters long, and the rarest word she played had a rarity of 21.1. This was the first version she played that had the submit prompt window. At first it caused confusion, and she thought that submitting one word would end the game. However, after figuring out how it actually worked, she thought that the display of the rarity of a word was a helpful feature. When asked how she felt about the prompt window, she said that it made the game harder to play. She thought it added too many steps and made the gameplay slower. She also thought that some of the sound effects like the explosion sound when submitting a word created some anxious feelings while she was playing.

The last version Zoey played was productive obstruction. She played 9 words for a total score of 74 with an average rarity of 3.57. Her longest word was 3 letters long, and the rarest word she played had a rarity of 20.3. She played the most words in this round most likely due to her growing familiarity with the game. She once again commented that she preferred the more "comfortable" sound effects in this non-juicy version. She found this game to be more relaxing than other games she liked to play like competitive MOBA games because she did not have to "think a lot."

Chapter 7: Discussion

7.1 Introduction

This dissertation explored the effects of productive and unproductive frictions in a casual word game, *Brave New Word*, designed specifically for this work. This study was informed by the findings on the potential positive impacts of intentional design frictions in applications—games or otherwise [30, 91]. In addition, this work drew from studies exploring the benefits and disadvantages of juicy game design [56, 67]. Based on such studies, this dissertation asked how intentional design frictions in the form of obstructions and juice affect players, and whether those frictions can be designed in such a way to make them productive in some sense.

Researchers have been working to apply knowledge from the fields of human-computer interaction, usability, and user experience to the area of video games for nearly 2 decades [10, 68, 105]—if not longer. In UX circles, the concept of benevolent and intentional design frictions is still relatively unexplored. In the realm of PX, although concepts like "challenge" and "ease-of-use" are widely accepted to be important to game experience, there is a lack of scholarly work investigating the effects and mechanisms of design frictions in video games. Thus, this work contributes to the growing literature that is attempting to understand how intentional design frictions can improve user experiences [30, 91]. Specifically, it attempts to address the gap in understanding how these frictions work in the context of video games.

To explore productive frictions, two specific game design concepts were chosen: obstructions and juice. Obstructions on game actions could be used as microboundaries to support "mindful" interaction [30]. Juicy design, on the other hand, intro-

duce friction in the form of potential distractions but are theorized to improve overall player experience and (perceived) player competence levels. These two examples fit the mold of productive friction as defined in this work, so they provide a good opportunity to test the potential benefits of productive friction in this specific game context. Researchers have applied the approach of microboundaries on products such as smartwatches [24] and digital self-control apps [80], but there is a lack of literature on how they apply to video games in particular. Juicy game design is a popular topic amongst game developers and designers in mainstream media [43, 64, 120], but structured, empirical evidence for the benefits of juice is still severely insufficient considering the mainstream appeal of the concept [56, 67, 72].

A digital word game was designed for mobile devices and the web browser to investigate the impact of productive friction in games. Multiple versions of the game were created for the study, each with different obstructive or juicy design elements. It was conjectured that testing these different designs of the game could help better understand how specific features can influence players. Chapter 3 described each feature in detail and summarized how the game was intended to explore the effects of the productive and unproductive variations on obstructions and juice. By testing and comparing the results from each of the versions, the goal was to discover the individual impacts of obstruction and juice on player experience and game outcomes.

The game, *Brave New Word*, was designed to be a simple word search game where players select adjacent letters on a board to form English words. The letters on the board were randomized and every time a word was found and submitted, those letters disappeared and new ones filled in their space. The base game without any variations in the controls or extra audio-visual effects was considered the *control* version. The *productive obstruction* version added a button that the player must press in order to submit their word. In addition, the player had to interact with a confirmation prompt

window before submitting a word. The *unproductive obstruction* version also included the button to submit a word, but instead of the confirmation prompt window, the user had to click on each letter individually to select. In contrast, the *control* version allowed players to select words by simply clicking and dragging over the letters. For the *productive juice* version, various particle effects, background animation, screen shaking, and more energetic audio were added to the game. The *unproductive juice* version included everything in the *productive juice* version with the addition of more extreme particle effects, excessive screen shaking, and more intense audio effects and background music. Lastly, one version combined the *productive obstruction* and *productive juice* versions, while another version combined the features in the *unproductive obstruction* and *unproductive juice* versions.

As described in Chapter 4, the work was comprised of two studies: a large scale analysis of telemetry data from people who downloaded *Brave New Word* from app stores and a qualitative analysis of seven participants who played through all versions of the game. Logistic regression was used in the former to analyze the impact of different variables on game outcomes, while verbal data from think-alouds and interviews were collected during the qualitative study sessions. A short post-task questionnaire was also conducted during the study sessions to collect demographic information about each participant. The think-aloud data and interview transcripts were coded for references to the frictional design patterns and how they affected the participants. The coding was then further analyzed for insights into the hypotheses on how juice and obstructions could impact players. By viewing all of this data along with the quantitative analysis from the first study, a more holistic understanding of the frictional designs emerged.

Chapters 5 and 6 presented the results of the quantitative and qualitative studies, respectively. This chapter will highlight the key findings from those results. The first

key finding is the role of obstructions on player behavior in the specific context of this game, including unintended side effects of certain design decisions. The next finding is on the impact of juice on player experience, and its apparent lack of impact on game performance metrics. Another finding includes some advantages and disadvantages of the chosen methodology for this study. The hypotheses proposed in Chapter 1 will be revisited and appropriately updated. Finally, the chapter discusses the implications of these findings on potential future work related to these topics.

7.2 Key Findings

One of the primary aims of this study was to understand *how productive and unproductive design frictions could influence players' behaviors and experiences*. It will be beneficial to review the nature of the data collected to help with this task. As described in Chapter 5, five dependent variables were selected from the game performance metrics: *Avg_Rarity*, *High_Rarity*, *Avg_Word_Len*, *Max_Word_Len*, and *Wds_Per_Round*. Several logistic regression models were built with these variables and the versions of the game (e.g. productive obstruction, unproductive juice). For the seven qualitative study sessions discussed in Chapter 6, every participant played one round of each of the seven versions of the game. For each round, the players participated in a think-aloud and semi-structured interview that was then coded for reactions to the frictional designs in the game (e.g. obstructive button, juicy animations). Players' verbal data were coded for positive and negative reactions and explanations for why they reacted to the game in the way that they did.

This methodological approach showed that one of the general findings of this study is that *productive frictions in the form of obstructions have the potential to change player behavior by introducing microboundaries that encourage mindful inter-*

actions. Another general finding is that *productive frictions in the form of juice have the potential to improve player experiences by creating engaging feedback but do not seem to impact game performance.* Therefore, players can benefit from the inclusion of certain kinds of design friction in games. These findings are in line with the work in current UX and PX literature, which have shown that design frictions and juicy design can positively affect player behavior and experience [56, 91].

7.2.1 The Role of Obstructions on Player Behavior

The two design frictions presented here, obstructions and juice, serve different purposes in game design. Productive obstructions, defined as microboundaries meant to promote mindful interactions in gameplay, can help users play with more reflection. Reflective gameplay can lead to better objective gameplay outcomes in certain cases, such as longer words in the context of this study—though it can naturally also slow down the gameplay. It can also lead to more thoughtful gameplay and changes in strategies. Game designers should be careful when creating microboundaries and obstructions, however. If the designs are too disruptive, they can dampen the player experience and lead to a less enjoyable experience as shown in the findings related to unproductive obstruction.

The results from logistic regression (Table 5.8) showed that people playing both the productive and unproductive obstruction versions played longer but fewer words each round. The lack of difference between the productive and unproductive obstruction versions suggests that both versions contained design friction that affected player behavior in a similar way. The friction could have stemmed from the same source (such as microboundaries), or independent obstructive elements could have coincidentally led to similar outcomes. In this case, the former explanation seems

more likely. Both versions contained the Play Word button as an obstruction, and this design could have impacted players as a microboundary. This result was surprising because the Play Word button was not intentionally included in the final design to be a microboundary that promoted System 2 behavior by itself. The button was supposed to be a design that merely increased friction. In comparison to the study by Hedeen et al. [54], the button could be seen as a successful "passive" design friction relative to the prompt window. The button being a productive design friction would make sense because it could be considered a microboundary, but there are some questions as to how the other frictional elements in the unproductive obstruction design contributed to the outcomes.

Study 2 helps us understand the results in Table 5.8 and suggests some ways that productive and unproductive obstruction affected players. Intentional design friction as discussed by Mejtoft et al. [91] consists of three categories: microboundaries [30], slow design [127], and uncomfortable interactions [11]. The design friction of the button and prompt window could be described as either microboundaries or slow design. Productive obstruction introduced a microboundary in the form of a prompt window before the player could submit a word. As per Cox et al. [30], the purpose of this microboundary was to promote the user to switch from System 1 behavior to System 2. The momentary pause required could create a more mindful reaction by the player as they considered their options as they selected words.

The prompt window could also be described as slow design. According to Strauss & Fuad-Luke, one of the core principles of slow design is to "induce contemplation and 'reflective consumption'" [127]. The intent of including the prompt window was precisely to promote contemplation and reflection. Although there was only one participant who explicitly used the prompt window to compare the rarity values of words before choosing which one to submit, that moment exemplified the potential

for reflection. Even some of the participants who personally thought that the prompt window was an unnecessary obstacle commented that the prompt window could be useful in certain situations. Despite recognizing the potential usefulness, nearly all of the participants commented on how slow the game felt in the obstruction versions. PER:SLOW was one of the most common codes in the analysis of the obstruction versions.

