Level Up: Supporting In-Game Skill Development

A dissertation submitted to the
College of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Computer Science
University of Saskatchewan
Saskatoon

By
Colby Johanson

Permission to Use

In presenting this dissertation in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my dissertation work was done. It is understood that any copying or publication or use of this dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my dissertation.

Disclaimer

Reference in this dissertation to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement, recommendation, or favoring by the University of Saskatchewan. The views and opinions of the author expressed herein do not state or reflect those of the University of Saskatchewan, and shall not be used for advertising or product endorsement purposes.

Requests for permission to copy or to make other uses of materials in this dissertation in whole or part should be addressed to:

Head of the Department of Computer Science
176 Thorvaldson Building, 110 Science Place
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5C9 Canada

OR

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9 Canada

Abstract

Video games are challenging and complex. They require players to master a diverse set of skills to succeed. Through play, players acquire and eventually master these skills, transitioning from novice to expert through skill development. Making progress and performing well in a game is directly tied to a player's ability to master in-game skills, so players are strongly motivated to get better at the games they play. Games can do a good job of supporting a new player's learning, but too often they leave a player to work out for themselves how to improve and get better at the game. The problem is that game designers do not always know how to support skill development in their games. To solve this problem, we need to better understand how skill learning occurs in games, as well as explore specific new approaches for supporting skill learning in games. Games are not the only context in which skill development and high performance is important — the field of human performance already explores this in detail and provides many theories to apply to this new domain. Inspired by these theories I explore different ways of supporting players' learning at two different stages of skill development. First, I explore how early learning can be supported through the use of *guidance* and explore how later learning can be supported by *modifying practice*. Testing out the effects of guidance by providing new players with different levels of navigation guidance and evaluating how well they were able to learn the environment, I found that guidance improved a player's immediate performance and allowed them to complete tasks within the game more effectively. I evaluated the idea of modifying practice by applying spaced practice (having players take breaks when playing) in two different games, as well as by adding checkpoints to a side-scrolling platform game. I found that having players take breaks improved players' immediate performance and allowed them to make more progress within the game and that a variety of break lengths were effective. I found that checkpoints allowed players to make progress in the game and learn the game just as effectively as when checkpoints were not present. Overall, this research adds to our understanding of how skill development occurs in games and provides some concrete examples of how support methods used in other contexts (such as in sports) can be applied to digital gaming.

Acknowledgements

I will be forever grateful for the support and guidance provided by my supervisors, Dr. Regan Mandryk and Dr. Carl Gutwin. None of this would be possible without your continued patience, wisdom, and feedback. Thank you.

I am also grateful to my family for their support and encouragement in this journey. Thank you, Kristin, Gary, Riley, Kellie, Gracie, Elliot, and Samuel.

Finally, I am grateful to my collaborators and the many friends I made along the way, many of whom provided me with further guidance, support, and feedback in their own way. Thanks to Ansgar, Brandon, Jason, Julian, Katelyn, Martin, Matthew, Max (B and F), Michael, Phaedra, Rod, Sami, Susanne, and anyone else I may have missed.

Contents

List of Tables

List of Figures

List of Abbreviations

ACS	Attentional Control Scale
ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
DDA	Dynamic Difficulty Adjustment
FKS	Flow Scale Short
FPS	Frames Per Second
FPS	First-Person Shooter
FSS	Flow Short Scale
GPS	Global Positioning System
HIT	Human Intelligence Task
IMI	Intrinsic Motivation Inventory
ITQ	Immersive Tendencies Questionnaire
KP	Knowledge of Performance
KR	Knowledge of Results
RPG	Role-Playing Game
MANCOVA	Multivariate Analysis of Covariance
MTurk	Mechanical Turk
NPC	Non-Player Character
OMSA	Object-Manipulation Spatial Ability
PXI	Player Experience Inventory
QCM	Questionnaire on Current Motivation
RM-ANCOVA	Repeated Measures Analysis of Covariance
RT	Reaction Time
SA	State Anxiety
SIMS	Situational Motivation Scale
SNA	Spatial Navigation Ability
SOQ	Sport Orientation Questionnaire
TLX	Task-Load Index
VE	Virtual Environment
VM	Visual Memory

Part I

Introduction

1 Introduction

1.1 Problem and Motivation

In digital games, learning and skill development are critical aspects of a game's experience. In fact, the commercial success of a game is often related to how easy it is for new players to learn and improve their performance within the game [54]. Consider, for example, how an early game designer of coin-operated arcade games from Atari, Nolan Bushnell is quoted as saying: "All the best games are easy to learn and difficult to master. They should reward the first quarter and the hundredth." In the days of arcade cabinets, if a game was too complicated to learn then players would simply move on to one of the many other cabinets in the arcade [20]. Once players began playing, the game had to maintain a player's interest by providing the right balance in terms of challenge [237]: If the game was too easy, players would get bored; if the game was too hard, players would get frustrated. Players motivated to succeed wouldn't be deterred by a failure, and would instead willingly spend coin after coin to play again, with the hope of progressing further than last time, gaining mastery with each attempt.

Therefore, players remain engaged with a game if the game successfully allows the players to learn and master the various skills they require to succeed. Today's digital games may be more complex than those of the past, but they are still designed with skill development in mind. Doing this successfully for today's games may not directly result in coins being inserted into a machine, but the commercial implications are just as, if not more present. Video games are a big business: in the United States, the video game industry generated \$90.3 billion in annual economic output in 2019, supporting nearly 429,000 jobs [91] (in Canada in 2021, the industry generated \$3.7 billion CAD in revenue while employing 32,300 people [92]). Games that can keep players engaged and provide a positive experience are likely to receive better ratings. This can attract new players and keep players coming back to play the game over the long term. This affects players' purchasing decisions, whether that is the initial purchase of the game, the desire to purchase additional content or sequels, or to buy in-game items. Some games even require players to pay to continue playing the game over time — for example, *World of Warcraft* [31] — requires players to pay a monthly subscription fee.

Players are highly motivated to develop in-game skills because successfully learning those skills is required to make progress. Making progress in the game is important because it acts as a source of intrinsic motivation for players. This is due to how progress is able to fulfill one's need for competence and how that interacts with the need for autonomy [246]. In the case of arcade games, pacing the game's challenge to align with the player's expertise level promotes feelings of competence by providing challenges that allow players to get a bit further in the game with each attempt [237]. This works because it is combined with opportunities to acquire new or improve existing in-game skills, and players are given positive feedback when accomplishing in-game goals [247]. These feelings of competence must be accompanied by feelings of autonomy in order to foster intrinsic motivation — players must be able to attribute their in-game success to their own actions [246] and the player must have the freedom within the game to accomplish their goals [247].

Skill development is therefore a significant reason why players continue to engage with a game. Skill development is the process whereby one initially learns how to carry out and then improve at the execution of a skill [256, 302]. *Skill* refers to the ability to carry out a specific task in order to achieve a specific goal [98, 302, 256, 89].

Skill development is important to the success of a game, and this process of learning skills is accomplished by directly playing the game. Games are therefore often considered to be effective learning environments [167, 178, 114] that often do a good job of teaching players the skills required to improve in-game performance [178]. A well-designed game has clear goals with strong feedback [297], and is a task that players are willing to give their complete attention to [52] — factors that have been shown to facilitate continued skill development [95]. Game designers also care quite a bit about supporting learning and so design games with this in mind, such as by introducing skills one at a time and only when they are applicable [114, 295], or by providing hints that consider the state of the game and are delivered at a time when they could be beneficial [143, 250, 295, 331]. Once a player has learned a new skill, the primary way in which they improve their performance in that skill is through practice; that is, by playing the game. Players must repeatedly apply that skill to overcome challenges, allowing them to practice under a variety of conditions and in different contexts [167, 114, 342, 143]. These challenges start out easy and gradually increase in difficulty [143, 342, 297, 114]. Failure is also an important component of this, as this failure acts as a form of *feedback* that encourages players to adjust their approach and try again [158, 120]. Despite game designers recognizing the importance of skill development and learning, there are still many scenarios in which games do not provide sufficient support to players. Evidence for this comes from players who turn to social media for help. For example, many players post questions on the online discussion website Reddit or watch videos on YouTube that provide video walk-throughs of the game or explain in-game strategies.

As a concrete example, consider a type of game that used to be quite popular in the past but is slowly declining in popularity: the arena shooter. As of January 2023, *Quake Champions* has an all-time peak of only 17,000 players[1]. Meanwhile, the most popular first-person shooter game on the same distribution platform, *Counter-Strike: Global Offensive* [130] has an all-time peak of 1.3 million players [2]. Further, *Fortnite* [93], possibly the most popular shooter game at the moment, has peaked at a record of 15.3 million concurrent players [106]. This might be because *Fortnite* is the easiest of the three games to learn while *Quake Champions* is the hardest. Contrast this to the mid-1990s, where arena shooters such as *Quake* and *Doom* were the most popular games. At that time there were no other, easier-to-learn options, so players stuck with the game and learned how to play. Some players lament the gradual decline in popularity that arena shooters are currently going through (e.g., [310]), which they can recognize is due in large part to new players being unable to learn the game and compete with the experienced players. If these games supported player

[1] See `https://steamdb.info/app/611500/graphs/`.
[2] See `https://steamdb.info/app/730/graphs/`.

learning effectively, then new players may be afforded the opportunity to enjoy the depth and complexity offered by these difficult-to-master games.

The problem is, therefore, that *players want to get better at the games they play but game designers do not always know how to explicitly support skill development in their games*.

1.2 Solution and Contributions

In this dissertation, my[3] overall goal is to increase our understanding of the approaches that game designers can support players in their skill development. From the examples already mentioned within this introduction, it should be apparent that this is something that game designers and players are very interested in. Therefore, there have been many approaches to solving this problem already. However, there is also the possibility of developing new methods that game designers can incorporate into their games to support a player's skill development. I accomplish this by first exploring the existing methods for supporting skill development in games and working towards understanding the underlying theories of skill learning they may be leveraging to be effective. In particular, I make use of the existing theories of human performance and perceptual-motor skill learning that exist in other contexts, such as in sports. The field of human performance is concerned with how one learns new skills and improves over time [104], which is exactly what players of digital games are doing when they play a game and improve over time.

By examining the literate on skill development in other contexts, I found that there are many different ways to support a learner's skill development, and that what a learner is focusing on changes over time. In the early stages of learning, the player requires some instruction or guidance [104] to learn *what* they should be doing and *how* to do it. Later on, the player may know what to do but need to refine their approach [104], so a different support method is required. This work takes several steps toward increasing our understanding of the approaches that game designers can use to support skill development at different parts of the learning process. First, for the early parts of a game where more significant help may be necessary, I found that *guidance* can be an effective tool for helping new players. Unlike in other contexts (where one can become reliant on guidance [257, 16, 274, 135, 182, 343]), my work suggests that guidance in games can improve players' immediate performance and allow them to carry out tasks within the game more effectively, without compromising learning. This goes against many prior findings of guidance, which found that a learner may start to ignore feedback inherent to a task and focus solely on the provided guidance [257]. This appears to not be true in navigation, and may not be true in other game contexts. Guidance is already sometimes found within games, and not just in the context of navigation. For example, many games provide explicit instructions to players to explain important game mechanics or provide hints [143, 295]. It is possible that players may not become reliant on these aids, just as they did not seem to be reliant on the navigation aids within my studies.

Second, for the scenarios where players understand what to do but are unable to execute the skill effectively, I found that *modifying practice* is an effective way of helping players make progress within the game. I found that modifying

[3] Throughout this dissertation, I use the pronouns "I" or "my" outside of the manuscripts and "we" or "our" within the manuscripts. Although I am the sole author of this document, the work presented in each manuscript was completed in collaboration with co-authors. Before each manuscript, I introduce the co-authors and describe what my specific contributions are to the manuscript.

practice by applying spaced practice (whereby one breaks up tasks with rest periods) can help players make progress within a game through increased performance after returning to the game from a break. These performance gains from the break seem to be temporary, however, as performance differences between players who took breaks and those who didn't equalize after both groups take a one-day break from the game. That is, whether or not a player takes a break has little effect in the longer term, but in the short term, it can help a player progress further in the game. I also explored the robustness of spaced practice as a mechanism for improving performance and whether this would help players make progress within a game. I found that short breaks of less than ten seconds could still be effective, that breaks do not need to be presented on a fixed schedule, and that the improved performance can help players make progress within the game.

Finally, I compared my new method of supporting players (game-integrated breaks) to an existing method. I explored how checkpoints affect performance and learning and explained using theories of skill development how checkpoints might be helping players learn the game. Checkpoints modify practice in a way that directly helps players make progress in the game (by minimizing how much progress is lost each time a player fails). More progress within the game means that players are exposed to a greater variety of tasks during practice, and not having to repeat earlier sections of a level means that players are given a chance to focus on sections of the level that they struggle with. I confirmed that checkpoints do in fact help players make progress, but do not have any long-term effects on learning. In other words, playing a game with checkpoints present is more beneficial than playing without checkpoints, as players can make progress in the game more easily and experience more game content, while still effectively learning how to play the game. Finally, compared to checkpoints, game-integrated breaks helped players make progress in the game in a similar way. This shows that novel techniques for supporting development can be designed and integrated into games.

These contributions increase our understanding of how skill development occurs in games, and how game designers can help people become better gamers and become better at the games they play.

1.3 Outline

In Chapter 2, I review relevant theories of human performance, including what a skill is, how skill learning occurs, how skill learning is measured, and different ways in which skill learning can be supported.

There are times in which a player's need to learn an important aspect of a game is very apparent. One of the most apparent times is when a player is first learning how to navigate the game environment. It is through this lens that I explore how a player's early learning might be supported, and this is done within two manuscripts presented in Part I of this dissertation. In Manuscripts A and B I explore *guidance* as a mechanism for supporting the development of skill in the context of navigating through virtual environments. This skill is fundamental to a player's success — any experienced player can navigate through the environment with little difficulty — and yet there are many novices who find this to be difficult. In Manuscript A (Chapter 4), I explore the performance and learning effects of two different types of visual navigation assistance in navigating to different parts of the environment. In Manuscript B (Chapter 5), I extend this work by introducing mechanical guidance where players automatically get taken along a set path to target landmarks and explore how this affects performance in learning.

In addition to issues with learning skills in a game early on, there are many scenarios in games where players know what needs to be done and have already acquired the necessary skills required to overcome the challenge presented by the game, yet they still struggle to succeed. In this scenario, a player may be tempted to simply give up. Part II of this dissertation explores ways of supporting players in this scenario, via the approach of *modifying practice*. In Manuscript C (Chapter 7) and D (Chapter 9) I apply spaced practice in two different games with two different approaches. Manuscript C focuses on determining if spaced practice can be applied to games like it is applied to other tasks, and whether there is a particular rest length that works best. In Manuscript D, this work is extended to consider how spaced practice might be more naturally scheduled within a game and if it can truly be leveraged to aid a player's in-game progress. In Manuscript D, I also explore one other approach to modifying practice that is common to many single-player games — *checkpoints*. Checkpoint systems prevent a player from losing progress and so can improve a player's performance or allow them to make progress within a game. However, they adjust practice in ways that could interfere with how a player learns and improves at the game.

Finally, Chapter 11 summarizes the work presented in the manuscripts and the ways in which I have contributed to our understanding of supporting skill development in games. Because there is still a lot we do not understand about skill development in games, Chapter 12 discusses work that could be done in the future to further increase our understanding of how to support skill development in games.

2 Background

In this dissertation, each manuscript contains its own related work section containing the background information required to understand the experiments presented. Despite this, there are several overarching concepts that are not fully described within the manuscripts — the concept of what a skill is, the process of learning a skill, and the way in which performance and learning of skills are measured. Even though the manuscripts do not explicitly explore or test these fundamental topics, a basic understanding of how and why people get better at skills is helpful for explaining the results of the manuscripts, and I will re-introduce these topics in the discussion.

This chapter opens with a section defining what "skill" means, followed by a section outlining some of the reasons that players are able to improve at carrying out skills. Section 2.3 introduces a concept that was very important for my experiments, that learning can be inferred from different types of tests of performance. Section 2.4 introduces various ways in which skill learning can be supported (guidance, directing attention, part-task practice, practice variety, and spacing practice), as well as factors that affect how learning occurs (such as feedback or where a learner's attention is directed) and some of these topics *are* tested and evaluated within the manuscripts. Finally, Section 2.5 provides an overview of how skill development can continue in the long-term.

2.1 What is a Skill?

"Skill" has been defined in a variety of different ways. For example, many researchers describe skills in terms of an activity or task whose purpose is to achieve a goal (e.g., [185, 98, 89]). Other definitions emphasize the idea that a skill is something where performance is desirable (e.g., [103, 235, 253, 302]), for example, Tomporowski [302] defines it as "the ability to use one's knowledge effectively and readily in execution of performance". For the purpose of this document, I define a skill as **a task carried out to accomplish a specific goal with efficiency and a high success rate**. Therefore, someone is "skilled" at a particular skill when they can carry out a task with a high degree of proficiency [183, 302] (e.g., professional gamers can be considered to be "skilled").

When applying this definition to games, we cannot simply say that, for example, playing *Super Mario Bros.* [219] is a skill. In reality, each game consists of many tasks that the player must carry out, and often several tasks are performed in concert to effectively play the game. For example, to be successful at playing a first-person shooter game, a player must be capable of carrying out the tasks of navigating their environment, locating opponents, targeting opponents, and completing any relevant objectives. Each of these tasks (skills) can be further broken down into their smallest components, known as "skill atoms". Skill atoms describe a feedback loop between the user and the system — a player performs an action that the game's systems processes and turns into on-screen feedback [58, 75, 137].

Furthermore, in this dissertation, I similarly do not consider a game as a whole to be a task. I consider a task to be an activity that a player engages with within a game, not a game or an environment as a whole. It may seem that task and skill could be used interchangeably, but a skill is more specifically a task where performance is a desirable goal. For example, avatar customization is a task within a game that is not a skill. However, avoiding enemies is a skill in a game that is also a task.

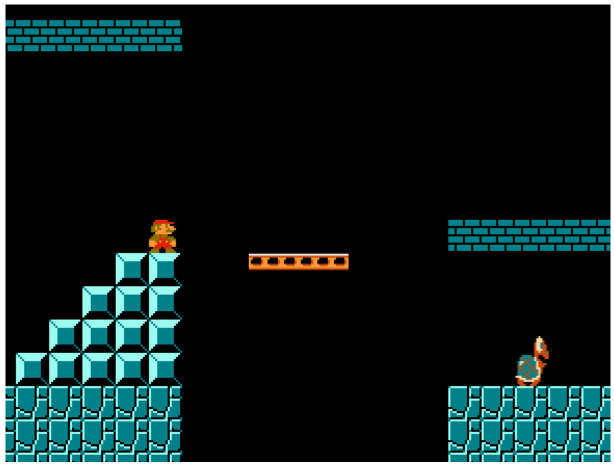


Figure 2.1: Mario must jump on the platform to get past the gap in the floor. From *Super Mario Bros.* [219] (all game screenshots in this dissertation were captured by me unless stated otherwise).

For example, in the side-scrolling platform game, *Super Mario Bros.* [219], the player is tasked with the goal of moving from left to right on the screen to reach the end of the level. As they do so, they encounter many obstacles along the way. In one scenario, the player may need to make Mario jump on a platform to clear a gap in the level (Figure 2.1). This requires the players to make use of the "jumping" skill atom (pressing a button makes Mario jump) to accomplish this goal.

Skills can be further classified in a number of different ways. The most high-level classification is whether the skill is primarily cognitive or motor. Cognitive skills involve the learner's intellectual capabilities and the focus is on choosing a correct response rather than carrying out that response [136]. Skills of this type can include problem-solving, memory skills, language skills, or social skills [98, 302, 184]. Motor skills involve the learner using their muscles and limbs to make a physical response and examples include throwing, catching, or running [136].

Despite the separate classifications, it is often insufficient to classify a task as being simply cognitive or motor — in reality, many tasks require components of both cognitive and motor skills to be completed successfully [302, 184].For example, in a Mario game, stomping on enemies requires a player to determine if the enemy is in range and safe to jump on (requiring cognition) and then requires a player to correctly time their button press (requiring motor movements). Generally speaking, the *perception* and *decision-making* components end up being important to the success of a motor response. As stimuli are picked up by various sense organs [89, 236], they must be identified and processed so that an appropriate response is selected and carried out [89, 340, 112, 236, 302]. The importance of *perception* and *cognition* in general is reflected in the common label for skills involving aspects of both cognitive and motor skills — they are referred to as being either *perceptual-motor* or *psychomotor*.

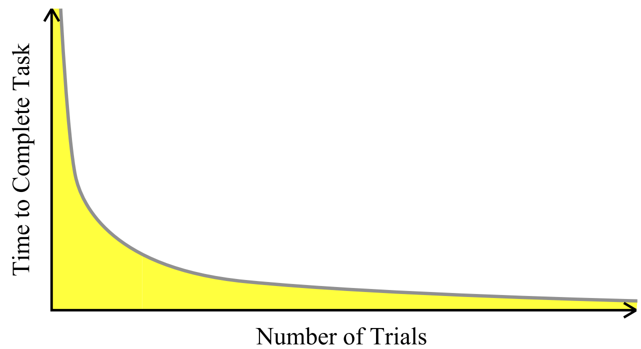


Figure 2.2: The power law of practice. Adapted from [338].

2.2 Why Players Improve: Perceptual-Motor Skill Learning

Humans have the capability to improve at skills due to *learning*. Learning involves the change of behaviour over time, as a result of experience [6]. For skills, learning occurs primarily due to *practice* [302]. With enough practice, eventually, one becomes an expert and is able to execute a skill with high performance [185]. This concept of becoming an expert over time can be found in the context of digital games, and we see players who can continue to make performance improvements over dozens, hundreds, or even thousands of hours [33, 173]. In fact, with some games it is possible to play a game as one's job and continue to see performance improvements [17, 84]. These changes in performance are described by the *power law of practice* [277, 214, 257, 302].

The *power law of practice* describes the predictable way in which a learner makes improvements when learning a new task. The rate of improvement follows the shape of a power law, which can be written mathematically as:

$$T_n = T_1 n^{-\alpha}$$

where T_n is the time required to complete a task on the n^{th} trial, T_1 is the time required to complete a task on the first trial, and α is a constant [44]. A representation of this power law is shown in Figure 2.2.

The greatest changes in performance take place early on in the learning of a new skill. As performance improves, the rate of improvement slows. Eventually, an effective asymptote is reached and performance improvements slow down and potentially stop entirely. When a learner reaches this asymptote, they are said to have reached a *skill ceiling* [256]. The cause for these ceiling effects is either due to physical constraints of a system (such as the cycle time of a cigar-making machine [61]), physiological constraints (it was once thought that a four-minute mile was the limit for runners [256]), or psychological constraints (some learners simply aren't motivated enough to put in the effort to make continual improvements [159]). Conversely, the floor refers to the minimum possible performance — if a task has a 30-second time limit, then the minimum performance that is possible while still completing the task is 30 seconds [256].

In practice, the power law works when the difficulty of the task is constant. Games often vary in difficulty over the duration of the game. This is discussed in Section 2.3.

2.2.1 Feedback

The reason that practice leads to improvements is due to the presence of *feedback* [89]. Feedback refers to the information that one receives about their attempts to carry out a task [253]. As one attempts to complete a task, feedback allows one to detect errors and evaluate their performance [272, 288, 256]. After trying different approaches and observing the resulting feedback [104, 302], one is able to refine or optimize their response by selecting the more effective approach [344, 248, 272, 166].

The feedback that occurs as a direct result of one's own actions and is picked up through one's sense organs is considered to be *inherent* (or "intrinsic") feedback [253, 89]. This type of feedback generally occurs *after* an action and so is also called "response-produced" feedback [253][1]. For example, one knows that they failed to jump over the gap in *Super Mario Bros.* when they see Mario plummeting to his death. In their next attempt, they might adjust their approach based on this feedback and get a running start or delay pressing the jump button until they are closer to the gap. As should be evident from this example, inherent feedback can be easy for a learner to evaluate — the error is often communicated very effectively [253]. However, there are situations where inherent feedback is not so obvious, and in these cases, the learner may be provided with augmented (or "extrinsic") feedback [253], which will be described more in Subsection 2.4.1.

Feedback in digital games takes many different forms, and because each game is a task designed by humans that takes place in a digital environment, all feedback found in games is technically augmented feedback. Feedback is often the result of an in-game reward or punishment; for example, the ubiquitous "game over" screen that is shown in arcade games is a type of punishment that indicates to the player that they have failed [158, 120]. Arcade games also provide feedback in the form of an in-game score or when a player loses one of their lives [4]. Aside from making progress in the game, other rewards include achievements, badges, leader boards, or the acquisition of virtual items or customization options and all serve as feedback that tells the player that they are doing well [75]. These rewards can motivate a player to continue playing during more difficult parts of a game [342].

Feedback is therefore also a source of motivation for players [158, 342, 226, 237]. This is particularly true if players feel that failure was a result of their own actions and that they can do better on future attempts [158]. Feedback also provides a player with information regarding how well they are doing, allowing them to evaluate whether their actions were effective [297]. This can satisfy one's need to feel a sense of competence if the player is overcoming the challenges found within the game [237]. Thus, feedback is not just important for learning, but for continued learning over the long term. If the game continues to increase in difficulty and provides players with challenges that are just outside a player's comfort zone, then the feedback of failing at the challenge will prompt the player to adjust their approach [95, 162]. Otherwise, if the challenge of the game is too easy, then the player may adopt a strategy or response that is acceptable for the difficulty of the task, but is sub-optimal compared to other approaches or when the task becomes more difficult [260, 95] — a phenomenon known as "satisficing" [271].

The next two subsections focus on how perceptual-motor skill learning occurs from two different perspectives. Sub-

[1] Sometimes one can detect that something has gone wrong in the middle of an action [253], but I will discuss that more when covering augmented feedback in Subsection 2.4.1.

section 2.2.2 focuses on learning and improvements to performance purely from the perspective of stimulus-response pairings while Subsection 2.2.3 focuses on learning and performance improvements from the perspective of what the learner is doing or focusing on throughout the process.

2.2.2 Stimulus-Response Learning

One of the simplest forms of learning is the forming of associations between a stimulus (S) and a response (R) as a result of experience [302, 89]. This sort of learning, also known as conditioning, does not occur unless feedback is present as it is through this feedback that a learner is able to gain the experience required for learning [135, 302]. This model of learning can be applied directly when considering the responses that someone has learned when perceiving various stimuli [340]. For example, stopping a car in response to a stop sign, jumping on top of an enemy when they get close in a platform game, or firing at an opponent when they enter a player's field of view in a first-person shooter game.