In addition, three of the seven participants in study 2 found the obstruction versions to be harder to play. The slower controls and more difficult gameplay could explain why the obstruction versions led to fewer words played per round.

On the other hand, the friction in unproductive obstruction could be described as being just bad design because it added extraneous steps to take an action without intentionally invoking mindful System 2 behavior. In that case, why did both unproductive and productive obstruction result in longer words? This might have been the unintentional result of the friction from the Play Word button. Some participants noted that it was easier to make modifications to their selection in the obstruction versions because they could let go and still make changes before pressing the Play Word button. In contrast, letting go would automatically submit the word in the other versions. In fact, there were many missed opportunities during the think-aloud sessions where participants could have easily added an extra letter or two to the end of their word, such as an 'S' or 'ED.' Although unintentional, the microboundary of adding a button made it more affordant for players to search for more letters they could add to their selection before submitting it. The micropause from the button created an affordance for users to be more mindful in their selection of words. Although the unproductive obstruction version was meant to be a bad design without productive friction, the button may have actually been productive enough to result in System 2 behavior.

Although both productive and unproductive obstruction versions resulted in longer words compared to the control, the means for average word length and maximum word length were not the same. The mean *Avg_ Word_ Len* for productive obstruction was 3.826 and 3.688 for unproductive obstruction. The mean *Max_ Word_ Len* for productive obstruction was 6.267 and 5.763 for unproductive obstruction. The odds ratio for productive obstruction's *Avg_ Word_ Len* was 4.857 while the odds ratio for unproductive obstruction's *Avg_ Word_ Len* was 2.669. This difference could be due to the lower friction involved in selecting letters in the productive version compared to the unproductive version. The ability to swipe and modify the selection after letting go seemed to be the sweet spot for participants. Although unproductive obstruction also allowed for modification after selection, having to click on each letter individually hampered the flow for many participants. One participant, Toby, even said that he felt "inclined to choose smaller words just to do less work," which could explain why unproductive obstruction might lead to less longer words than productive obstruction despite both versions having longer words than the control version.

7.2.2 The Influence of Juice on Player Experience

Juiciness did not seem to affect gameplay performance metrics unless the juicy feedback was especially distracting or disruptive. In study 1, productive juice showed no significant differences in game performance measures from the control version. This aligns with what other initial empirical studies have found on the effects of juicy design on games [56, 67]. Hicks et al. [56] showed that visual embellishments (visual juicy design) had no measurable impact on game performance by comparing a juicy version of a game and a non-juicy version. Most participants in study 2 preferred the

productive juice version over other versions. Although this is not an explicit measure of their player experience, it does suggest that the overall experience of productive juice was relatively high for the majority of the participants. This is not surprising based on the existing user experience literature. Research suggests that visual embellishments can have a positive impact on player experience by improving the aesthetic appeal and perceived usability of a system [52, 53, 56, 61, 88]. Hicks et al. also demonstrated that it is possible in certain scenarios for juicy designs to increase players' perceived competence levels [56]. Additionally, Kao provided evidence for how juiciness in a game could lead to an increase in performance [72].

Two participants, Toby and Donna, enjoyed the uptempo music in productive juice more than the calmer music in the non-juicy versions. Toby stated that it felt "better" when his gameplay matched the uptempo energy of the music. Donna said that it helped her focus. It should be noted that many of the participants simply turned the audio volume down when playing the game, so any potential effects of audio and music may have been missed. That uptempo music would have a different effect than calmer music in a player's experience is unsurprising. There is abundant literature showing links between music and a person's affect in diverse contexts including video games [115, 117]. Cassidy & MacDonald showed that high-arousal music correlated with increased driving speed and errors made while playing a driving simulator [22, 23]. Rogers, Jörg, & Weber also found that the presence or absence of background music could impact the amount of risk taken by players [115]. In this case, it appeared that the presence of uptempo music influenced participants' behaviors and affective states. The impact seemed to be highly subjective, as music can be, which is in alignment with existing literature [25, 102]. Some participants found the background music in the juicy versions highly distracting while others like Donna found it easier to focus and enter a flow state.

Unproductive juice showed a lower average *High_ Rarity* compared to the control version. This result corroborates the initial findings by Juul & Begy [67] where players performed worse in the juicy version of their game, although their results were not statistically significant. According to Juul & Begy, the distractions from juice could "split the attention of a user, increasing cognitive load" [67]. Participants from study 2 certainly noted the high number of distractions in the unproductive juice version. One of the biggest complaints from participants was that the extra juicy elements were distracting and sometimes even caused anxious feelings. These distractions and anxiety-inducing features could have led to worse performance in the game through the increase in cognitive load. This is in line with the split-attention effects found by Kalyuga, Chandler, & Sweller [71] and the findings of Kao's study on an action RPG [72]. As mentioned before, background music can have an impact on the affective psychology of players as well, though the effects can be highly mediated by the subjective perception of the music [25]. For example, the uptempo background music made some participants feel "rushed" as they played the game, which could have led to the anxious feelings. This is similar to the findings by Cassidy & MacDonald in their study on music and driving games [22, 23].

Another reason why unproductive juice could have negatively impacted gameplay performance is because of the features that slowed down or disrupted the player. For example, the screen shaking whenever they selected a letter created a natural barrier to quickly selecting words. This slowdown was unlike the microboundaries and slow design features of productive obstruction in that it did not prompt mindful interactions and reflection; instead, it usually led to frustration. This aligns with the findings of Borgo et al. [14] who demonstrated that visual embellishments led to slower visual search speed.

7.2.3 The Lack of Interaction Effects

In the versions where obstructions and juice were combined, there were no surprising interaction effects. That is, the logistic regression results for productive obstructions and juice were essentially a combination of the results for productive obstruction and productive juice, and likewise for unproductive obstruction and juice. This is overall unsurprising given that the frictional elements in obstruction were entirely separate from the design elements of juice. All of the juicy features were audio and visual in nature, while the obstruction features affected the controls and actions of the game. There was no reason to believe that combining the features would result in some sort of emergence, e.g. "the whole is greater than the sum of its parts" phenomenon.

7.2.4 Gender in Context

Study 2 also allowed for an examination of the effects of gender on this research. Historically, video games have had unequal participation by gender, at least in part because game manufacturers did not initially target girls [103, 110]. Certain popular genres and styles of games have been found to be more appealing to boys and men, such as first-person shooters and other games with violent feedback [69]. There also exists the sociocultural context of gatekeeping in gaming culture [112]. For example, the controversy surrounding 'Gamergate' in 2014 was symptomatic of the toxic and frequently misogynistic culture that can be prevalent among gamers on the internet [34].

Casual mobile games like match-3 puzzle games and word games have offered a more welcoming environment for women than the traditionally masculine oriented games that existed on consoles or PC [28, 139]. This inclusivity was one of the reasons why a puzzle word game was chosen to be the game implemented for this

study. Although these "casual" games are video games by all technical definitions, 'gatekeepers' categorize them as fundamentally different precisely because they are played by a diverse (or majority non-male) audience. This 'gatekeeping' perspective is so pervasive that it was reflected by a participant who identified as a woman, Claudia, when she responded to the post-task questionnaire. She indicated that she did not play video games often, but at the same time responded that she regularly played mobile word/puzzle games (or 'casual' games). This suggests that she did not consider mobile games to be 'real' video games, suggesting that she had internalized gatekeeping discourse.

7.3 Hypotheses Revisited

In Chapter 1, several hypotheses related to each research question were presented. In light of the findings, each hypothesis will be re-examined and adjusted if necessary.

7.3.1 Player Behaviors and Experiences

The first main research question concerned the influence of productive and unproductive design frictions on players' behaviors and experiences. The hypotheses were broken down into predictions of how productive obstructions, unproductive obstructions, productive juice, and unproductive juice would affect player behavior and experience.

The first hypothesis was that *productive obstructions will lead to more reflective gameplay behaviors*. The findings supported this prediction, though perhaps not in the intended manner. The prompt window was designed to create a microboundary and promote reflective behavior; however, there were only a few explicit examples

of this in the qualitative study sessions. Originally, it was thought that the prompt window would have more of an effect than it appeared to have. The most influential obstruction that caused players to change their behavior by pausing and selecting longer words was the simple addition of the Play Word button. The button changed the controls so that players were afforded the opportunity to more easily modify their selection. In this case, the microboundary of the button combined with the affordance of an alternative action created an opportune context for players to change their behavior. This finding could extend the idea of microboundaries with the implication that creating alternative action pathways for users is important if the goal is behavior change.

The second hypothesis was that *unproductive obstructions will worsen the overall player experience*. The evidence from the findings supported this hypothesis as well. The analysis of the qualitative study sessions showed that the vast majority of players found unproductive obstructions unappealing in *Brave New Word*. The obstruction of having to select each letter individually proved to be too significant of an impediment. Players found that it excessively slowed down the gameplay and made the game unenjoyable. This was an important finding because it confirmed that obstructions for the sake of obstruction do not necessarily have a positive impact. Interestingly, the inclusion of the button feature in the unproductive obstruction version still led to the mindful interactions seen in productive obstruction. This provided evidence that the button feature was the main contributor to behavior change.