Operant conditioning (also known as instrumental learning) theories suggest that learning occurs through a trial-and-error process whereby a learner experiments with different responses to a stimulus and observes the results [340, 302]. Operant conditioning occurs over multiple trials or attempts, so it is possible to make use of a process called *shaping*, or "learning by successive approximations" [340]. Wingfield [340] gives an example of an infant learning to say the word "mommy." The infant will make random sounds until they incidentally make a sound starting with "ma," which the infant's parents will then reward. Eventually, the infant will say "mama", which is further rewarded, until finally, they say the correct word, "mommy."

Many games are expected to be played over and over again, and even those that aren't designed this way will tend to have mechanics that require a player to repeat part of the game [120, 158]. Because of this, trial-and-error approaches to learning the game are commonly employed by gamers, as the punishment of death discourages them from repeating the behaviour that led to the death. There also tends to be many rewards for a player eliciting the correct behaviour, such as defeating a boss, finishing a level, or levelling up in a role-playing game (RPG). These rewards and punishments encourage players to adopt responses to stimuli that allow them to make progress within the game.

The Information Processing Model

The three-stage information processing model describes the theoretical processes that occur within a person that takes place between presenting environmental stimuli and carrying out a response. It is thought that there are three stages of processing [236, 264, 302, 256]: *stimulus identification*, *response selection*, and *response programming*. The relation between these stages, the stimulus, and the response are shown in Figure 2.3. Under this model, improvements in reaction time (RT) are due to a person moving through these three stages more quickly.

The *stimulus identification* stage is where an environmental stimulus is detected and identified by a person. The stimulus is presented as light or sound (when considering digital games) and transformed into neurological impulses which are then processed by the brain [256]. The stimulus is thought to be processed until it arouses some associated memory and is then identified [256].

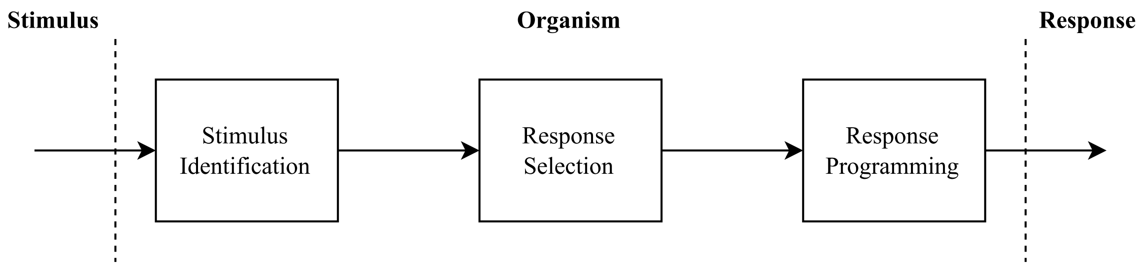


Figure 2.3: The three-stage information processing model.

The ability to identify a stimulus affects reaction time. For a visual stimulus, an increased *"stimulus clarity"* [256, p. 58] leads to a shorter RT. Similarly, the *"stimulus intensity"* [256, p. 58] considers the relative brightness or loudness of the stimulus and also affects processing time. Finally, the *"modality"* [256, p. 58] of the stimulus — whether or not is it visual or auditory — also affects processing time. It takes longer for someone to respond to visual stimuli when compared to auditory stimuli [347].

The *response selection* stage is where a person decides what to do in response to the identified stimulus [256]. To make this decision, a person must retrieve information stored in memory [302]. Multiple responses may be considered before making a decision [236], and that decision might even be to not respond to the stimulus [256].

The ability to select an appropriate response is affected by the number of stimulus-response alternatives, as well as stimulus-response compatibility. A greater number of stimulus-response alternatives is known to increase a person's reaction time in a relationship described by Hick's law [129] — where the reaction time increases approximately constantly every time the number of stimulus-response alternatives doubles. Stimulus-response compatibility refers to how "natural" the relationship is between a set of stimuli and the set of responses [256]. For example, given two lights and two buttons, it is more natural to press the button on the left if the left light turns on than it is to press the button on the right [256]. Another example is the *Stroop effect*, where increased reaction time is observed when participants have to name the colour of the text where that text is the name of another colour [286].

After a stimulus has been identified and a response has been selected, the body must then prepare for the movement that will achieve the desired response in the *response programming* stage [256, 302]. More complex responses increase total reaction time regardless of the stimuli [128]. *Complexity* in this context refers to the number of components to a skill, the amount of accuracy required, and its attentional demands [185, 256].

2.2.3 The Three Stages of Skill Learning

The stimulus-response model of learning as well as the information processing model describes what is happening *within* a learner that leads to them responding to stimuli more proficiently as they gain experience. The three stages of skill learning, on the other hand, describes the *journey* that a learner goes on as they first encounter a new skill, find an appropriate response, and refine that response over time to improve their performance. This model was originally described by Fitts [103] and has been expanded by Proctor, Reeve, and Weeks [236].

In the *cognitive* stage, a learner is trying to understand how to execute a skill that is new to them [104]. This is often a trial-and-error process where many mistakes are made. Expertise is gained when the learner observes feedback that allows them to adjust their execution of the skill [344, 248, 272, 166, 341, 302, 113]. In this stage, executing the skill is characterized in terms of *controlled interaction*, which is slow, attention-demanding, and intentional [56, 261]. Part of the reason for this is that the learner is considering their declarative knowledge (symbolic or verbalizable information) of the skill while attempting to carry it out [164, 104, 256, 302].

Performance improvements in the cognitive stage can occur rapidly [214, 257, 104, 103, 236, 235, 164]. These improvements are due in part to the formation of stimulus-response codings that allow the learner to move through the three stages of information processing more quickly [236]. They can also be attributed to the formation of *procedural knowledge* [164, 339] — a type of knowledge that can be applied directly in executing a task [10], is more robust to decay [302, 164], and allows for quicker response times and improved performance over *declarative knowledge* (i.e., factual information [163, 292]) [191, 27]. One final reason for performance improvements is that early on, novices tend to carry out skills in small steps. With experience, these individual steps are transformed into a single production that yields the same result [10].

In the *associative* stage of skill development, a learner has determined an effective way to complete the task and can now focus on making subtle adjustments to their response to improve their effectiveness or performance [302]. Their attention has transitioned from *what* actions they need to perform in response to stimuli, to *how* they are performing those particular actions [302, 256]. This stage is characterized by a dramatic reduction in errors, greatly improved performance, and gradual but continued improvements [214, 104, 302, 257].

There are several reasons for performance improvements in this stage. In terms of the learner's attention and what they are doing during this stage, they are determining the responses that work but are not effective, and eliminating them in favour of more effective alternatives [302]. In terms of information processing, the learner has a better understanding of the stimuli or cues important to the skill and is more able to differentiate between those that might simply be distractions and those that are important to their performance [302]. Additionally, the response-selection time and response programming times have decreased significantly — Proctor, Reeve, and Weeks describe this as the codings being developed to the point where there are now "direct stimulus-response associations" [236, p. 225] and the correct response is ready immediately instead of needing to be processed [333]. Finally, considering the transition from declarative to procedural knowledge, the learner no longer needs to use the declarative knowledge of the task [164] and can instead leverage procedural knowledge and "direct stimulus-response associations" [236].

In the *autonomous* stage, the learner performs the skill with coordination, smoothness, and accuracy [302]. It is characterized by few errors, stable expert performance, and little further improvement in the motor domain [256, 104, 214]. The learner can respond to stimuli with automaticity — that is, they can do so without conscious thought [104, 302, 236, 261] — and in some cases, this can be done in parallel with a variety of other tasks, or in the presence of distractions or other activities [261, 104]. While continued improvement at this stage is difficult, it can be accomplished by making use of deliberate practice [95] — a type of practice where the learner avoids acting with automaticity and focuses on trying alternative high-level strategies [302] to make further improvements or focus on specific aspects of skill execution that can be improved [95]. However, doing this can result in a temporary drop in performance as the learner directs their attention inward [27, 349, 354].

2.3 Evaluating if Players Improve: Measuring Performance and Learning

Skill learning cannot be observed directly, and so is inferred by examining the learner's performance [257, 302, 134, 278]. However, the learner's performance during or immediately after training is often an insufficient measure of learning because transient factors such as fatigue may be at play during the training session [257, 278]. Learning must therefore be evaluated by having the learner perform an additional test after training. If the training involved some type of manipulation (e.g., one of the support methods described in the next section), then this needs to be absent from the test [257, 278]. The test can be either a *transfer test* (where the learner completes a task different from the training task) or a *retention test* (where the learner completes the same task but after a delay) [257, 278]. It is important to note, however, that in the game domain, we are interested in both immediate improvement (i.e., as the player is repeatedly attempting the objective) as well as longer-term retention of skill.

In the game domain, another important consideration is that the difficulty of the task does not always remain constant. For example, many single-player games are designed in such a way that they increase in difficulty as a player makes progress [296, 342]. Other games make use of dynamic difficulty adjustment (DDA) — a system that is designed to dynamically manipulate the challenge of the game such that the player is facing challenges that are not too easy or too difficult [357]. And finally, for multiplayer games, the game's difficulty is determined by the skill level of the other player. These games often make use of matchmaking systems that match similarly skilled players together [226]. Therefore, in the presence of DDA or matchmaking, player performance might appear to remain consistent, even if the player is learning the game. The manuscripts presented in this dissertation do not make use of DDA or matchmaking, though Manuscripts C and D utilize games that increase in difficulty over time.

2.4 Helping Players Improve: Supporting Skill Learning

Sometimes players need additional help to learn the skills required to play a game. This can be provided in a variety of different ways. In this section, I'll introduce the concepts of guidance, augmented feedback, spaced practice, part-task practice, directing attention, and variety in practice. The manuscripts in this dissertation explore only *some* of these support methods. Manuscripts A and B look at the role of *visual and mechanical guidance* in skill learning, in the context of learning how to navigate a new virtual environment. Manuscripts C and D look at *spaced practice* in two different games. Manuscript D also looks at a common way of modifying in-game practice (checkpoints) that might introduce some part-task practice as well as change the variety of things practised during a play session.

2.4.1 Guidance and Augmented Feedback

Learning can occur without providing a learner with any explicit support, through a trial-and-error approach where a learner makes errors and leverages inherent feedback to observe the results until the correct response is acquired

[135, 272, 192, 234]. An alternative to this approach is to provide support to the learner in a way that they can carry out skills with a reduced or restricted number of errors [272, 256]. This type of learning has been called "guided" learning [274, 275], "errorless" learning [234, 273, 192, 139], or "error-free" learning [273, 272, 192, 155]. Therefore,

However, we know that inherent feedback is essential to the learning process [89], and the information provided by guidance may limit how much a learner attends to the information provided by any inherent feedback. There has therefore been some debate as to whether or not making mistakes is essential to learning psychomotor skills [135, 274, 272, 258, 192, 155, 256]. Several learning theories (such as operant conditioning [341] and experiential learning [166]) suggest that learning occurs when a learner finds a correct response after attempting a variety of different responses — by learning from their mistakes.

Like feedback, guidance is a way of providing *information* to a learner [215]. In this dissertation I use "guidance" to refer to any information or assistance given to the learner by some external source provided *before* performing an action or provided concurrently to action being executed [253]. In this dissertation, I use "feedback" to refer to any information provided concurrently to the task if it is inherent to the task (not from an external source), or provided after the task, regardless of whether it is provided by an external source. "Augmented feedback" in particular is when that information is provided after executing a task by an external source[2].

Guidance

Guidance can be provided in a variety of ways, for example, by verbally talking someone through the steps involved in completing a task, by pulling a learner's limbs through a movement, or by adding restrictions to an apparatus that prevent or constrain errors [253]. Therefore, guidance generally benefits performance significantly, although those benefits do not tend to persist into transfer and retention tests [253]. Guidance can also assist learners less directly, as it is known to have motivating effects for novices [302].

More generally, guidance can be presented visually, verbally, or mechanically. *Visual guidance* is guidance that is presented to the learner visually and the intent of this guidance is to help the learner develop a mental image of the task as well as how to complete it [136]. *Verbal guidance* is any type of information that is provided verbally, or is verbalizable (i.e., can be presented via text) [136], and is most commonly presented before task execution [216], and can also motivate learners [302]. *Mechanical guidance* is any type of guidance that introduces a mechanical restriction on the learner to minimize errors or force a particular response [216, 136].

Augmented Feedback

As defined earlier, feedback refers to information one receives about their attempts to carry out a task. *Augmented* feedback refers to feedback that is provided by an external source, such as a system or a coach [253, 185, 89]. This sort of feedback can be provided under circumstances where errors are difficult to detect through inherent feedback [253]. Augmented feedback can be categorized as knowledge of results (KR) or knowledge of performance (KP).

[2]Other sources use "augmented feedback" to also refer to concurrent but external feedback (e.g., [253]), however, I use the term "guidance" for any external information provided concurrently to an action. The reason for this is that manuscripts A and B explore support in the context of navigation, where "guidance" is a better-understood term than "concurrent feedback".

- *Knowledge of results* (KR) is externally provided feedback about the *outcome* of a response [249, 253, 185]. KR provides a learner with information about the outcome of a response in relation to a goal, and so is provided after an action [253]. Additionally, KR is *verbalizable* [249]. KR can be found in games, for example, at the end of a round of a first-person shooter game a player might be notified of their accuracy, indicating what percentage of their shots hit a target. At a minimum, KR tells a learner whether or not they succeeded at their goal [185].
- *Knowledge of performance* (KP) is externally provided feedback about the way the response led to a particular performance outcome [185]. This sort of feedback is intended to help a learner correct or improve their response [253] and so can be presented regardless of the outcome [139]. While KR is only presented verbally, KP can also be presented visually, for example by showing someone a recording of what they just did [185]. KP can be found in games, for example, a player might review a replay in the real-time strategy game, *Star-Craft 2* to spot opportunities to improve their response. Navigation guidance (like is found in Manuscripts A and B) could be considered to be concurrent KP, as it allows one to evaluate when they've made a wrong turn and provides more information about how their decisions made during navigation have led to getting closer to or further away from a target.

Feedback and guidance are highly related concepts because they both focus on providing a learner with information that will help them execute a skill successfully. In fact, based on my definitions I would actually consider KP to be guidance if it were provided concurrent to carrying out a task. Physical guidance could also be considered to be KP. For example, Howard [139] wrote that "concurrent continuous feedback, such as physical guidance, is considered KP since the subject receives ongoing information about the success of the movement" (p. 2).

The next three parts of this subsection describe the *timing* of the information. Particularly, they describe what types and sources of information one might encounter when presented before, during, or after an action. The work in this dissertation (in Manuscripts A and B) explores concurrent guidance, so that paragraph therefore includes more details than the others.

Information Before Action

Ideally, one has at least some knowledge of a task and its goal before an attempt is made to complete it [215]. Information about an action that is provided before an action is generally provided visually or verbally. **Visual guidance** that precedes an action takes the form of videos, charts, visual aids, or demonstrations [136, 216]. For example, one might have watched another person complete the task before attempting to complete it themselves. **Verbal guidance** that precedes an action comes in the form of verbal instructions provided by a coach, or some variety of written instructions [215].

Preceding visual guidance often takes the form of a demonstration. Studies where participants view expert or novice performers demonstrate a skill find similar benefits to both types of demonstrations over no demonstration [232, 196], although a combination of both expert and novice demonstrations is more effective than either on its own [243]. The benefits of these demonstrations seem to depend on whether the demonstrations reduce uncertainty on the part of the learner [216], and one study found that participants who were informed about what they were about to

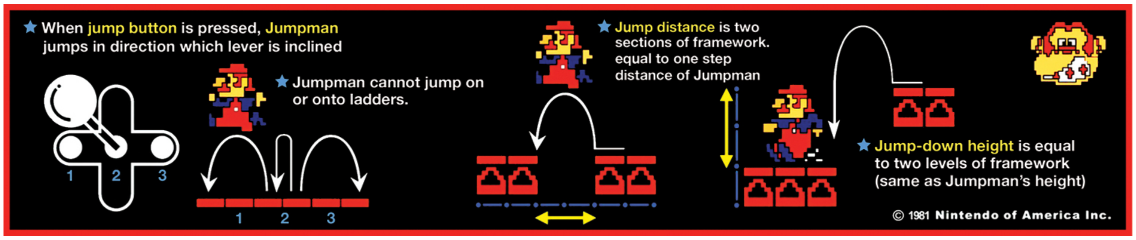


Figure 2.4: Instructions for the *Donkey Kong* (1981, Nintendo) arcade game. From [14].

Figure 2.5: The 1500-m event from Microsoft's *Olympic Decathlon* game (1982, Microsoft).

see *before* the actual demonstration performed better than if they were informed *after* the demonstration [12] — the learner's attention was therefore directed towards relevant aspects of the demonstration.

In digital games, preceding visual guidance is also typically presented as a demonstration. For example, the instructions for the *Donkey Kong* [223] arcade game, intended to be read before play, included visual aids to instruct the player on the rules for jumping within the game (Figure 2.4). The same arcade game also featured an "attract mode", which visually demonstrates what the game looks like as it is being played [306, 295]. Often, printed manuals or third-party help websites include visual aids that also serve as preceding visual guidance. Similarly, watching someone play the game (for example, on Twitch or YouTube) could also serve as a demonstration of the game which could act as preceding visual guidance.

For preceding visual guidance in the form of a demonstration, Pollock and Lee [232] tested the effect that watching another person play a game had on performance. They compared watching a novice play, watching an expert play, and not watching either. They used Microsoft's Olympic Decathlon [276] game, specifically the 1500-m dash event (Figure 2.5). This game involves steering an arrow to move counter-clockwise along a track, using the W, A, S, and Z keys to adjust the arrow's direction up, left, right, or down respectively. They found that the two groups that watched another player before playing themselves achieved a significantly higher level of performance compared to the group that did not watch another player; however, whether or not an expert or novice was watched made no difference.

Preceding verbal guidance generally precedes task execution and is most effective when used to call attention to

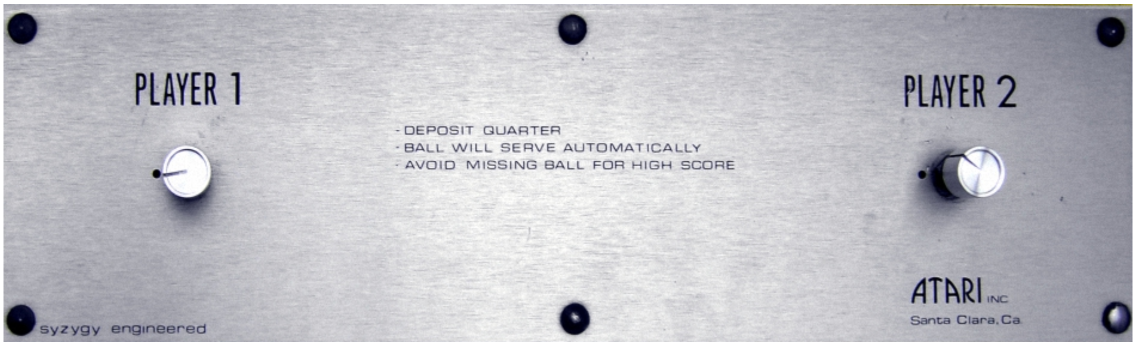


Figure 2.6: Instructions for the *Pong* (1972, Atari) arcade game. Image from [171].

perceptual information or cues [136], or direct a learner's attention towards specific aspects of a task or their actions [69]. The effectiveness of this approach appears to be affected by the amount of instruction. Learners must not be provided too much information or out-of-context information [352]. One suggestion is to provide the bare minimum of information needed at the start [256]. Providing learners with too much information is thought to prompt the learner to execute the skill in a way where they focus on using their declarative knowledge of the task and so hinders the development of procedural knowledge (the knowledge that must be developed to be able to transition to the associative stage of skill development) [164, 236]. Attempts to use declarative knowledge when executing a skill tend to mean that the learner performs the skill in a very controlled, conscious, and deliberate way — and this actually results in a less successful and reliable execution [27]. With many skills, a learner can learn to execute it implicitly — without being formally exposed to verbal instructions [267]. Too much instruction is thought to interfere with this process [191, 267, 192].

In digital games, preceding verbal guidance can be provided to the player either before attempting a task within the game or before attempting the game at all. In early arcade games, the number of instructions required to begin playing was often minimal, in line with the suggestion that games be "easy to learn, difficult to master" [334]. For example, the instructions for one of the first arcade games, *Pong* [18], were simply printed onto the cabinet and players were expected to read them before playing (Figure 2.6) [295].

Some early games for home computers were complex enough to require long instruction manuals. For example, an early computer role-playing game, *Temple of Apshai* [21] featured a 54-page instruction manual that contained important information describing sections of the game and how to play [295]. Today's digital games rarely include instruction manuals [70] and instead, instruction is integrated into the game itself. [3].

Information Alongside Action

Information about an action that is provided concurrent to or alongside that action can be provided visually, verbally, or mechanically. This sort of information often guides one toward doing something specific. For example, concurrent visual guidance can prompt a user to respond in a particular way (e.g., choosing to move in a particular direction due

[3] Although reading about a game is still common and instead of instruction manuals there are now community-edited wikis.

Figure 2.7: Visual guidance designed to help players identify audio cues. Image from [154].

to signs in an environment) or at a particular time (e.g., pressing a button in response to a light turning on [234]). Concurrent mechanical guidance can essentially force one to choose a particular response (e.g., a control stick moves on its own and one simply follows along [16]) or limit the error one can make (e.g., physical blocks restricting the amount that one can move a sleeve along a rod [135]). Concurrent guidance, whether provided visually, verbally, or mechanically can be considered knowledge of performance as it provides additional information that allows one to assess how their response led to a specific outcome in terms of performance.

Visual guidance that is applied concurrently can take many forms. For example, a light may illuminate to indicate that it's time for the user to press a button [234]. The effects of concurrent visual guidance can largely be summarized by stating that it significantly improves *performance* when it is present, but any gains are temporary and disappear when the guidance is removed [257, 16]. In fact, on retention and transfer tests, participants who receive concurrent visual guidance perform worse than or the same as participants who are never given guidance [16, 234, 274].

Incidentally, my Master's thesis [154] provided players with visual guidance to aid with the task of parsing audio cues to estimate the location of opponents. The guidance worked by displaying icons within the game's interface that corresponded to the sounds that players were hearing (Figure 2.7). The goal was to help new players parse the meaning of the sound effects, as those sound effects often gave away an opponent's position. This work found that players were better able to locate opponents through audio cues when visual guidance was present.

In games, one common way that visual guidance is included in games concurrently is to visually call attention to something in the game. This is done to increase discoverability, prompt the player to look in a particular direction, or provide a navigation aid [78]. For example, making an important item glow, colouring an item differently, placing lights in the environment to mark the desired path, or even placing a glowing particle trail on the ground that the player can follow (the effects of which are explored in Manuscripts A and B).

Because games are digital, **verbal guidance** can be provided by the system in a way that is concurrent to action or is close to concurrent by being provided just before it is needed. For example, in *Gauntlet* [19], every time a player encounters a new game element, a text box shows on-screen instructions and the text is spoken by a synthetic voice (see Figure 2.8).

A similar approach to this is hint systems that are integrated into the game [143, 250, 295, 331]. Hints can be

Figure 2.8: Instructions in *Gauntlet* (1985, Midway).

designed to be context-sensitive, provided on-demand, or introduced to fill idle time (such as on loading screens) [143, 8, 295, 307]. Context-sensitive hints consider the state of the game to provide relevant instructions. They are often displayed without the user requesting them, for example, when the player fails to make progress after a set amount of time, after a player fails a challenge, or whenever a game designer deems it appropriate (such as if a new challenge or game mechanic is being introduced) [307, 143, 295]. In games with problem-solving challenges such as puzzles, on-demand hints may be offered to players to help them find the solution to a puzzle [295].

The efficacy of just-in-time hint systems was evaluated by Andersen et al. [8], who implemented just-in-time hints (they called them "context sensitive" hints) by introducing concepts as closely as possible to when the player needed the information. This was done by displaying notifications or pop-up messages that could be dismissed. These hints were compared to context-insensitive instructions, which were implemented by providing multi-page manuals, organized by topic and presented before the player could interact with the game. Andersen et al. found that for the two "casual"[4] games they tested, there were no significant effects of context-sensitivity on their measures of time played or levels completed, which parallels the results they found when examining tutorial presence. For the complex game, *Foldit* [318], context-sensitive tutorials significantly improved the time played and levels completed.

Mechanical guidance is primarily provided during an action [216] and almost guarantees that performance will be high [343]. Like concurrent visual guidance, mechanical guidance significantly improves *performance* when it is present, but any gains are temporary and disappear when the guidance is removed, as evaluated by retention and transfer tests [257, 330, 135, 182, 16, 343]. Aside from this overall finding, there were also several secondary findings. In particular:

- Forcing a response always results in worse learning performance than restricting or constraining errors [135, 16, 139].

[4] In this context, "casual" implies that the game mechanics used within the games were commonly found within other games.

- Constraining errors in a “bandwidth” method so that the learner can make some amount of error results in better learning [16, 139].
- A transition from easy-to-difficult should be preferred over a transition from difficult-to-easy [182, 274, 343].
- Providing guidance on a fading schedule is better than providing guidance on every learning trial [274, 343].

It must be noted that the above findings are for tasks that are relatively simple (e.g., reproducing specific patterns [16], or manipulating specific inputs with one’s hands and feet [274]). Such tasks are not difficult to learn using the inherent response-produced feedback of the task and therefore the findings may not apply to more complex tasks [350]. For example, findings for a study involving a ski simulator found that providing mechanical guidance benefited performance and learning [351], and similar findings exist in the context of musical training [119] and gymnastics [127].

Information After Action

Information about an action that is provided after an action is completed is augmented feedback in the form of knowledge of results (KR). Despite the timing of this information being different, it largely affects performance and learning in a similar way to information provided concurrently. In particular, the “guidance hypothesis” refers to how “strong” amounts of KR feedback results in similar learning and performance outcomes to those outcomes found when studying concurrent mechanical and visual guidance (e.g., [249, 255]). This hypothesis states that strong KR feedback results in the learner becoming reliant on the aid: performance is significantly improved while it is present, but learning is hindered compared to not providing the KR feedback in the first place (e.g., [135, 343, 258]). This is similar to the results of concurrent guidance on learning and performance when considering simple perceptual-motor tasks (e.g., [16, 234, 274]).

The Guidance and Augmented Feedback Found in Games

The support methods described in the previous parts of this subsection (e.g., instructions, demonstrations, concurrent guidance, and KR) have been presented largely in isolation from one another. This is a consequence of how these methods have been studied, as they have often been presented as part of an experiment where it is desirable [102] to isolate the effects of each support method so they can be evaluated in isolation. In practice, in settings commonly found in many video games, support methods are combined in various ways. For example, written instructions on game mechanics are often accompanied by visual aids.