The third hypothesis was that *productive juice will improve the overall player experience and aesthetic appeal*. The findings mostly supported this hypothesis. The analysis of the verbal data from the interviews and think-alouds suggested that the majority of participants preferred the productive juice version over the other variations of *Brave New Word*. This was unsurprising given the other studies that

have reported on the effects of juice [56, 72].

The fourth and final hypothesis for this research question was that *unproductive juice will be distracting and worsen the overall player experience*. The findings were very much aligned with this hypothesis. Participants frequently complained about the distractions in the unproductive juice version of *Brave New Word*. Excessive screen shaking was one of the most commonly disliked features. The analysis of the interview and think-aloud data showed that most participants preferred not to have these features in the game.

7.3.2 Game Performance

The second main research question revolved around the effects of productive and unproductive design frictions on game outcomes. The four hypotheses for this research question were also separated into the four types of design frictions introduced in the game.

The first hypothesis was that *productive obstructions will result in higher game performance metrics*. The findings from the logistic regression analyses suggested that some game metrics were improved by productive obstructions. For example, productive obstructions were correlated with longer average words. At the same time, fewer words were played in each round. Unsurprisingly, obstructions seem to affect various game metrics differently. Thus, it is important to consider the relationship between an obstructive design and the metric that is being targeted. As mentioned in the previous section, it appears that the effects on game performance came mostly from the button feature and not the prompt window.

The second hypothesis was that *unproductive obstructions will result in lower game performance metrics*. The findings from the logistic regression analyses showed

that there were mixed results in terms of game performance—some metrics were lower while others were higher. There was a correlation with lower *High_ Rarity* values, longer average words, and fewer words played in each round. In light of the analysis of the verbal data, the longer average words are most likely a consequence of the microboundary created by the Play Word button. However, the lower *High_ Rarity* values could be attributed to the unproductive obstructions since this relationship did not exist in the productive obstruction version. Assuming that the higher game performance was a result of the unintended productive obstruction in the form of the button, these results suggest that this hypothesis remains viable.

The third hypothesis was that *productive juice will result in higher game performance metrics*. The findings from the logistic regression analyses suggested that this prediction was not accurate. In fact, the findings showed that productive juice was not correlated with any changes in game performance—positive or negative. This contrasted with the results of Kao's [72] study on juiciness which found increased performance for games with moderate amounts of juice. This difference may be a result of the game genres used in the studies—Kao's study used a 3D action RPG while this study was on a casual 2D word game. Other factors could have included the degree and/or quality of the juicy effects implemented for *Brave New Word*. Future research questions should explore the space of what amount and/or quality of juice might be required to result in higher game performance.

The last hypothesis was that *unproductive juice will result in lower game performance metrics*. The logistic regression analyses suggested that this prediction was accurate. The *High_ Rarity* values were negatively correlated with unproductive juice versions of the game. None of the other variables were significantly correlated with unproductive juice. This finding is congruent with other works that have studied the effects of distraction on cognition and performance [106, 107] and the effects of ex-

treme amounts of juice in games [72]. Unsurprisingly, too much juice and distracting feedback seemed to hinder players' performances.

7.4 Advantages and Disadvantages of Methodology

7.4.1 Gathering Telemetry Data

The second research question for this dissertation examined what effects productive and unproductive design frictions might have in a casual word game. In order to address this question, this study gathered game log data from over 200 participants. By collecting individual game actions at a detailed granularity, it was possible to reconstruct the gameplay of each individual at a fine level of detail, including at which millisecond they selected a specific letter on the board. To allow for action organization and filtering, each logged entry had a "parent key" and "key" that identified the category of action that was being logged. This made it much easier to analyze the data at different levels of granularity. This was important for the study because it is difficult to foresee what types of data will end up being essential to a particular research question. In this case, the most useful data was at the *word submission* level of granularity; it was not important to analyze which letters were being selected. However, it is entirely possible that the existing data could be analyzed at different levels to answer other research questions.

While the amount of data collected led to insights about player performance, there were also disadvantages to this method. First, the recruitment of participants was not a truly random process. The game was advertised through popular social media channels, so there was a natural "convenience sample" of participants who were either personally close to the researcher or only a couple of degrees away in the

social network. Second, there could have been some amount of selection bias even in the participants who were not personal contacts of the researcher, since there is likely a particular type of person who would opt-in to play a mobile game for a scientific study. Finally, because the data collection for this section was anonymous even to the researcher, there was no demographic information to connect to the gameplay data. Therefore, it was impossible to make connections to potentially key factors such as previous experience with word games. However, due to the fairly large sample size, the hope was that even with the convenience sample there would be a diverse enough sample to counteract any potential bias in the data.

7.4.2 Measuring Player Experience in the Study Sessions

In reference to the first research question about how productive design frictions influence players' behaviors and experiences, one of the biggest challenges was determining how to measure player experience, especially in the qualitative study where each participant played all seven versions of the game. The participants could have been given a standard PX survey after each round like the Player Experience of Need Satisfaction (PENS) [113] or the Player Experience Inventory (PXI) [134]. However, it seemed likely that players would quickly tire of taking the same survey repeatedly within a short period of time. The study thus relied mainly on interviews and think-alouds for measuring PX. As discussed in Chapter 4, these are entirely appropriate and standard methods that fit the context of this study well.

On the other hand, this reliance on interviews and think-aloud data made it difficult to precisely compare the various frictional versions of the game. One of the advantages of using a survey is that it provides quantifiable results for characteristics such as "ease-of-use" or "aesthetics" which would have facilitated making comparisons.

Verbal data from interviews and think-alouds are much more subjective (though it should be noted that surveys themselves can be quite subjective as well despite providing numerical data [1, 65]). Future studies could incorporate PX surveys in the data collection or develop novel methods for quantitatively measuring player experience during a think-aloud / interview session. Accurately quantifying specific aspects of player experience without surveys or questionnaires would be a significant contribution to the field of player-computer interaction.

7.4.3 Qualitative Coding Methodology

In terms of the first research question, First Cycle and Second Cycle coding stages as described by Saldaña [118] were used to describe the verbal data segments from the interviews and think-alouds. The participant data were segmented according to which version of the game they were playing at the time. By organizing the data by game version, it was possible to distinguish patterns in how players reacted to the features in each version. The qualitative coding scheme also labeled participants' reactions to each individual frictional feature. By doing this, it made the process of analyzing the mechanism of each design friction much simpler. Instead of painting a broad picture of the participant's response to each version, the codes were able to provide richer insight into the impacts of individual features.

Some disadvantages to this qualitative coding methodology relate to the topics chosen. Since the coding granularity focused on the individual frictional features, other aspects of player experience and behavior may have been missed. While distinguishing players' responses to each feature was useful (e.g., the player's reaction to screen shakes), coding in this manner could have obscured some facets of the data. For example, data about the Play Word button could have encompassed various as-

pects of the gameplay (e.g., "how long it takes to press the button" or "the controls for selecting letters are different because of the button"), which could invite different analyses and interpretations. However, this nuance of information was usually captured well by other types of qualitative codes, such as the perception and emotion codes.

In addition, the coding only analyzed the verbal data. No attempts were made in this study to analyze video data, whether it be detailed game actions or facial cues. Video coding and analysis could have revealed further insights into player experiences and behaviors that were unable to be made from the verbal data alone. However, due to resource limitations, only the transcriptions of the think-alouds and interviews were able to be analyzed.

7.5 Future Directions

The findings in this study provide implications for future work concerning both obstructions and juicy design in games. The primary aim of this dissertation was to understand the relationship between productive friction and games, and that led to the specific investigation of obstructions and juice.

So far, the findings indicate that it is worth exploring the concept of obstructions and microboundaries in games further. While this study explored the general effects of obstructions and served primarily as a proof-of-concept for the idea of productive obstructions, future work could form specific research questions concerning the mechanisms of *how* obstructions can cause reflection. There are also unanswered questions about what the guiding principles should be to ensure that the obstructions are not overly burdensome to the players and cause negative experiences. There is also the question of generalizability: how would productive obstructions fare in more

complex games or other genres of games? Cox et al.'s [30] work on microboundaries also identified three potential benefits, but this work only demonstrated one of those benefits clearly taking effect—guiding the user towards a specific action. Future studies could focus on game designs intended to demonstrate the other two potential benefits, (1) increased accuracy from speed accuracy trade-offs and (2) avoiding actions that might not align with personal values.

The findings also show that juicy game design is a topic in need of further research. Despite the intuitive belief in mainstream circles that juiciness is key to a good game, there was only minor evidence suggesting that productive juice led to improved player enjoyment. In contrast to the original hypothesis, productive juice did not lead to any changes in game performance—positive or negative. These two findings are mostly consistent with existing literature that show aesthetic improvements through juice but limited effects on objective game outcomes [56]. However, Kao [72] found that moderate amounts of juice improved game performance compared to having no juice or extreme amounts of juice. This difference could be explained by the differences in game genre or perhaps the control version for *Brave New Word* being more juicy than originally thought. Either way, this facet of juiciness should be explored further. Another finding was that unproductive juice negatively affected players' experiences of the game by creating massive distractions during gameplay. Future works could investigate the fine line separating engaging juicy feedback and excessively distracting juice. Principles to guide game designers who wish to avoid negatively impacting players through excessive juice would be a great contribution to the field. A complementary investigation could find specific principles for designing juicy feedback that improves player experience. Other studies could further explore whether productive juice can lead to improved game performance by not only increasing players' *perceived* competence but their *actual* competence levels as well.

Related to this, there is a need for more efficient methods of accurately measuring various player experience attributes. A major limitation in this study with respect to juicy design was the lack of precise player experience measures. The analysis of the interviews and think-alouds showed that there seemed to be a preference for versions with productive juice; however, it was difficult to gain more insights beyond that due to the design of the study. Also, the lack of quantitative player experience data made it hard to make precise comparisons among the seven versions of the game. An invention of a novel player experience measurement method would be a great contribution to the field and allow for improved study designs in the future.

In addition, future work could examine the effects of productive obstructions and juice on player engagement through a longitudinal study.

Chapter 8: Conclusion

8.1 Productive Friction in Game Design

The main objective of this study was to investigate the ways in which productive and unproductive design frictions in a game influence players. For this purpose, two primary research questions were examined:

1. How do productive and unproductive design frictions in a casual word game influence players' behaviors and experiences?
2. What effects might productive and unproductive design frictions in a casual word game have on game outcomes?

Specifically, two kinds of productive design frictions were investigated, obstructions and juice. It was hypothesized that productive obstructions would act as microboundaries and prompt players to have "mindful" interactions in the game by engaging in System 2 thinking. Additionally, it was predicted that these productive obstructions would lead to an improved game performance. It was found that some obstructions were more effective than others at bringing about change, and the affordances of alternative action pathways appeared to be related to this outcome. Another conjecture was that appropriately juicy design would enhance the overall player experience and potentially boost game performance as well. Although it appeared that juiciness did improve overall PX, there was not enough evidence to suggest that it increased game performance.

This study was originally motivated by findings related to frictional game design patterns [90], microboundaries [30], and juiciness in games [56, 67, 72]. Research on these topics has indicated the complex array of factors that go into designing a

game, especially when the designer has specific aims in mind with respect to affecting player experience and behavior. Whether the goal is to record learning activities, promote mindful interactions, or improve player experience, a designer has many tools at their disposal, including productive friction. While previous work has shown the effectiveness (or at least the theoretical ability) of obstructions and juiciness in games, there is still the open question of *how* these frictions might be designed to optimize for these effects. In addition, this study attempted to create a category called productive friction to unify these distinct design concepts under one umbrella.

The framework for productive friction was drawn from many sources, including literature in the learning sciences, human-computer interaction, and game design. The theoretical perspectives proposed by education scholars such as Kapur's productive failure [73, 77, 78] and VanLehn's impasse-driven learning [13, 135] provided a helpful lens through which to view how typically negative experiences can have cognitive benefits. Similarly, some scholars of UX and HCI have explored alternative philosophies of design that emphasize mindful and reflective interactions as opposed to completely frictionless engagement [30, 50, 123]. For example, some research focused on what could be called *intentional design frictions* [30, 91] and how they could guide users toward mindful interactions and enhanced user experience. However, these studies have focused mainly on either learning environments or non-game software.

This work sought to extend these ideas to the realm of video games. Integrating the perspectives of productive failure (and its related theories) and intentional design frictions into the concept of obstructions in games [9] allowed for a clear structure to reason about the mechanisms and situations in which obstructions can be beneficial to players. Also in the area of game design, juiciness has been a popular topic of discussion in media [43, 64, 120] but empirical studies are much more difficult to

find [56, 67, 72]. These ideas of productive challenges, intentional design frictions, and juicy game design inspired the framework of productive friction in the context of games.

To investigate productive friction, a casual word search game called *Brave New Word* was designed and developed from the ground up. It was custom designed so that two examples of productive friction could be studied: obstructions and juiciness. As described in Chapter 3, several frictional features were designed and implemented for different versions such as productive obstruction and productive juice. Study 1 was a between-subjects experimental design, and participants were randomly assigned to one of seven versions so that their performances could be compared across all versions. In contrast, Study 2 utilized a within-subjects approach so that all participants played through all seven versions.

The main findings of this dissertation helped to clarify what effects productive frictions have on player outcomes. They also hinted at the potential challenges of designing for productive friction in games. There were many factors to consider including not overwhelming the player's cognitive load and aligning design goals with game strategies. In addition, this study revealed several difficulties when measuring player experience. In particular, getting accurate estimations without the use of surveys proved to be challenging. Interviews and think-alouds provided valuable PX data, but in some ways they were lacking. Alternative methods of precisely measuring PX without resorting to burdensome questionnaires would be a valuable asset. The results also showed how important it is to have a game that can support multiple viable strategies when studying productive frictions (or any design feature that could affect players' behaviors). As such, future work interested in changing player behavior through specific game design mechanics would benefit from exploring games that involve a wide variety of options for play style.

8.2 Relation to Other Work

Several studies related to juiciness and design frictions that were carried out independently contain many similarities to the approach used in this work. For example, in an effort to determine the effects of varying levels of juiciness in a video game, Kao [72] conducted a quantitative investigation of players interacting with an action role-playing game that had four different versions—each with varying degrees of juiciness, from no juice to extreme amounts of juice. The study employed Amazon's Mechanical Turk system to recruit thousands of participants to play the game and answer questionnaires to measure their PX. The study's author also logged game metrics and player actions so that the results could be compared across versions. Kao's experiment found that having no juice and extreme amounts of juice were consistently worse than having moderate or high amounts of juice in terms of game outcomes and player experience.

Although Kao's study shares many features with this current dissertation, there are important distinctions that make them complementary with respect to their research goals. For instance, Kao's work sought to answer the questions of *what* effects varying amounts of juiciness in games had on players, while this dissertation also aimed to answer qualitative questions such as *how* certain game elements impacted players. The qualitative aspect of this present work complements and adds to the work done by Kao and others who have contributed to the empirical study of juiciness.

8.3 Limitations

There are limitations that should be taken into consideration when interpreting the findings in this dissertation. Although all attempts were made to make sure the studies were carefully controlled, they were not ideal in design. First, the telemetry data were not collected from a truly random sample, so care should be taken when generalizing results from Study 1. Second, there was no reasonable option for logging whether or not players had their devices on silent or had turned off the audio. Thus, it is challenging to make any concrete conclusions about the effects of audio on players in this study.

For Study 2, as with all other qualitative research that investigates a small sample size, the qualitative study sessions should not be used to generalize to a larger population. Rather, the study sessions should be seen as examples of how productive friction designs can impact players and the mechanisms by which they could work. These findings serve to inspire future work that could explore these results in further detail. Also, because all the students recruited for Study 2 were all college students from the same university, this is a common case of a convenience sample. As with all occurrences of convenience sampling, the results should not be extrapolated to more general populations. This was already the case given the small sample size, however. In addition, convenience samples can make studies difficult to replicate. This provides good reason for why future studies should attempt to replicate and verify the findings from this work by expanding the demographics of the samples studied beyond college students. However, this does not invalidate the findings of this study since the qualitative study sessions were meant to examine the details of specific cases and not generalize to the broader population.

Another limitation is that players were not given validated surveys to measure player experience. Future studies could explore larger sample sizes of participants and compare direct measures of player experience using questionnaires like the Player Experience and Needs Satisfaction (PENS) [113] and the Player Experience Inventory (PXI) [134]. Expanding the types of games by exploring games in different genres (e.g. first-person shooter, role-playing games), varying degrees of complexity (e.g. casual mobile games, games with complex systems and mechanics), and educational games would be beneficial for understanding the impact of productive friction more generally.

8.4 Conclusions

Three primary conclusions can be drawn from this study regarding the challenges of designing productive frictions, the challenges of studying player experience, and the importance of designing for multiple game strategies in studies like this one.

8.4.1 Challenges of Designing Productive Frictions

The findings showed that many factors had to be weighed when designing productive frictions. For productive obstructions that were designed to act as microboundaries, explicit prompts were insufficient. Instead, affordances for alternative action paths were significant for creating successful productive obstructions. For instance, the original hypothesis was that a prompt window asking the user to consider their word choice before submitting would engender a desire to find rarer words. In actuality, players seldom modified their selection after seeing the prompt window. Rather, the selection controls of the button naturally made it easier for players to modify their

selections, and they did so more frequently due to the button. This is an example of how an intended design can fail and how another design feature can satisfy a design objective unintentionally.

When designing productive juice, the player's cognitive load and personal preferences had to be taken into consideration. For example, many of the participants in Study 2 thought that the particle and sound effects when submitting a word were satisfying. However, some players considered the effects to be too distracting and overwhelming. What some players thought was a good amount of juice, others might find to be too much. Furthermore, despite relevant literature suggesting that productive juice can increase players' performance [72], that was not found to be the case in this study. The reasons for this discrepancy could be investigated in future works.