When Guidance and Augmented Feedback Are Not Necessary

It is not always necessary for games to include explicit support systems or instruction for new players. Andersen et al. [8] experimentally evaluated the impact that tutorials have when comparing three games of differing levels of complexity. One of their hypotheses looked at the presence of tutorials: “Games with tutorials will exhibit better player engagement and retention than games without tutorials" [8, p. 59]. Their results compared the inclusion of tutorials of any kind versus no tutorials of any kind. They found that for a complex puzzle game (*Foldit*, Figure 2.9) involving mechanics not found in any other game, players played significantly longer and played more levels when tutorials were present than when they were not present. For their other two more "casual" games (where "casual" implies that the mechanics are present in other common games), the inclusion of tutorials did not significantly affect time played or levels completed, and in one case the inclusion of tutorials even lead to a lower return rate.

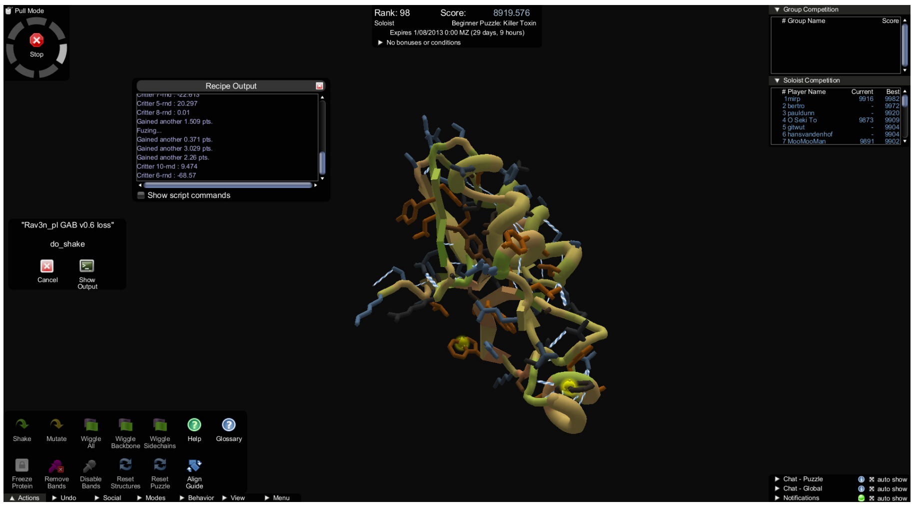


Figure 2.9: Screenshot of *Foldit* [318]. Image from [337].

Whether or not explicit support will be appreciated is also dependent on the individual player's prior experience [205]. For example, experienced problem solvers can make use of their past problem-solving experience to aid them in solving future problems [98] (that is, experienced players might not need as much assistance as new players). To investigate whether or not this idea also applies to problem-solving in games, Moirn et al. [205] conducted a study where players played a free-to-play match-3 role-playing game, *Saber's Edge*. The study used two between-subject factors — presence of a tutorial, and the expertise of the player ("casual" or "hardcore") — and used micro-transaction purchase intention and continued use intention as outcomes. They found support for the idea that "casual" gamers benefit from tutorials, both in terms of purchase intention and the intention of continued use. When considering "hardcore" gamers, they found no benefit of a tutorial when considering purchase intention or intention of continued use.

To summarize, both Andersen et al.'s [8] and Moirn et al.'s [205] work suggest that there are situations in which including explicit support would not be beneficial. Andersen et al. found that for simpler games where the rules or mechanics can be discovered with a small amount of trial and error, it is not necessary to include tutorials. Moirn et al. found that more experienced gamers do not benefit from the inclusion of a tutorial, at least when considering a casual free-to-play mobile game. In general, these findings suggest that providing players with additional information in the form of guidance or feedback will not always lead to beneficial effects in terms of performance and learning. This is similar to the suggestion by Wulf and Shea that support may be better suited to complex tasks where intrinsic feedback alone is not enough to learn how to complete the task successfully [350].

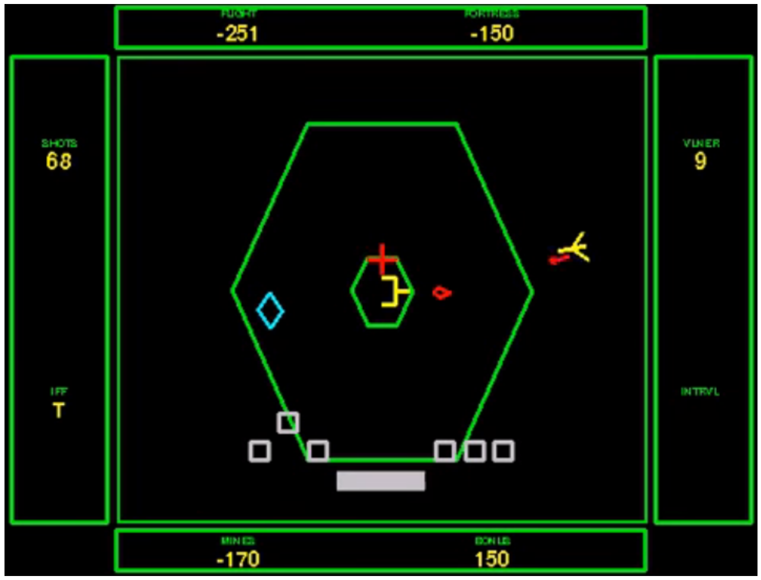


Figure 2.10: The *Space Fortress* game. Image from [57].

2.4.2 Spaced Practice

Spaced practice (also known as "distributed practice") means scheduling periods of rest to break up periods of work within a training session [256]. This approach has been shown to improve performance during training and in retention tests [174, 81, 256]. There is no fixed timing for the rest periods relative to work periods [321], and any amount of rest compared to a continuous-practice condition is considered spaced practice [256].

Meta-reviews of spaced practice studies have shown strong overall effects for the technique [174, 81]. For example, Lee and Genovese found a large mean weighted effect size of 0.96 for training (immediate performance), and a medium effect size of 0.53 for retention [174] — this indicates that spaced practice has a large effect on performance but less of an effect on learning. However, Verhoeven and Newell suggest that the meta-reviews do not necessarily provide unequivocal support for the idea that spacing practice enhances learning compared to continuous practice, as there are aspects of practice that moderate the effectiveness of spacing (e.g., differences in the task or the learner) [321]. Additionally, there is little agreement as to what length of rest optimizes the effect [321]; some suggest that "longer"[5] breaks are more effective than "shorter" breaks [256], and others suggest that performance follows an inverted U function (with medium-length breaks resulting in the best performance) [48, 81].

Manuscripts C and D make use of spaced practice, and each manuscript includes a related work section that provides further information about spaced practice.

2.4.3 Part-Task Practice

When learning a skill, one can choose to learn it in its entirety, or by focusing on its component parts via part-task practice. Part-task practice is generally beneficial when the skill is high in complexity and low in organization [211,

[5]There is also little agreement on what is considered a "long" break or a "short break".

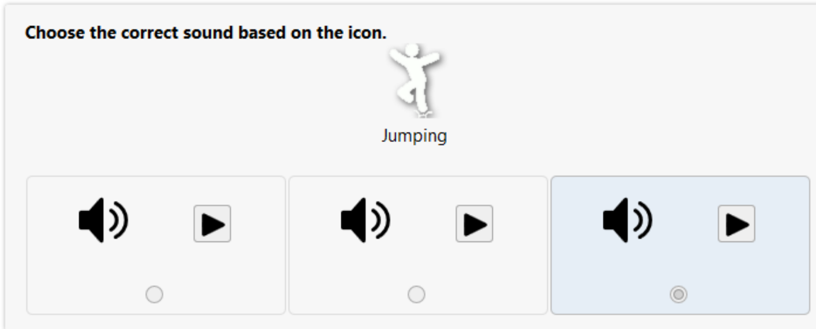


Figure 2.11: Training system to help players associate sounds with in-game actions. Image from [154].

105]. *Organization* refers to "the relationships among the component parts of a skill" [185, p. 406], so a skill being low in organization would mean that the components of a skill are relatively independent.

Therefore, for tasks that are complex and involve components that can be practiced independently, part-task practice is beneficial. For example, a different experiment using the game *Space Fortress* [79, 80] (Figure 2.10) decomposed the game into its component skills and developed specific training sub-tasks [107]. They found that it was not necessary to train component skills in context; their part-task approach led to increased performance over whole-task practice. However, participants who trained with part-task practice experienced initially worse performance upon transferring from their training tasks to the actual game. However, performance overall was higher than those who trained under whole-task practice by the end of training and in retention tests [107].

Incidentally, my Master's thesis work [154, 149] included part-task practice. In this work, participants were trained to recognize and identify sound effects from the game *Quake Live* [144]. Players trained this skill via a system that was inspired by the language learning service Duolingo[6] (see Figure 2.11). Participants were then tasked with estimating an in-game opponent's location by listening to and parsing the audio cues. This training significantly improved participants' accuracy in this task.

A common design pattern found in commercial games is to present a player with a fixed sequence of challenges to be completed one after another [335]. Therefore, a game designer knows what parts of the game the player has experienced and has yet to experience, and so can design challenges that are appropriate for the player at that point in the game. Well-designed single-player games optimize the first hour or so of the game to intrigue new players while instructing them on the skills required to play the game [54]. In the early parts of a game, players have fewer choices to make [342], the difficulty of the tasks is relatively low [342], and there is a reduced consequence of failure [297]. The intention is that the player should feel willing to experiment and observe the results of their experimentation, learning through trial-and-error [143]. As the player learns more about the game, the complexity increases and the tasks continue to be matched to the player's current skill level [297, 342]. If it works as intended, the skills players learn in this early part of the game continue to be used to solve later problems [114].

This approach is highly related to the concept of part-task practice because players are explicitly given opportunities

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

to learn skills in isolation and then practice those skills in increasingly more complex situations. Consider, for example, the highly successful single-player game, *Portal* [319]. A good portion of the game is explicitly presented to the player as a tutorial, with later concepts building upon earlier concepts, teaching the player how to make use of a novel game mechanic of placing and traversing through portals. The game also increases in complexity as players make progress; players start out only able to move through existing portals, then can place one portal within the environment themselves, and then finally can place two portals within the environment. Alongside this, the difficulty of the puzzles increases.

For another example, consider the first level of *Mega Man X* [43]. After a player has completed this first level, they have been introduced to all of the skills required to play subsequent levels [123]. This is because the player encounters different enemies or obstacles one at a time that each need to be bypassed or eliminated in different ways, in an environment with reduced risk [123]. This concept of instructing players through level design has been referred to by different names, including "conveyance" [160], "rational design" [197], or as a "fish tank" [114]. To accomplish this, designers design the game to call attention to a specific game mechanic or system by removing all other distractions [197] or showing the player the mechanic in a simplified setting where players can more easily see how something is accomplished [114]. The player, therefore, learns the mechanic in isolation and then is tasked with carrying out the mechanic in increasingly difficult settings, so that learning the game and playing the game become the same activity [160].

Introducing Part-Task Practice in Non-Linear Games

The above examples of part-task practice from *Portal* and *Mega Man X* work because those games are designed to be completed in a fixed sequence and the game designer, therefore, knows how much experience a player is likely to have (assuming the player has never played the game before). These days, more single-player games are often "open world" or non-linear and so players move through the game in an order unique to the individual player [335]. Multiplayer games are also often not linear in a way that allows for part-task practice. With these less linear experiences, a game designer does not necessarily know how much experience a player has at any particular point in the game, yet the performance expectations placed on the player are the same — regardless of how much experience the player has they are expected to be able to carry out the tasks presented to them.

To get around this issue in both open-world and multiplayer games, game designers often introduce a linear section of the game to instruct new players how to play. For example, in *Zelda: Breath of the Wild* [222], the early part of the game has players starting out on a large plateau, and players must complete a series of quests before they are able to leave and explore the rest of the world. These quests introduce important concepts and game mechanics that the player will need to make use of later on in the game, instructing the player on concepts such as stealth, combat, using items, and abilities that are used to solve later in-game puzzles [356]. In multiplayer games, a common solution to this is to ask the player to complete a tutorial before participating in online play. In *Age of Empires IV* [239], for example, players are placed into an approximately twenty-minute tutorial before even seeing the main menu of the game. After this tutorial, players are then encouraged to play through several "art of war" missions, in which they can learn further skills that will help them be successful at playing the game.

2.4.4 Variety in Practice

For many skills, the ability to cope with various or novel situations is important to success. The ability to do this successfully can be improved by making use of practice that contains increased variety or variability [254] when compared to tasks that remain constant, as evaluated by transfer tests involving novel variations of a task (e.g., [195, 270]). In addition to this, benefits to increased variety in practice are also found in retention tests, where the task tested is one that both the variable and constant groups had trained on (e.g., [265, 266]). However, these benefits are subject to factors such as the nature of the task (e.g., benefits of variable practice are found most commonly with simple tasks) and the expertise of the learner (e.g., increased variability may be most beneficial for early learning) [350].

There is one drawback to this approach to this type of practice: it results in initially worse performance as the learner is training with increased variety, due to increased errors [185]. While this is a temporary problem, it could be problematic in digital games, as a player's enjoyment of and progress in a game is tied to their current performance.

2.5 How Players Improve in the Long-Term

For as long as people have been playing games, people have been competing and striving to reach new high scores or prove they are better than the competition. For example, in the 1980s, players competed in Video Game Masters Tournaments to set new world records [336]. Players can continue to make performance improvements over dozens or hundreds of hours [33, 173], and with some games, it is possible for a player to play a game as if it were a full-time job and continue to see performance improvements over time [17, 84] (as many esports professionals do).

Performance improvements over time are generally described by the power law of practice (Section 2.2), where performance improvements gradually reduce over time until an effective asymptote (or skill ceiling) is reached [277, 214, 257, 302]. However, practice alone is often not enough: due to phenomenons such as satisficing [271], a player's performance may level off despite there still being potential performance gains possible. For a non-game example, consider an office worker who never learns how to touch type (typing without looking at the keyboard) [71, 23] despite many hours of practice each day. Essentially what happens is that the learner has adopted an inferior strategy for completing the task, or their learned response is flawed in some way. Further practice of this flawed technique may improve performance slightly, but it generally only serves to reinforce the flawed technique [95]. Practice in this situation can be considered to be "naive practice", which involves one "doing something repeatedly and expecting that the repetition alone will improve one's performance" [95, p. 14].

In games, ineffective strategies are eventually eliminated because the difficulty of the game will tend to increase over time [296, 342]. If the game's challenge never increases, then players have no reason to adopt better strategies or attempt to improve their existing strategy. If the challenge is too difficult, then players are likely to feel overwhelmed. A challenge that is well-matched to a player's ability provides a suitable setting for learning [332], and if that challenge is at the outer edge of a player's current ability then it may encourage them to improve at the game.

2.5.1 Adjusting a Game's Challenge

As described earlier (Section 2.4.3), linear single-player experiences can generally be designed such that players are only encountering challenges well-matched to their ability. In multiplayer games, the challenge is determined primarily by the other players, but the balance can be manipulated by providing *dynamic difficulty adjustment* or *skill assistance*.

Dynamic difficulty adjustment involves modifying the game's parameters to give a struggling player some advantage [226, 353]. This approach has been found to increase engagement [353] and so is effective when it can be applied. Aids for players based on in-game performance have been utilized since at least 1999 (in *Unreal Tournament*) [226]. There is unfortunately no study that looks at the effects of dynamic difficulty on *learning*, so it is not clear whether it would aid players in refining and improving their skills. However, it would likely keep players engaged with the game and playing under conditions in which they are more likely to be learning.

Skill assistance can be considered mechanical guidance (described in Section 2.4.1) because it directly helps the player with a skill they haven't yet mastered, while difficulty adjustment cannot because it simply makes the challenge that the player must face easier to overcome. As a concrete example, consider the skill assistance provided by the racing series, *Forza* (e.g., [305]). The player has the option of enabling a variety of driving assistance. They can enable automatic braking and steering — two examples of mechanical guidance — as well as an ideal driving line showing them the optimal path to take — an example of visual guidance. Aim assistance is an example of mechanical guidance that is provided in countless first-person shooter games, whether it is to mitigate the limitations of analog sticks on game controllers [111, 322, 210] or to provide a balanced challenge for players of disparate skill levels [322].

Because the goal of assistance is to only provide a short-term improvement to performance, there has not been much work on the effect it has on long-term learning. Gutwin et al. [122] looked at the effect of aim assistance over a five-day study that had players walk through a short single-player level with or without aim assistance. They evaluated learning by having participants complete a shooting range without assistance at the start of the study — before the walk-throughs — as well as at the end of the study. Overall, they found no significant changes in either performance gains or experiential measures, comparing between assisted and unassisted groups. These results could be due to a few reasons. First, the assistance technique did not aim for the player. Players still had to perform the motor task of aiming as they normally would. The assistance used — bullet magnetism — simply enhanced the participant's aim by altering the path of bullets in flight. Second, the assistance may have allowed for a type of easy-to-difficult transition to occur, which can have positive effects on learning [182, 274, 343]. Alternatively, it may have allowed the player to direct their attention towards improving on tasks other than aiming. Unfortunately, these results focused on novices and not on experts, so it is unknown whether such an approach would be of any benefit to players wanting to improve their skills over the long term.

Aside from dynamic difficulty and skill assistance, one further way that game balance can be adapted in multiplayer games is through *matchmaking*. Matchmaking has become the accepted approach to this problem and has been applied to online games since at least 2001 (in *Internet Backgammon* [202]). Matchmaking works by giving each player a skill rating [226]. When a player indicates that they want to play, they must wait until a similarly rated player also indicates they want to play. However, the longer a player waits, the less strict this matching becomes, and the matched players may be of wildly difficult skills if few players are online [233]. Further, the system has difficulty matching new players who do not have a rating and so these players' first matches are likely to be unbalanced and therefore frustrating [233]. Presumably, if a player matches with other players who are at similar skill levels, the player will be in a better learning environment (due to the difficulty match [332]) in which they are motivated to improve the execution of their skills; however, this has not been demonstrated in any study so far. Another approach that is used is to have new players queue up to face off against AI-controlled opponents (e.g., in *League of Legends* [241]). This can give players a taste of the game in an environment where the difficulty could be manipulated via an algorithm tuned by the game developers. This is also a potentially beneficial learning environment and a good way to initially gauge a new player's skill level.

2.5.2 Deliberate Practice

Game designers can encourage players to improve by increasing the challenge in a way that matches a player's learning. A player can also make an intentional decision to improve by making use of *deliberate practice*.

Naive practice — the idea that simply repeating a task and expecting one's performance to improve — works for a time but eventually, performance improvements cease [95]. Therefore, an alternative to this is deliberate practice [94, 96, 95]; a type of practice that is designed specifically to increase a learner's performance. Even in scenarios where naive practice still results in performance improvements, Ericsson and Pool [95] suggest that deliberate practice is considerably more effective in maintaining continued long-term improvements, and can allow one to reach a level of performance far beyond what might be reached with naive performance. They also make a point of stating that they have "found no limitations to the improvements that can be made with particular types of practice... people in every area of human endeavor are constantly finding ways to get better, to raise the bar on what was thought to be possible, and there is no sign that this will stop" [95, p. 113].

Deliberate practice is based on the idea of intentionally avoiding acting with automaticity [94]. This automaticity is what characterizes the autonomous phase of skill development and the concept that a learner is able to execute the skill with little to no conscious thought is what allows one to achieve a high level of performance [104]. Ericsson and Pool [95] suggest, however, that this is a problem when the learned automatic response is flawed or in some way ineffective, and therefore any practice of the skill that utilizes this learned response serves only to reinforce those flaws, making the learner better at executing the flawed technique or executing that response more automatically[7].

Ericsson and Pool [95] describe several conditions that must be present for deliberate practice to take place:

- There must be well-defined, specific goals.
- There must be feedback present.
- The task must be just outside of the learner's comfort zone.
- The task must be given the learner's complete attention.

[7] These automatic responses are still appropriate in situations where high performance is the primary goal, just not when the goal is *improvement* [94]

Deliberate practice is not something explicitly designed into games, but it is an approach that many players adopt in order to improve. For example, professional *StarCraft* [30] players use deliberate practice to refine how they execute their actions during the early part of each match to ensure that they start producing their army as fast as possible [285]. They will also review recordings of previously played matches specifically to learn from their mistakes and look for ways to improve [285, 141]. This deliberate practice is also something done by video game speedrunners — players who play through the same game many times with the aim of completing it as fast as possible [308].

Part II

The Manuscripts

3 Introduction to Manuscript A

As described in Section 2.1, every digital game contains a variety of tasks that a player must master. One task that is common to many games is the skill of *navigation* — to be successful at a game, one must be able to memorize and navigate new game environments to discover the location of in-game objectives or resources. A variety of aids exist that players can make use of when navigating a virtual environment, for example, directional signs within the environment, maps, paths within the environment, or guided tours from a non-player character (NPC), to list a few [208]. Of the skills found within games, navigating an environment is slightly unique in that it is also a skill found in one's everyday life. Yet, despite most people being capable of navigating the physical environments they live in, this is something that many struggle with once the environment becomes virtual [68, 67].

In Subsection 2.4.1 I introduced the idea of supporting learning by providing one with additional information, either before, during, or after an action. The goal of this information is to reduce the number of errors that one makes as they are completing a task. The task of navigating a virtual environment is one where a lot of well-established support methods already exist. And given that these support methods are so common [188, 208], it is likely that they are effective at supporting navigation. In the terminology of Section 2.3, these aids likely have strong effects on *performance*, but how might they affect *learning*? That is, how does making use of these aids affect how well one can navigate an environment once they are taken away?

We chose to try three different navigation aids: A map showing the routes through the environment, a map that additionally showed the player's current location, and augmenting the environment by adding trails that show where to go. All three of these can be classified as concurrent visual guidance, with the intended difference between them being the amount of effort that must be put in to make use of them. Research suggests that when more effort is put into learning something, a learner should be able to better remember that information later on [238]. Furthermore, Subsection 2.4.1 describes how information provided alongside an action can often become relied upon — it can often lead to improved performance when it is present but tends not to lead to any lasting improvements to performance if it is taken away. Based on these ideas, we hypothesized that we would find that the assistance that led to the least effort during navigation (i.e., the trails) should result in worse learning compared to the level of assistance with the most effort (i.e., the map).

The experiments in this manuscript provided an expected result. In terms of performance, our findings matched the hypothesis: the "stronger" forms of visual guidance — map with the player's position and trails — resulted in improvements in navigation performance compared to just having a map. However, when considering learning, we found no statistically significant differences between the three levels of concurrent visual guidance. This is a surprising result for two reasons. First, it contradicts the findings of past work that evaluated the efficacy of concurrent visual guidance for learning (as summarized by Subsection 2.4.1), which suggests that people can learn a task more effectively without concurrent visual guidance (at the cost of reduced performance while learning). Second, the group that only received the map spent considerably more time within the virtual environment during training, and it is surprising that this increased exposure to the environment did not lead to any more spatial knowledge about that environment.

Considering the theories presented in Chapter 2, part of the reason why concurrent guidance can interfere with learning is that it could cause a learner to disregard the inherent feedback of a task. Inherent feedback is how a learner can learn a task through a trial-and-error approach — by attempting the task and observing the result [135, 272, 192, 234]. But what is the result of turning left or right or choosing not to turn, and how might someone navigate to evaluate whether they have made a wrong decision? It might be that navigation is simply a task where inherent feedback occurs less often, so providing additional information does not interfere with learning as much as it might with other tasks. Another potential explanation for these results, as discussed within the manuscript, is that spatial learning occurred incidentally. Therefore, because our participants experienced the route in its entirety (i.e., they were not teleported to their destination) they were able to learn the route through passive observation.

3.1 Methodological Clarifications

3.1.1 Sample Size and Effect Sizes

Participant counts were determined by resource constraints rather than an a-priori power analysis [170]. Therefore, here I will report the effect sizes we found for the training and testing completion times for Studies 1 and 2.

In Study 1, I found a strong effect size of training with assistance (1.06), but a small effect size of assistance during testing (0.17). In Study 2, I found a strong effect size of training with assistance (1.21) and a small effect size of assistance during testing (0.16).

3.1.2 Covariate Selection

In Studies 1 and 2, I collected information regarding the individual differences of the participants to be used as covariates. Due to the number of participants in the studies, including every variable as a covariate would have resulted in a loss of statistical power [170], so I selected covariates based on whether they correlated with our outcome variables. All measures of individual differences and outcome measures were put into a Pearson correlation matrix. A measure of individual differences was chosen to be a covariate if it was significantly correlated with the outcome measure of the test.

3.1.3 Participant Payment

In Study 1, we paid participants $6 USD and the study took on average 42.7 minutes to complete (SD=15.6). In Study 2, we paid participants $3.50 USD for day 1, which took on average 21.7 minutes to complete (SD=8.1), $3.00 USD for day 2, which took an average of 14.8 minutes to complete (SD=10.0), and $4.50 for day 3, which took an average of 30.5 minutes to complete (SD=9.5).

This works out to $8.43 USD per hour for Study 1 and $9.85 USD per hour for Study 2, both of which are more than the United States' federal minimum wage of $7.25 an hour, where our participants were from.

3.2 Additional Analyses, Results, and Figures

When writing Manuscript A, there were some results and figures that did not make it into the final paper due to space constraints. These have been included in the appendix, in Section E.1. This includes:

- A figure showing the map used for the Furious Heights environment from *Quake Live*.
- Reporting the statistical effects of the covariates used.
- Reporting on possible gender differences.
- Reporting on the debrief responses.

3.3 Publication and My Contributions

This manuscript was published as [153]:

> **Johanson, C.**, Gutwin, C., & Mandryk, R. L. (2017). The Effects of Navigation Assistance on Spatial Learning and Performance in a 3D Game. *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 341-353. https://doi.org/10.1145/3116595.3116602

This work involved contributions from my supervisors, my contribution included: determining the overall design for the experiment and guidance conditions, implementing the digital system, performing data analyses, and writing the majority of the manuscript. I also presented this work at the conference at which it was published.

4 Manuscript A

The Effects of Navigation Assistance on Spatial Learning and Performance in a 3D Game

4.1 Abstract

Navigation in 3D game environments is often difficult for novices, who may get lost and be unable to reach game objectives. Many games provide navigation assistance (e.g., mini-maps, directional markers, or glowing trails); however, there is a risk that players will become reliant on an aid and fail to develop a mental model of the map. To investigate, we carried out two online studies in which people carried out training tasks with varying navigation assistance. After training, they navigated the map with assistance turned off. In both studies, we found that assistance improved training performance, but found no harmful effect of assistance on performance after it was removed, even when comparing between those who received glowing trails to follow and those who received no assistance. We show that navigation assistance in 3D games is effective, and that it does not necessarily harm development of a novice's spatial learning.

4.2 Introduction

There are many skill differences between novices and experts in 3D games, including targeting, evading enemy attacks, understanding weapon capabilities, and the ability to memorize and navigate the game map [154, 322]. Map knowledge — knowing one's own location as well as knowing game locations and routes — can dramatically affect play experience: whereas experts quickly learn where resources and enemies are located [289], novices find themselves continually lost and unable to get to objectives.