8.4.2 Challenges of Studying Player Experience

Confidently and precisely measuring player experience is a top concern for many game designers and researchers. Due to the subjective nature of PX, it is still an ever evolving concept that requires adjustment and growth as the field progresses. One of the challenges faced in this work was accurately measuring PX to allow for comparisons across game versions. Although an argument could be made for using one of the standard PX questionnaires in this study, the decision was made to exclude surveys due to the cumbersome nature of the within-subjects methodology. Since each participant had to play through all seven versions of the game, forcing the player to take the same questionnaire seven times seemed like it would be too burdensome. Despite the verbal data from interviews and think-alouds providing valuable insights, more nuanced knowledge could have been gained if there were alternative methods

for quantifying PX that were less of a burden on the participants.

Another limitation in this work was how PX was measured in the qualitative study sessions because those observations were not longitudinal in nature. Long term engagement could not be measured because the study sessions were only an hour long. On the other hand, gathering anonymous survey data from the participants who downloaded the game in Study 1 may have been a viable option, but the worry was that a survey would be a big barrier to entry. Thus, PX surveys or questionnaires were not included in Study 1.

8.4.3 Designing for Multiple Game Strategies

To properly study productive frictions and changes in player behavior or experience, it is beneficial to design a game that inherently has multiple optimal or near-optimal strategies that players can choose among based on their preferences. This way, microboundaries or distracting juicy effects might be designed in such a way to guide players toward those strategies. If there exists only one clearly optimal strategy, then even with strong interventions, it is highly unlikely that players would want to switch to a different strategy. Thus, it is important that multiple strategies and pathways are available to the player.

In the case of *Brave New Word*, there were various strategies that players could theoretically adopt during the gameplay. For example, a potential strategy would be to find as many short words as possible so that many words could be submitted in a short amount of time. Another strategy—promoted through the productive obstructions—was to search for rarer or longer words to try and score more points with each word. The availability of multiple viable strategies increased the likelihood that players would be willing to change their behaviors.

A conclusion that can be drawn from these findings is that the pausing and reflecting caused by productive frictions should bring into focus what it is that a player should be learning. The obstructions needed to afford the capacity to do the things that were necessary to score higher points (e.g. longer words, rarer words). People learned how to play the game better and searched for more complex words with certain obstructions. For example, although the prompt windows presumably could encourage players to reflect on a word's rarity and score, they did not afford the opportunity to modify the words. On the other hand, the Play Word button allowed players to easily modify their selections; thus, it led to longer words. In order for a productive friction to be successful, it not only emphasized the aspects of the game that the designer is trying to get the player to focus on, but also provide an easy path for the user to engage productively. This resonates with Kapur's work on productive failure [73–78]; Kapur found that failure while exploring multiple representations and solution methods led to greater long term success. In this context, productive friction led to increased game performance and player satisfaction through small obstacles and not necessarily by reaching a metacognitive (or at least player-recognized) "failure state."

8.5 Theoretical Implications

8.5.1 Inherent Strategies Can Override Microboundaries

As mentioned in the previous section, the existence of multiple practical strategies is important for microboundaries to be effective. Although the original objective of the submit prompt windows in productive obstruction was to guide players toward playing rarer words, players resisted this push because of the inherent strategy in the

game that the majority of participants thought was better or easier: play more words, even if they are shorter/common words. They ended up developing this strategy even if at the beginning they stated that they thought they should prioritize playing rarer words because it was easier to find short letter combinations. Microboundaries in games must either align with an inherent strategy that seems feasible, if not optimal, to the player, or else they will resist the attempt to guide their behaviors.

In the end, the microboundary that affected players the most was the micropause created by the Play Word button. The controls with the button allowed users to pause and easily modify their selection before submitting a word. These opportunities gave players a chance to scan the board for better words or additional letters they could add to their selection. This type of affordance is critical for designing effective microboundaries in games.

8.5.2 There Is Such a Thing as Too Much Juice

As the title of this section suggests, excessive juice was found to have a negative impact on players, both in terms of player experience and game outcomes. Extreme visual embellishments and intense audio effects were often distracting and disruptive for players as predicted. Players were unable to concentrate on the task at hand (finding words) and sometimes found the mechanics of the game more challenging (difficult to select letter tiles because of all the animations). This finding was not surprising given the various theoretical perspectives that predict it [56]. In fact, this study provides further confirmation of the theories surrounding juice, and at the same time provides new evidence that juice is only harmful at excessive levels, since players in the productive juice condition showed no decreases in game performance. It also serves to counter certain "extreme" ideas (that are potentially straw man arguments)

in game design circles that idolize juiciness and overemphasize its necessity in games to the point of encouraging excessive amounts of juice. The study by Kao [72] also aligns with these findings that too much juice can have a negative impact on players' experiences and performances.

As important as juicy game design can be to improving the aesthetics of a game and thereby boost the player experience, the findings in this work and other similar studies provide evidence that too much juice can lead to negative outcomes. Game designers should be warned that juiciness can be a double-edged sword that pleases and satisfies some players if the feedback is used in moderation but disrupts and overwhelms players if overdone.

8.5.3 The Many Factors of Player Experience

Game designers and researchers must keep in mind that there are many exogenous *and* endogenous factors that determine player experience. The external stimulus of the game was not the only determinant of what the player will feel, since PX is also largely subjective. Individual preferences, skill-levels, and moods are also major variables that should be taken into consideration when attempting to measure player experience. This was highlighted in the qualitative study sessions when Donna indicated that she preferred the unproductive juicy features despite the fact that they were intentionally designed to be unappealing. When game design elements—such as juicy music or animations—are viewed through the lens of aesthetic appeal, it is clear that there is not a one-size-fits-all policy that will be guaranteed to work for every player. One potential solution to this would be to allow players to personalize game feedback systems to their level of preferred juiciness.

However, this does not mean that game designers and researchers should not

pursue overarching design principles for things like productive friction. As with other areas of design, some design principles and guidelines can be developed to work for a majority of people a majority of the time. As such, future studies could explore juicy design frameworks [55] and investigate appropriate juicy game design principles.

8.6 Practical Implications

8.6.1 Designing Obstructions for Alternative Pathways

A challenge faced by both educational and non-educational game designers is how to direct players toward certain behaviors without completely restricting gameplay or forcing players against their will. Although the general goals may be different, both categories of game designers sometimes need to guide or motivate players to act in specific ways. Productive obstructions offer a potential solution to this challenge through the mechanics of microboundaries and reflective pauses [30]. However, the findings of this study suggest that simply adding a prompt window to invite reflection is not enough to change players' behaviors. That is, a pause or obstruction in and of itself is not enough to achieve a goal. Rather, there should also be a corresponding promotion of alternative pathways for players to engage. For example, although the prompt window with the word score and rarity information was not enough by itself to prompt significant changes in behavior, the Play Word button controls created an environment in which it was more affordant for players to modify their selections. It was this unintentional design that resulted in the most significant effects according to the qualitative data. Thus, designers and researchers should consider if their obstructions can create opportunities for other actions and not just impede

the player's current actions.

8.6.2 Improving Player Experience Through Juicy Design

Though productive juice did not seem to impact game performance, player experiences were improved. Participants found that certain juicy feedback made the game feel more satisfying, such as audio effects in reaction to clicks and explosion particle effects when submitting a valid word. These minor aesthetic improvements leave a major impact on players and can significantly increase their opinions of the game. The extra polish in details like sound effects and animation can take a game from feeling like a prototype to a finished product. Juice, when appropriately implemented, is ultimately a boon for the game designer. Although other studies have shown that juice can also improve performance [72], this work did not find any evidence that suggested the same. There could be many reasons for this, including the genre of the game or the types of juicy effects implemented.

These findings could also be extended to general use software not meant as games, such as educational applications, productivity products, etc. Although games were the main focus of this study, the design principles of juicy feedback and productive friction could be applied to other areas of HCI, though there would have to be other considerations depending on the various contexts. For example, productivity software generally focuses on efficiency and accuracy, so there is potentially a lower threshold for acceptable distractions from juice. It is important to note, however, that this should not be taken as a justification for why this study should be considered important, as there is a long history of researchers being marginalized for studying games for the sake of studying games [20]. Games are such a large part of today's culture and society that researchers should be able to pursue the study of games

without needing to translate its significance to another field.

8.6.3 Moderating Juicy Feedback and Distractions

Despite the potential benefits of juicy design described in the previous section, there appears to be a threshold where the juiciness becomes more of a distraction than an improvement of the player experience. Too much juice can detract from the PX, and as the findings of this study suggest, it can lead to worse game outcomes as well. As Kao [72] concludes, having a moderate amount of juice appears to achieve maximum results. Both having no juice at all or too much juice appear to diminish player outcomes and experiences. This study confirms that finding, and so game designers should use a moderate amount of juicy features when attempting to create the ideal player experience. Unless there is a specific game design reason for incorporating an extreme amount of juiciness, the recommendation would be to aim for a moderate amount of juice.

There are many aspects of juice that require continued investigation. Further work should explore *why* certain levels of juice have these effects. The theories, mechanics, and implications of juicy game design are still relatively unexplored in scholarly literature, which makes it a prime area for future study. In addition, aspects of juiciness in relation to game genre, play style, and other platforms such as AR/VR need to be investigated. Different categories of juice such as audio and visual effects could also be studied independently and jointly in further detail to see if there are any variations in outcomes based on the type of juicy effect.

Appendix A: Post-Task Questionnaire

1. What best describes your gender?
2. How often do you play video games?
   (a) I don't usually play video games
   (b) Less than once a month
   (c) A few times per month
   (d) A few times per week
   (e) Almost always daily
3. When I play puzzle/word games similar to "Brave New Word", I like to (choose as many options as you want):
   (a) Maximize my point total
   (b) Relax
   (c) Waste time and/or distract myself from other activities
   (d) Mess around
   (e) Find the best strategy
   (f) Be better than other players
   (g) Gather achievements
   (h) Solve complex challenges
4. How often do you play similar puzzle/word games (i.e. Candy Crush, Boggle, Wordscapes, Words with Friends, etc.)?
   (a) I don't usually play video games
   (b) Less than once a month
   (c) A few times per month
   (d) A few times per week

(e) Almost always daily

Appendix B: Consent Form

Computational Thinking in Games in Informal Settings - Adult Consent Form
Audio and Video Recording Allowed - No Incentives

Complex Play Lab @ UW-Madison
Research Information & Adult Consent Form
AUDIO AND VIDEO RECORDING ALLOWED
No Incentives

Title of the Study: Computational Thinking in Games in Informal Settings

Dear Participant,

We would like you to be a part of a research project that looks at whether playing educational video games helps people learn, especially in the area of computational thinking (e.g., logical thinking, pattern matching, programming). You will be asked to play one or more of the games within the suite of computational thinking games developed by the Complex Play Lab at UW-Madison. For further information about the games, please either email mberland@wisc.edu or visit the CPL website at complexplay.org. While we do not know if the games will promote learning yet, we are exploring the potential for games to be used as a tool to help people understand computer science content, or change their attitudes about learning computational thinking skills.

Each session will take roughly sixty minutes; you may be asked to play one or more games. Before and after playing the game(s), you may be asked to take part in a short interview and to respond to a short survey. The interview and survey will be about your thoughts and feelings toward games, and your thoughts and feelings toward computer science. The survey will also ask you about what you learned from game play. We may also ask you to 'talk-aloud' during game play and describe your thoughts to us while playing the game. You may be asked to participate in an associated learning activity for each game that you play. If you are, we will record anything you create during these activities.

If you decide to participate, you will be asked to be audio recorded, photographed, and videotaped at times during game play. We will only take recordings if you give us written approval to do so. If you would like to participate, but do not wish to be recorded, that is okay. There is no drawback to you if you are not recorded.

Researchers may also take notes about your activities during the study. We may record identifying information in these notes, however we will replace your name with a pseudonym before storing the notes.

Researchers will have access to data about your game play and will be able to see how you played the game(s). In some cases, if you play more than one of our games, your anonymous user ID may be linked with data across games so that we can see what game play and learning looks like across different games. For example, you may use a different strategy in one game versus another. Some games encourage exploration, while others are more linear.

To help us see whether our games are effective for all players, we will ask you for non-sensitive information that you feel comfortable sharing regarding your views toward school, computer science, and video games.

Only researchers involved with this project at the University of Wisconsin-Madison will be able to see identifiable information that you share with us. Anonymous or de-identified details may be shared in reports or presentations. This information will be used to understand how people play and learn with our games. This research will be conducted at the University of Wisconsin-Madison.

All information collected will be safely stored at the university on password protected and secure servers that few people can access; namely, the researchers involved with this study and IT support. Researchers may use the data gathered for other studies.

Computational Thinking in Games in Informal Settings - Adult Consent Form
Audio and Video Recording Allowed - No Incentives

Participation in this research project is voluntary and involves no unusual risks to you and you may withdraw your permission at any time with no negative consequences.

There are no direct benefits to taking part in this research. However, by taking part in our project, you may help us learn how to improve the development of educational video games. We hope to learn how using a new technology and using games can impact computational thinking skills and attitudes.

If you agree to participate, please indicate this decision below on this page.

If you have any questions about this research or would like to review the game prior to agreeing to participate, please feel free to contact me at mberland@wisc.edu or at (608) 263-7379. If you have questions about your rights as a research subject, you may contact the University of Wisconsin-Madison Social Science Institutional Review Board at (608) 263-2320.

Matthew Berland
Associate Professor, Curriculum and Instruction
University of Wisconsin-Madison

~~~~~~~~~~

Please sign below if you consent to participate in this research project:

I consent to participate in this research project.

_____
(PRINTED name)

_____ _____
(SIGNATURE) *Date*

Please initial below if you consent for your audio/video/photos to be recorded and used in reports.

_____ YES – You may record audio of me during this study and use that data in your research papers and presentations (pseudonymously).

_____ YES – You may record video/photos of me during this study and use that data in your research papers and presentations (pseudonymously).
~~~~~~~~~~

Bibliography

[1] Albaum, G. The likert scale revisited. *Journal of the Market Research Society 39*, 2 (1997), 1–21.

[2] Alexander, C. *A Pattern Language: Towns, Buildings, Construction*. Oxford University Press, Oxford, UK, 1978.

[3] Alha, K., Koskinen, E., Paavilainen, J., and Hamari, J. Free-to-play games: Professionals' perspectives. In *Proceedings of Nordic DiGRA, 2014* (2014).

[4] Alhadreti, O., and Mayhew, P. Rethinking thinking aloud: A comparison of three think-aloud protocols. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada, 2018), ACM Press, pp. 1–12.

[5] Anderson, C., Dalsen, J., Kumar, V., Berland, M., and Steinkuehler, C. Failing up: How failure in a game environment promotes learning through discourse. *Thinking Skills and Creativity 30* (2018), 135–144.

[6] Arnab, S., Lim, T., Carvalho, M., Bellotti, F., de Freitas, S., Louchart, S., and De Gloria, A. Mapping learning and game mechanics for serious games analysis: Mapping learning and game mechanics. *British Journal of Educational Technology 46*, 2 (2015), 391–411.

[7] Atanasov, S. Juiciness: Exploring and designing around experience of feedback in video games. Master's thesis, Malmö University, Malmö, Sweden, 2013.

[8] Ašeriškis, D., and Damaševičius, R. Gamification patterns for gamification applications. *Procedia Computer Science 39* (2014), 83–90.

[9] Barr, P., Khaled, R., Biddle, R., and James, N. Get out of my way! exploring obstruction in popular video games. In *International Academic Conference on the Future of Game Design and Technology* (2006).

[10] Barr, P., Noble, J., and Biddle, R. Video game values: Human–computer interaction and games. *Interacting with Computers 19*, 2 (Mar. 2007), 180–195.

[11] Benford, S., Greenhalgh, C., Giannachi, G., Walker, B., Marshall, J., and Rodden, T. Uncomfortable interactions. In *Proceedings*

of the 2012 CHI Conference on Human Factors in Computing Systems (2012), p. 10.

[12] Björk, S., and Holopainen, J. *Patterns in Game Design*. Charles River Media, Boston, MA, 2004.

[13] Blumberg, F. C., Rosenthal, S. F., and Randall, J. D. Impasse-driven learning in the context of video games. *Computers in Human Behavior 24*, 4 (July 2008), 1530–1541.

[14] Borgo, R., Abdul-Rahman, A., Mohamed, F., Grant, P. W., Reppa, I., Floridi, L., and Chen, M. An empirical study on using visual embellishments in visualization. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2759–2768.

[15] Brühlmann, F., and Mekler, E. Surveys in games user research. In *Games User Research*. Oxford University Press, Oxford, UK, 2018, pp. 141–162.

[16] Buckthal, E. Juiciness in citizen science computer games: Analysis of a prototypical game. Master's thesis, California Polytechnic State University, San Luis Obispo, California, 2014.

[17] Cairns, P., Cox, A., and Nordin, A. I. Immersion in digital games: review of gaming experience research. *Handbook of digital games 1* (2014), 767.

[18] Calvert, S., Appelbaum, M., Dodge, K., Graham, S., Nagayama Hall, G., Hamby, S., and Hedges, L. The american psychological association task force assessment of violent video games: Science in the service of public interest. *American Psychologist 72*, 2 (2017), 126–143.

[19] Cao, J. How to reduce friction with good design, Mar. 2015.

[20] Carter, M., Downs, J., Nansen, B., Harrop, M., and Gibbs, M. Paradigms of games research in HCI: a review of 10 years of research at CHI. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play - CHI PLAY '14* (Toronto, Ontario, Canada, 2014), ACM Press, pp. 27–36.

[21] Carvalho, M., Bellotti, F., Berta, R., De Gloria, A., Sedano, C., Hauge, J., and Rauterberg, M. An activity theory-based model for serious games analysis and conceptual design. *Computers & Education 87* (2015), 166–181.

[22] Cassidy, G., and MacDonald, R. The effects of music choice on task performance: A study of the impact of self-selected and experimenter-selected music on driving game performance and experience. *Musicae Scientiae 13*, 2 (2009), 357–386.

[23] Cassidy, G., and MacDonald, R. A. The effects of music on time perception and performance of a driving game. *Scandinavian journal of psychology 51*, 6 (2010), 455–464.

[24] Cecchinato, M. E., Cox, A. L., and Bird, J. Always on (line)? user experience of smartwatches and their role within multi-device ecologies. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (2017), pp. 3557–3568.

[25] Chanda, M. L., and Levitin, D. J. The neurochemistry of music. *Trends in cognitive sciences 17*, 4 (2013), 179–193.