Spatial abilities in games are of particular interest because unlike many game skills, learning and navigating 3D environments is something that people do in the real world throughout their lives. Even though there are individual differences in spatial ability, most people are fully able to function in the environments around them, and can easily find their way through houses, office buildings, or shopping centres. This does not appear to be the case in 3D games: Several researchers have noted the difficulty of navigating virtual environments [68, 67], and there are many examples of novice players discussing difficulties with navigation, or videos making fun of their poor navigational skills. For example, a player of Final Fantasy Online posted in a forum:

> **Noob, Lost, Cannot find my way around the map to the point of AHHHHHHH!!!** I am super frustrated. I am a noob to the game. I am somewhere on some steps... I can get to all the levels and outside

> but I cannot find the Drowning Wench. [...] Once you stop laughing at how much of a moron I am for not being able to find my way out of a paper bag could someone please give me some tips or something. (forum.square-enix.com/ffxiv/threads/90497)

Problems with game navigation — that is, getting lost or not knowing how to get to an objective — can be extremely frustrating, and may contribute to a novice player's decision to quit the game. To reduce navigational difficulty, some games add directional assistance to the game environment, such as markers telling the player where to go (Figure 4.1), or routes drawn on the game map (Figure 4.2). However, these assists have one significant limitation — they only work when the desired destination is well known. There are many situations in which a player needs to adapt to changing conditions in the game and stray from the well-marked path: for example, to find an alternate route that flanks an enemy, or to look for treasure in the environment. In fact, wayfinding itself can be a compelling challenge for many players [229], and some even feel that the joy of discovery is stolen from them when an assist led them to that discovery [311].

Although providing novices with assistance can improve early performance and experience, designers may be setting players up to fail later on. There is a concern that a player will become reliant on an aid, and when they are eventually required to navigate on their own (e.g., to play competitively), they will be unable to do so. This phenomenon of players relying on external feedback is known as the *guidance hypothesis* [249, 258], which suggests that greater effort during training (i.e., *intentional* learning) will lead to better retention and understanding [88].

In contrast to the guidance hypothesis, however, is research suggesting that spatial knowledge in 3D environments can be gained through *incidental* learning [11, 320, 126]. Incidental learning occurs simply through exposure to an environment — for example, people may learn the layout of a building even if they are being led by a guide. This natural ability to learn about an environment may arise because an understanding of our surroundings was critical for the survival of early humans. There is debate about whether location learning can occur as an incidental process, however, and there are studies that point to problems in spatial learning caused by navigation aids such as GPS (e.g., [42, 146, 176]).

These competing theories about location learning mean that it is difficult to predict the effects of navigational assistance techniques in games. To determine both the benefits and risks of navigation assistance in 3D games, we carried out two studies in which novices completed several route-finding tasks with different forms of navigation aid, and then completed test routes without any assistance. The training tasks used one of three assistance levels: *no assistance*, *moderate assistance* (a mini-map and pop-up map with the player's current location), or *strong assistance* (a path drawn on the ground of the game world itself). The testing routes required novices to navigate un-aided to a particular landmark. Both studies used the same testing and training routes, and the same two real game environments: the "Gold Rush" map from *Enemy Territory* [280], and the "Furious Heights" map from *Quake Live* [144].

In the first study, participants completed sixteen training routes, then eight test routes, all in the same session. In the second study, participants completed sixteen routes in each of three sessions on three successive days, as well as eight test routes on the third day. In addition to the in-game navigation test, we asked participants questions to assess their knowledge of landmarks, routes, and distances on the two game maps; and we also asked them to rate their play experience after training and after testing.

- Navigation assistance helped substantially when it was present: in both studies, having more assistance allowed participants to complete significantly more routes, and to complete them in significantly less time.
- There were no performance differences in testing: despite the strong effects in training, there were no significant differences in performance on the test tasks — including the number of test routes completed, the distance travelled, or the completion time — regardless of whether or not people had assistance during training.
- Subjective responses mirrored performance results: the presence of navigation assistance significantly reduced perception of effort, frustration, and mental demand. When the assistance was taken away in testing, these differences disappeared.

Our findings confirm that navigation assists substantially help novices, and adds to evidence [122] that early guidance does not necessarily hinder performance when taken away (even with strong assistance such as glowing trails). Our results can be used by game designers in several ways: first, designers should be aware of the difficulties that novices have in learning 3D game environments; second, designers can substantially improve novices' initial play experience with navigation assistance; third, navigation assistance can be used as a player-balancing mechanism for social-play situations; and fourth, taking navigation assistance away will not necessarily result in the player becoming lost.

4.3 Related Work

4.3.1 Navigation in Real and Virtual Environments

A wide variety of research has been carried out to investigate the ways that humans learn and perform navigation in real-world environments — for example, researchers have looked at the development of spatial knowledge in children (e.g., [124]), sex differences in navigation (e.g., [53, 172]), and theoretical models for navigation (e.g., [53]). One major focus in navigation research is on *wayfinding*, the process by which people orient themselves to an environment and move from place to place. Early work identified three kinds of knowledge that are important for wayfinding, and that are associated with increasing spatial understanding [301, 300, 299]:

- *Landmark knowledge* involves remembering specific objects or settings in an environment — such as a statue or a building in a city centre.
- *Route knowledge* is understanding how to navigate between specific locations, and the actions required to reproduce a specific path between them. Route knowledge often builds on landmark knowledge (e.g., by linking different landmarks together).
- *Survey knowledge* is a map-like mental representation of an environment, and is the highest form of spatial understanding. Survey knowledge allows people to navigate skillfully, estimate relative distances, and choose alternate routes to objectives.

There are two ways in which people can gain this spatial understanding of an environment [68]. First, people learn through direct exposure to their surroundings — that is, simply being in an environment and moving through it. Second, external information sources such as maps provide other forms of spatial learning. When used in an actual navigation task, maps require that users identify their own location on the map, and then translate orientations, directions, and distances from the map representation to the actual environment.

Researchers have also studied a variety of navigational tools and aids in real-world wayfinding. The most common tool is the map, and researchers have looked at several aspects of map use, such as the differences between "track-up" and "north-up" orientations [15]. Recent research has also looked at the effects of guidance systems such as GPS, and has found that people can become overly focused on the directions provided by external guidance, hindering the development of their spatial knowledge (e.g., [42, 146, 176]).

Navigation in virtual environments has also been extensively studied. One main interest is in whether virtual environments can be used as training simulations for real-world navigation [328], and whether spatial knowledge and wayfinding ability transfer to real environments. Researchers have also identified that navigational difficulties are common in virtual environments (e.g., [68, 88]): "Virtual world navigators may wander aimlessly when attempting to find a place for the first time. They may then have difficulty relocating places recently visited. They are often unable to grasp the overall topological structure of the space" ([68], p. 166).

To combat these difficulties, previous work has also looked at a variety of navigational aids. The value of landmarks has led researchers to consider the idea of allowing users to place visual markers, having the system create a visual trail showing where users have been, or having a fixed marker to provide a consistent indication of north [68, 67]. Results with these forms of assistance are mixed, however: adding a simple compass did not substantially improve navigation performance [65], and trails can quickly clutter an environment. To our knowledge, no studies have looked at the effects of navigation assistance on spatial performance once the assistance has been removed.

4.3.2 Incidental vs. Intentional Spatial Learning

A continuing debate concerns the relationship between spatial knowledge acquisition and intentionality. Studies indicate that at least some aspects of location learning occur automatically [11, 126]. For example, one study showed that recall of word locations was unaffected by the difficulty of a concurrent task [11]. Other work, however, shows the importance of intention; studies have shown that when people focused their attention on a route through a building, they were better able to draw a map of that path [320], and that even long experience with an environment may still result in poor survey knowledge [50].

Similarly, Ehret [88] suggests that the amount of attention, repetition, and practice during training will affect the degree to which an object's location can later be retrieved from memory. Ehret suggests that because explicitly remembering locations requires effort, people will choose a lower-cost strategy when possible, impairing their learning. This phenomenon is related to the guidance hypothesis and effort-retrieval hypothesis, described further below.

Figure 4.1: The quest marker on Skyrim's compass display, showing the direction to the next objective.

4.3.3 Navigational Assistance in Games

From subtle signs or arrows, to obvious glowing trails, many games feature navigational aids or assistance. Two common navigational assists are compasses and mini-maps [229]. Both are used to display locational information to the player. An example is the quest marker [229, 348] — icons that indicate the start, end, or intermediate goal of a quest. For example, *Skyrim* [29] (Figure 4.1) includes a compass at the top of the screen with markers for selected quests and icons for points of interest. Mini-maps are used similarly: *Counter-Strike: Global Offensive*'s [130] mini-map shows nearby teammates and the location of objectives.

A stronger aid that is found in games is a highlighted trail in the environment. The effect is often implemented as a smoke or particle trail along the recommended path. For example, *Fable II* [179, 293] (Figure 4.2) and *Neverwinter* [62] have particle trails which can be turned on or off through the user interface. In other cases, the trail may be implemented diegetically: in *Skyrim*, magic users have access to a "clairvoyance" [29] spell that temporarily highlights the route to a quest marker with a smoke trail.

4.3.4 Skill Development in Games

Games have gained a reputation as powerful learning tools; they are able to transform novices into experts through an engaging, motivating, and enjoyable experience [114, 167]. Many theories apply to learning in games. We describe Kiili's experiential gaming model, the guidance hypothesis, and the retrieval effort hypothesis.

Kiili's experiential gaming model [162] is based on *experiential learning theory* [166], flow theory [52, 63], and the zone of proximal development [325]. In experiential learning theory, a person forms a prediction based on their prior experiences and then tests those predictions on new experiences. Games provide many opportunities for players to test their ideas and predictions [114]. *Flow theory* describes a state in which one becomes so immersed in an experience that they notice nothing outside of that experience. Well-designed games are able to keep players in the flow state for long periods of time by providing challenges that are well-matched to the player's ability [114, 162, 167]. The *zone*

Figure 4.2: Glowing "breadcrumb trail" in *Fable II*, showing the path to the start of the quest (or the next objective).

of proximal development describes what a learner can accomplish if they are given some guidance. Often, the systems within a game can provide that guidance, scaffolding a player to overcome a challenge just outside their current skill level [167].

Kiili incorporates ideas from these three theories to form his experiential gaming model [162]. As it has been shown that being in the flow state benefits learning [332], Kiili proposes that the flow state can be extended with the zone of proximal development — if a player's ability is scaffolded so that they can complete challenges just outside of their ability, maximum learning will occur. Kiili's model is based on continual re-learning, and includes two phases. In the *ideation* phase, the learner generates ideas and potential solutions while in the *experience* phase, the learner tests their ideas while attempting to overcome in-game challenges. After testing, the learner can then incorporate their new experiences and generate new ideas.

The *guidance hypothesis* refers to the phenomenon where a learner starts to rely on extrinsic feedback provided by a system. The type of feedback which leads to this phenomenon is Knowledge of Results (KR) [249]. KR refers to extrinsic feedback indicating task success in response to a learner's actions [258]. A few researchers have studied whether assistance systems in games result in reliance on the assistance. For example, Gutwin et al. [122] investigated whether providing players with aim assistance in a first-person shooter (FPS) would hinder aiming ability without the assist, or affect their ability to learn other FPS skills. They found that players were not hindered in skill development for either skill when aim assist was present — in opposition to what the guidance hypothesis suggests.

The *retrieval effort hypothesis* refers to the relationship between the amount of effort involved in memory retrieval and the development of memory: "given that retrieval is successful, more difficult retrievals are better for memory than less difficult retrievals" [238]. In other words, increased effort when trying to remember should lead to a better memory of the retrieved information. This difficulty can be operationalized in many different ways, such as spacing out the time between retrievals [238], or making users use less representative symbols in a mental mapping between symbol and colour [88].

4.4 Study 1

We conducted an online experiment to explore whether the amount of assistance provided to a player when navigating an unfamiliar environment would affect route-finding ability and player experience, both when the assistance was present and after it was removed.

4.4.1 Study 1 Experimental Design

Navigation Task

Participants were asked to navigate from a starting position to an end location in a 3D environment. The game environment and the navigation tasks were implemented using the Unity game engine and were deployed online as a browser-based WebGL game displayed on a computer monitor. For each task (also referred to as a route), participants were placed at a predefined starting point, and instructed to move to a location indicated on the map.

Both training and testing phases involved navigating a 3D environment, but there were three different interfaces used during the training phase that provided different amounts of navigation assistance. These interfaces included several elements (Figure 4.3): a mini-map in the top-right corner of the screen, a full-screen map that was accessed by pressing the M key, and a route path that was displayed in the game world. In training tasks, the destination location was indicated with a flag on the maps and in the environment. In testing tasks, none of the three assists were used, and the destination was indicated at the start of the task as an image of an in-game landmark (such as a tank, bars of gold, or an in-game powerup) that the player needed to reach. This destination was also marked in the environment with a flag. To ensure that any spatial learning was acquired incidentally, there was no prior indication before or during training that participants would be tested on their spatial knowledge. In both training and testing, participants had to touch the flag to complete the task. All tasks had a 90-second time limit, after which the system would move to the next route (optimal times to traverse the routes ranged from 5-25s).

Assistance Levels (Training Phase)

The assistance groups were designed to vary the amount of navigation effort required by the participant. The system had three assistance levels, from no assistance to strong assistance, as shown in Figure 4.3.

No Assistance. With no assistance, the player saw the normal first-person view, and had access to a full-screen pop-up map (invoked with the M key) that showed the target destination. In this condition, participants had to identify

Figure 4.3: Left to right: Strong assist, moderate assist, and no assist, and testing UI (showing the "Gold Rush" map).

their own position on the map, plan a route to the destination, and translate directions and distances from the map view to the first-person environment.