[26] Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., and Boyle, J. M. A systematic literature review of empirical evidence on computer games and serious games. *Computers & education 59*, 2 (2012), 661–686.

[27] Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., Leaver-Fay, A., Baker, D., Popović, Z., and players, F. Predicting protein structures with a multiplayer online game. *Nature 466*, 7307 (Aug. 2010), 756–760.

[28] Cote, A. C. Casual Resistance: A Longitudinal Case Study of Video Gaming's Gendered Construction and Related Audience Perceptions. *Journal of Communication 70*, 6 (Dec. 2020), 819–841.

[29] Cox, A., Cairns, P., Shah, P., and Carroll, M. Not doing but thinking: The role of challenge in the gaming experience. In *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems - CHI '12* (Austin, Texas, USA, 2012), ACM Press, p. 79.

[30] Cox, A. L., Gould, S. J., Cecchinato, M. E., Iacovides, I., and Renfree, I. Design frictions for mindful interactions: The case for microboundaries. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16* (San Jose, California, USA, 2016), ACM Press, pp. 1389–1397.

[31] Csikszentmihalyi, M., and Csikzentmihaly, M. *Flow: The psychology of optimal experience*, vol. 1991. Harper & Row New York, 1991.

[32] Denisova, A., Guckelsberger, C., and Zendle, D. Challenge in digital games: Towards developing a measurement tool. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17* (Denver, Colorado, USA, 2017), ACM Press, pp. 2511–2519.

[33] Denisova, A., Nordin, A. I., and Cairns, P. The convergence of player experience questionnaires. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16* (Austin, Texas, USA, 2016), ACM Press, pp. 33–37.

[34] Derr, T. M. The road to gamergate: the history of gender stereotyping in video game culture, 1970-2000 and beyond. Master's thesis, Rutgers University, 2020.

[35] Deterding, S. The lens of intrinsic skill atoms: A method for gameful design. *Human–Computer Interaction 30*, 3-4 (May 2015), 294–335.

[36] D'Mello, S., and Graesser, A. Dynamics of affective states during complex learning. *Learning and Instruction 22*, 2 (2012), 145–157.

[37] Dormann, C., Whitson, J., and Neuvians, M. Once more with feeling: Game design patterns for learning in the affective domain. *Games and Culture 8*, 4 (2013), 215–237.

[38] Drachen, A., and Connor, S. Game analytics for games user research. In *Games User Research*. Oxford University Press, Oxford, UK, 2018, pp. 333–353.

[39] Drachen, A., Mirza-Babaei, P., and Nacke, L. Introduction to games user research. In *Games User Research*, A. Drachen, P. Mirza-Babaei, and L. Nacke, Eds. Oxford University Press, Oxford, UK, 2018, pp. 1–11.

[40] Dutton, F. Editors' choice: Why Dead Cells is one of 2018's best games, Dec. 2018.

[41] Electronic Arts. *The Sims 2*. Maxis, 2004.

[42] Firaxis Games. *Civilization III*. Infogrames, 2001.

[43] Gabler, K., Gray, K., Kucic, M., and Shodhan, S. How to prototype a game in under 7 days. *Gamasutra* (Oct 2005).

[44] Gage, Z., and Schlesinger, J. *SpellTower*. Zach Gage, New York, NY, 2011.

[45] Gamma, E. *Design patterns: Elements of reusable object-oriented software*. Addison-Wesley Professional, Boston, MA, 1995.

[46] Gauthier, A., and Jenkinson, J. Designing productively negative experiences with serious game mechanics: Qualitative analysis of game-play and game design in a randomized trial. *Computers & Education 127* (2018), 66–89.

[47] Gee, J. What video games have to teach us about learning and literacy. In *Computers in Entertainment (CIE)*, vol. 1. ACM New York, NY, USA, 2003, pp. 1–4.

[48] Girlfriend Reviews. Should your boyfriend play Celeste, Dead Cells, and Hollow Knight?, Dec. 2018.

[49] Grant, C. The game awards 2018: Here are all of the winners, Dec. 2018.

[50] Grosse-Hering, B., Mason, J., Aliakseyeu, D., Bakker, C., and Desmet, P. Slow design for meaningful interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13* (Paris, France, 2013), ACM Press, p. 3431.

[51] Har-Paz, M. M. Friction as a function in user experience: Sometimes a little extra cognitive load is a good thing., Apr. 2018.

[52] Hassenzahl, M. Aesthetics in interactive products: Correlates and consequences of beauty. In *Product experience*. Elsevier, 2008, pp. 287–302.

[53] Hassenzahl, M., and Monk, A. The inference of perceived usability from beauty. *Human–Computer Interaction 25*, 3 (2010), 235–260.

[54] Hedeen, L. When design friction is a good thing. https://uxdesign.cc/when-design-friction-is-a-good-thing-d4ab56c4049e, Oct. 2020.

[55] Hicks, K., Dickinson, P., Holopainen, J., and Gerling, K. Good game feel: An empirically grounded framework for juicy design. In *Proceedings of the 2018 DiGRA International Conference: The Game Is the Message* (2018), p. 17.

[56] Hicks, K., Gerling, K., Dickinson, P., and Vanden Abeele, V. Juicy game design: Understanding the impact of visual embellishments on player experience. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Barcelona Spain, Oct. 2019), ACM, pp. 185–197.

[57] Hicks, K., Gerling, K., Richardson, G., Pike, T., Burman, O., and Dickinson, P. Understanding the effects of gamification and juiciness on players. *IEEE* (2019).

[58] Hullett, K., and Whitehead, J. Design patterns in fps levels. In *Proceedings of the Fifth International Conference on the Foundations of Digital Games - FDG '10* (2010), pp. 78–85.

[59] Iacovides, I., Cox, A. L., McAndrew, P., Aczel, J., and Scanlon, E. Game-Play Breakdowns and Breakthroughs: Exploring the Relationship Between Action, Understanding, and Involvement. *Human–Computer Interaction 30*, 3-4 (May 2015), 202–231.

[60] IJsselsteijn, W., Poels, K., and De Kort, Y. A. The game experience questionnaire: Development of a self-report measure to assess player experiences of digital games. *TU Eindhoven, Eindhoven, The Netherlands* (2008), 42.

[61] Inbar, O., Tractinsky, N., and Meyer, J. Minimalism in information visualization: attitudes towards maximizing the data-ink ratio. In *Proceedings of the 14th European conference on Cognitive ergonomics: invent! explore!* (2007), pp. 185–188.

[62] Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., and Walton, A. Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies 66*, 9 (2008), 641–661.

[63] Johnson, D., Gardner, M. J., and Perry, R. Validation of two game experience scales: The Player Experience of Need Satisfaction (PENS) and Game Experience Questionnaire (GEQ). *International Journal of Human-Computer Studies 118* (Oct. 2018), 38–46.

[64] Jonasson, M., and Purho, P. Juice it or lose it. https://www.youtube.com/watch?v=Fy0aCDmgnxg, 2012. Accessed: 2019-07-15.

[65] Joshi, A., Kale, S., Chandel, S., and Pal, D. Likert scale: Explored and explained. *British Journal of Applied Science & Technology 7*, 4 (2015), 396–403.

[66] Juul, J. *A casual revolution: Reinventing video games and their players*. MIT press, 2010.

[67] Juul, J., and Begy, J. S. Good feedback for bad players? a preliminary study of 'juicy' interface feedback. In *Proceedings of the First Joint FDG/DiGRA Conference* (2016).

[68] Jørgensen, A. H. Marrying HCI/usability and computer games: A preliminary look. In *NordiCHI '04: Proceedings of the third Nordic conference on Human-computer interaction* (Tampere, Finland, 2004), Association for Computing Machinery, p. 4.

[69] Kafai, Y. B., Heeter, C., Denner, J., and Sun, J. Y. *Beyond Barbie and Mortal Kombat: New Perspectives on Gender and Gaming*. MIT Press, 55 Hayward Street, Cambridge, MA, 2008.

[70] Kahneman, D. *Thinking, fast and slow*. Macmillan, 2011.

[71] Kalyuga, S., Chandler, P., and Sweller, J. Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology 13*, 4 (1999), 351–371.

[72] Kao, D. The effects of juiciness in an action rpg. *Entertainment Computing* (2020), 100359.

[73] Kapur, M. Productive failure. *Cognition and Instruction 26*, 3 (2008), 379–424.

[74] Kapur, M. A further study of productive failure in mathematical problem solving: Unpacking the design components. *Instructional Science 39*, 4 (2011), 561–579.

[75] Kapur, M. Comparing learning from productive failure and vicarious failure. *Journal of the Learning Sciences 23*, 4 (2014), 651–677.

[76] Kapur, M. Learning from productive failure. *Learning: Research and Practice 1*, 1 (2015), 51–65.

[77] Kapur, M. Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist 51*, 2 (2016), 289–299.

[78] Kapur, M., and Bielaczyc, K. Designing for productive failure. *Journal of the Learning Sciences 21*, 1 (2012), 45–83.

[79] Kelle, S., Klemke, R., and Specht, M. Design patterns for learning games. *International Journal of Technology Enhanced Learning 3*, 6 (2011), 555–569.

[80] Kim, J., Park, J., Lee, H., Ko, M., and Lee, U. Lockntype: Lockout task intervention for discouraging smartphone app use. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (2019), pp. 1–12.

[81] Knoll, T. The think-aloud protocol. In *Games User Research*. Oxford University Press, Oxford, UK, 2018, pp. 189–202.

[82] Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E., and Sinnelä, A. UX Curve: A method for evaluating long-term user experience. *Interacting with Computers 23*, 5 (Sept. 2011), 473–483.

[83] Kühn, S., Kugler, D., Schmalen, K., Weichenberger, M., Witt, C., and Gallinat, J. Does playing violent video games cause aggression? a longitudinal intervention study. *Molecular Psychiatry* (2018).

[84] Lankoski, P., and Björk, S. Gameplay design patterns for believable non-player characters. In *Proceedings of DiGRA 2007 Conference* (2007), pp. 416–423.

[85] Lee, J., Kladwang, W., Lee, M., Cantu, D., Azizyan, M., Kim, H., Limpaecher, A., Gaikwad, S., Yoon, S., Treuille, A., and Das, R. Rna design rules from a massive open laboratory. *Proceedings of the National Academy of Sciences 111*, 6 (2014), 2122–2127.

[86] Lewis, C. Using the 'thinking-aloud' method in cognitive interface design. *Research Report RC9265, IBM TJ Watson Research Center* (1982).

[87] Lewis, C., Wardrip-Fruin, N., and Whitehead, J. Motivational game design patterns of 'ville games. In *Proceedings of the International Conference on the Foundations of Digital Games - FDG '12* (2012), p. 172.

[88] Mahlke, S. Visual aesthetics and the user experience. In *Dagstuhl Seminar Proceedings* (2008), Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

[89] McMillan, J., and Schumacher, S. *Research in education: A conceptual introduction*. Allyn & Bacon, London, England, 2001.

[90] Mechtley, A., and Berland, M. Frictional patterns in the design of games for learning. In *Proceedings of ICLS 2018* (2018), pp. 1487–1488.

[91] Mejtoft, T., Hale, S., and Söderström, U. Design friction. In *Proceedings of the 31st European Conference on Cognitive Ergonomics - ECCE 2019* (BELFAST, United Kingdom, 2019), ACM Press, pp. 41–44.

[92] Miles, M. B., Huberman, A. M., and Saldaña, J. *Qualitative data analysis: a methods sourcebook*, third edition ed. SAGE Publications, Inc, Thousand Oaks, Califorinia, 2014.

[93] Mitgutsch, K., and Weise, M. Subversive game design for recursive learning. In *Proceedings of DiGRA 2011 Conference* (2011), pp. 1–16.

[94] Motion Twin. *Dead Cells*. Motion Twin, Bordeaux, France, 2018.

[95] Myers, D. *The Nature of Computer Games: Play As Semiosis (Digital Formations; V. 16)*. Peter Lang Publishing, 2003.

[96] Nacke, L., and Drachen, A. Towards a framework of player experience research. In *Proceedings of the Second International Workshop on Evaluating Player Experience in Games at FDG, 11* (2011).

[97] Nielsen, J. Thinking aloud: The #1 usability tool.

[98] Nielsen, J. Usability 101: Introduction to usability.

[99] Nintendo Co., Ltd. *Super Mario Brothers*. Nintendo, 1985.

[100] Norman, D., and Nielsen, J. The definition of user experience (ux).

[101] Norman, D. A. *Emotional design: Why we love (or hate) everyday things*. Basic Civitas Books, 2005.

[102] North, A. C., and Hargreaves, D. J. Musical preferences during and after relaxation and exercise. *American journal of psychology 113*, 1 (2000), 43–68.

[103] Ogletree, S. M., and Drake, R. College Students' Video Game Participation and Perceptions: Gender Differences and Implications. *Sex Roles 56*, 7-8 (Apr. 2007), 537–542.

[104] Oliphant, T. E. *A guide to NumPy*, vol. 1. Trelgol Publishing USA, 2006.

[105] Pagulayan, R. J., Keeker, K., Fuller, T., Wixon, D., Romero, R. L., and Gunn, D. V. User-centered Design in Games. In *Handbook for Human-Computer Interaction in Interactive Systems*, J. Jacko and A. Sears, Eds. Lawrence Erlbaum Associates, Inc., Mahwah, NJ, 2003, pp. 884–906.

[106] Park, B., Flowerday, T., and Brünken, R. Cognitive and affective effects of seductive details in multimedia learning. *Computers in Human Behavior 44* (2015), 267–278.

[107] Park, B., Moreno, R., Seufert, T., and Brünken, R. Does cognitive load moderate the seductive details effect? a multimedia study. *Computers in Human Behavior 27*, 1 (2011), 5–10.

[108] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research 12* (2011), 2825–2830.

[109] Pew Research Center. *Americans' thoughts about video games.* Pew Research Center, Washington, D.C., Dec. 2015.

[110] Rabasca, L. The internet and computer games reinforce the gender gap. *Monitor on Psychology 31*, 9 (2000), 1.

[111] Reddit, Inc. Homepage - Reddit. https://www.redditinc.com/, 2020.

[112] Richard, G. T. Intersecting vulnerabilities in game culture: the effects of inequities and stereotype threat on player confidence, identification and persistence across gender and race. In *Proceedings of DiGRA* (2015), pp. 1–4.

[113] Rigby, S., and Ryan, R. The player experience of need satisfaction (pens) model. *Immersyve Inc* (2007), 1–22.

[114] Rockstar Games. *Grand Theft Auto: San Andreas.* Rockstar North, 2004.

[115] Rogers, K., Jörg, M., and Weber, M. Effects of Background Music on Risk-Taking and General Player Experience. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Barcelona, Spain, Oct. 2019), ACM, pp. 213–224.

[116] Rohrer, C. Quantifying and comparing ease of use without breaking the bank. https://www.nngroup.com/articles/pure-method/, Apr. 2017.

[117] Roschk, H., Loureiro, S. M. C., and Breitsohl, J. Calibrating 30 years of experimental research: a meta-analysis of the atmospheric effects of music, scent, and color. *Journal of Retailing 93*, 2 (2017), 228–240.

[118] Saldaña, J. *The coding manual for qualitative researchers.* Sage, 2015.

[119] Salen, K., and Zimmerman, E. *Rules of play: Game design fundamentals.* MIT press, 2004.

[120] Schell, J. *The art of game design: A book of lenses.* Elsevier/Morgan Kaufmann, Amsterdam ; Boston, 2008.

[121] Seabold, S., and Perktold, J. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference* (2010).

[122] Seif El-Nasr, M., Aghabeigi, B., Milam, D., Erfani, M., Lameman, B., Maygoli, H., and Mah, S. Understanding and evaluating cooperative games. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI '10* (2010), p. 253.

[123] Sengers, P., Boehner, K., David, S., and Kaye, J. J. Reflective design. In *Proceedings of the 4th decennial conference on Critical computing between sense and sensibility - CC '05* (Aarhus, Denmark, 2005), ACM Press, p. 49.

[124] Spradley, J. *The ethnographic interview.* Wadsworth, London, 1979.

[125] Statista. Mobile OS market share 2019. https://www.statista.com/statistics/272698/global-market-share-held-by-mobile-operating-systems-since-2009/.

[126] Statista. Most popular social media apps in U.S. 2019. https://www.statista.com/statistics/248074/most-popular-us-social-networking-apps-ranked-by-audience/, 2019.

[127] Strauss, C., and Fuad-Luke, A. The slow design principles. *Proceedings of the Changing the Change 14* (2008).

[128] Suits, D. B. Use of dummy variables in regression equations. *Journal of the American Statistical Association 52*, 280 (1957), 548–551.

[129] Swink, S. *Game feel: a game designer's guide to virtual sensation*. Morgan Kaufmann Publishers/Elsevier, Amsterdam; Boston, 2009.

[130] The pandas development team. *pandas-dev/pandas: Pandas*. Zenodo, Feb. 2020.

[131] Turoff, A. *Boggle*. Hasbro, Inc., 1972.

[132] Unity Technologies. *Unity*. Unity Technologies, San Francisco, CA, 2005.

[133] Valve Corporation. *Half-Life 2*. Valve Corporation, 2004.

[134] Vanden Abeele, V., Nacke, L. E., Mekler, E. D., and Johnson, D. Design and preliminary validation of the player experience inventory. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts* (2016), ACM, pp. 335–341.

[135] VanLehn, K. *Toward a Theory of Impasse-Driven Learning. In Learning issues for intelligent tutoring systems*. Springer, New York, NY, 1988.

[136] Vygotsky, L. *Language and thought*, 2nd ed ed. MIT Press, Cambridge, MA, 1986.

[137] Wijman, T. The world's 2.7 billion gamers will spend $159.3 billion on games in 2020; the market will surpass $200 billion by 2023, May 2020.

[138] Wiseman, S., Cox, A. L., and Brumby, D. P. Designing Devices With the Task in Mind: Which Numbers Are Really Used in Hospitals? *Human Factors 55*, 1 (Feb. 2013), 61–74. Publisher: SAGE Publications Inc.

[139] Wohn, D. Y. Gender and race representation in casual games. *Sex Roles 65* (Aug. 2011), 198–207.

[140] Young, V. Strategic UX: The art of reducing friction, 2015.

[141] Zagal, J., Björk, S., and Lewis, C. Dark patterns in the design of games. foundations of digital games, 2013.