Moderate Assistance. With moderate assistance, the interface showed an always-on mini-map in the top right corner of the screen. The pop-up map was also available. In addition to the target icon (a red flag), both maps included an icon indicating the player's current location and direction (similar to the pin icon used in Google Maps). In this condition, participants could see their dynamic progress on the map views — and if they focused on the map, there was less of a requirement to translate to the first-person view.

Strong Assistance. The strong assistance interface provided the same mini-map and pop-up map as described above, but additionally showed the path to the destination with a solid white line permanently drawn in the game environment. The line was a guide only — players could take any route they wanted. This visual effect is similar to the navigational aids used in several commercial games (as discussed above). In this condition, players had to expend far less effort than with the other interfaces — they did not have to identify their location or plan a route, and could simply follow the white line to the destination.

3D Game Environments

We used environments from two commercial 3D first-person shooter games. From *Wolfenstein: Enemy Territory*, we used the map "Gold Rush," and from *Quake Live*, we used the map "Furious Heights." These maps were extracted from the original games and recreated in Unity.

"Gold Rush" (Figure 4.3) is set in a fictional town in northern Africa, and most of the routes in the map take place outside. The town has a variety of streets, walls, buildings, passages, plazas, and staircases. There are several naturalistic landmarks such as palm trees in a town square, vehicles including carts and tanks, and multi-story towers.

"Furious Heights" (Figure 4.4) is set in a fictional multi-level castle, and all of the routes take place indoors. The castle has multiple distinct floors, and the most obvious landmarks in the map are artificial game objects (e.g., a glowing yellow first-aid symbol) that float above the floor. This environment also features two special navigation-related game mechanics: teleporters and jump pads. Both teleporters and jump pads allow players to travel to a higher floor, but are not required to travel through the environment.

Figure 4.4: "Furious Heights" environment from *Quake Live*.

4.4.2 Study 1 Procedure

At the start of the study, participants were told that the study would use WebGL to render 3D environments, and that they would need a relatively fast computer to participate. They were then asked to provide informed consent, and continued through the three phases described below.

Navigation Tutorial and Random Group Assignment

Participants were instructed to complete a simple navigation task in a separate tutorial level (not used in the rest of the study). The task had them walk down a hallway, jump over a small gap, travel through a teleporter, take a jump pad to a higher level, and finally, touch a flag to proceed. This tutorial was intended to give participants a chance to check their system's performance before getting too far into the experiment, and to introduce them to the controls they would use to navigate the virtual environment and the gameplay mechanics of the teleporter and the jump pad. After the tutorial, participants completed demographic and personality trait questionnaires. They were then randomly assigned to one of three assistance groups, and one of two orders.

Training Phase

Depending on their order group, participants started with either the Furious Heights map or the Gold Rush map. They carried out eight different route tasks (as described above). After completing the eight routes, participants completed a questionnaire about their subjective experiences. This training procedure was then repeated for the second map for their order group (for a total of 16 routes).

Testing Phase

After completing training routes and experience questionnaires in both maps, participants moved to a testing phase that involved questions about their spatial knowledge, and four additional route tasks in each map. First, the spatial-knowledge questions asked participants to locate four landmarks and three scenes on a 2D map that was similar to the pop-up map used in the training tasks, but with no marked icons. The landmarks and scenes had all been seen previously in the training phase. Participants answered the spatial-knowledge questions for both maps, in the same order as for training. Second, participants carried out four route tasks in each map: they were shown a picture of a landmark and instructed to go to it as directly as possible, but with no navigation aids (no mini-map, no pop-up map, no route line). The landmarks that were used as destinations for these tasks had all been seen previously in the training tasks (they were either used as destinations or were on a required route); however, since starting points were different for the test tasks, none of the routes had been used previously. After the four routes in each map, participants also completed the same experience questionnaire that was used during training. Participants completed test routes and questionnaire for the two maps in the same order as used for training.

At the end of the experiment, participants completed a debrief protocol and a final questionnaire, giving them an opportunity to provide comments about the experiment.

4.4.3 Study 1 Measures

At the start of the study, we collected measures of prior expertise and personality traits. During the study we collected information about navigation performance, spatial knowledge, and play experience.

Navigation Performance Measures

- **Route completion time.** The system recorded each participant's total time to complete the eight training routes in each map, and the four test routes in each map. The maximum time per route was 90 seconds.
- **Map review time.** The system recorded the total time that participants had the pop-up map open during training.
- **Distance travelled.** The system recorded the total 3D Euclidean distance travelled by the participant for each route (using Unity's default measuring system).
- **Route Completion.** The system recorded the number of routes where the participant reached the destination flag within the 90-second time limit.

Spatial Knowledge Measures

- **Scene-to-map translation:** We presented participants with three screenshots for each environment, and asked them to indicate the location of that scene by clicking on one of six labels (A-F) on a 2D map. The scenes were ones that players had seen during training (although the screenshot was taken with a wider camera angle to show more of the scene).

- **Landmark-image-to-map translation:** We presented participants with images of four landmarks from each environment, and asked them to indicate the location of the landmark by selecting one of six labels (A-F) on a 2D map. All of the landmarks had been seen during training.
- **Route duration estimate:** For each map, participants were shown a 2D map with a route marked on it (not one they had traversed). Based on their experience with navigating the environment, they were asked to estimate the time it would take someone to navigate that path directly.
- **Confidence:** After each spatial-knowledge question, we asked participants to rate their confidence in their answer.

Player Experience Measures

- **Task-Load Index (NASA-TLX)** [125]. The NASA Task-Load Index questionnaire is a widely-used [181] questionnaire to rate perceived workload when completing a task. We used the questionnaire's mental demand, performance, effort, and frustration scales.
- **State Anxiety (SA)** [189]. We anticipated that a participant's state anxiety could differ based on our conditions. We used Marteau and Bekker's six-item questionnaire of state anxiety to measure this construct.
- **Perceived Map Knowledge**. To measure their perceived map knowledge after training, we asked users to rate their knowledge of the layout of the map.

Prior Expertise and Personality Traits

Several factors have been shown to affect a person's navigation performance in virtual environments, such as prior experience with virtual navigation, wayfinding anxiety, and immersive tendencies [327]. To account for these individual differences, we collected the following measures.

- **Experience with our chosen environments**. Participants rated their experience with each of the two games (*Wolfenstein: Enemy Territory* and *Quake Live*) and each of the two maps (Gold Rush and Furious Heights).
- **3D gaming expertise**. We asked participants questions to establish their gaming expertise: how much they self-identified as a gamer, their experience with video games, their experience with keyboard-and-mouse input in games, their FPS experience, and their experience with 3D games.
- **Immersive Tendencies**. We used the Immersive Tendencies Questionnaire (ITQ) [346] to measure participants' tendency to experience presence in virtual environments. The questionnaire consists of three subscales: *involvement* (propensity to get involved with an activity), *focus* (ability to concentrate on enjoyable activities), and *games* (how much they play games and whether they become involved enough to feel like they are inside the game).

- **Wayfinding Traits**. We measured each participant's trait anxiety and tendency to use a "route-learning" strategy or an "orientation" strategy using Lawton and Kallai's [172] International Wayfinding Anxiety Scale and International Wayfinding Strategy Scale, respectively.

4.4.4 Participants and Recruitment

The experiment was deployed on Amazon's Mechanical Turk (MTurk) crowdsourcing platform. MTurk connects willing workers to paid *Human Intelligence Tasks* (HITs) — it has been used for research purposes before and has been shown to be reliable [190]. We had 42 participants complete the experiment. Participants were paid $6 USD for completing the experiment, which took 42 minutes on average.

We randomly assigned all participants to one of the three assistance levels, balancing for self-declared gender. We excluded 12 participants from our analysis for either rating themselves too high in prior experience (moderate experience or higher) with either of the two games or maps used, or for having too low a framerate on their system (<15 FPS). This left 15 female and 15 male participants (mean age 32.7, SD=7.56, min=20, max=53). Ten participants completed each of the three assistance conditions.

4.4.5 Data Analyses

For each participant, we aggregated performance data from both maps. This provided mean performance measures per participant for completion time, distance travelled, and routes completed (for each of training and testing). We also aggregated each participant's scores across both maps for the three types of spatial-knowledge question (scene location, landmark location, route duration estimation).

Due to our between-subjects design, we used covariates to acknowledge trait difference in anxiety and spatial ability in our participants. Covariates were chosen based on correlation between traits and dependent measures. For the subjective measures, four covariates were included: ITQ's focus subscale, ITQ's games subscale, wayfinding anxiety, and gaming expertise. For our objective measures, five covariates were included: ITQ's focus subscale, ITQ's games subscale, orientation wayfinding strategy, gaming expertise, and gender. Note that in both studies, one-way ANOVAs showed no significant group differences in the trait measures used as covariates, indicating that random assignment did not result in one of the assistance groups having skewed levels of any trait measure.

We expected differences between assistance groups during the training phase, and so performed two multivariate analyses of covariance (MANCOVA) using only the training data — one analysis for the subjective measures of experience (with four covariates), and one for the objective measures of performance (with five covariates). To test for effects of assistance level on performance and experience in the testing phase, we similarly carried out two MANCOVAs using the dependent measures collected during testing. Alpha was set at 0.05, and all pairwise comparisons used the estimated marginal means and Bonferroni corrections.

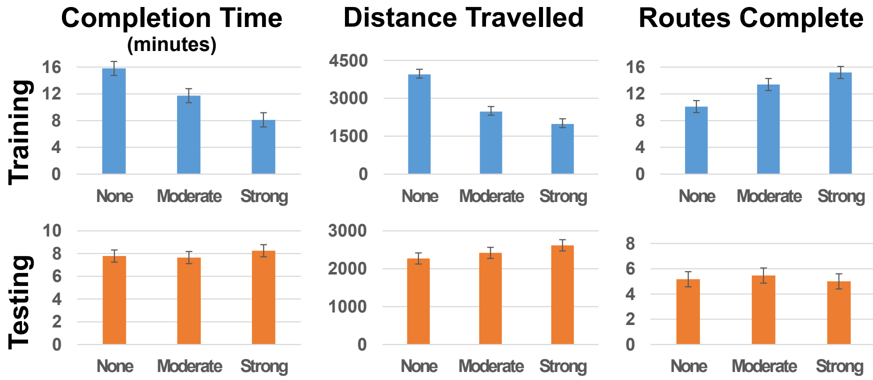


Figure 4.5: Descriptive statistic results for the performance measures of the 16 training routes and 8 testing routes. Values are estimated marginal means; error bars are ± s.e.

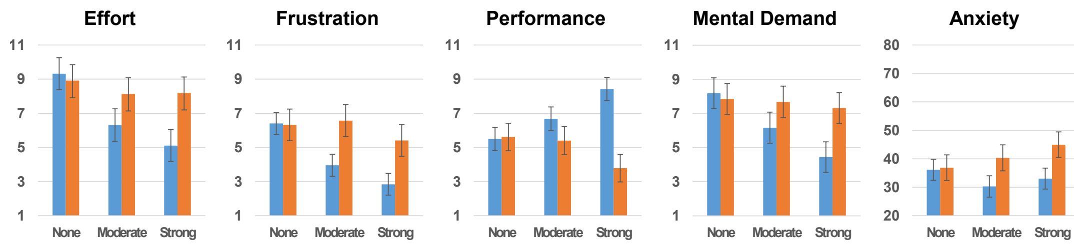


Figure 4.6: Experiential results after training (with assistance) and testing (with no assistance). Values are estimated marginal means, error bars are ± s.e.

4.4.6 Study 1 Results

We first look at how assistance affected participant subjective experience and performance during training, and then report whether the type of assistance used in training affected performance or experience in the testing phase (where all navigation assistance was removed). Note that because there were different numbers of tasks in training and testing, and because the actual routes involved different start and end points, it is not possible to directly compare each group's training performance to their testing performance.

Effects of Assistance Level in Training

For the performance measures, there were significant main effects of assistance level on the time taken to complete the training routes, the time spent reviewing the map, the distance travelled, and the number of routes completed within the 90-second time limit (see Table 4.1). Pairwise comparisons showed that participants with moderate or strong assistance spent less time completing the training routes ($p_m = .038$, $p_s < .001$), travelled less distance ($p_m < .001$, $p_s < .001$), and completed more routes within the time limit ($p_m = .044$, $p_s = .002$) compared to participants with no assistance. Furthermore, participants with strong assistance spent less time reviewing the map than participants with moderate assistance ($p = .050$) or with no assistance ($p = .005$).

For the experiential measures, there were significant main effects of assistance on effort, frustration, perceived performance, and mental demand, but not anxiety or self-rated map knowledge (Table 4.1). Pairwise comparisons revealed that participants experienced higher effort ($p = .013$), frustration ($p = .002$), mental demand ($p = .022$), and perceived performance ($p = .018$) with no assistance compared to strong assistance. In addition, participants experienced higher frustration ($p = .043$) with no assistance compared to moderate assistance. No other pairwise differences were significant.

Effects of Assistance Level in Testing

For the performance measures, there were no effects of assistance type on any measure, including completion time, distance travelled, and number of routes completed. There were similarly no main effects of assistance type on the spatial-knowledge questions, including landmark location, scene location, and route duration estimation. In addition, there were no differences in participants' confidence ratings for landmark or scene location. For experiential measures, there were also no significant effects of assistance type on any measure, including effort, frustration, perceived performance, mental demand, and anxiety (See Table 4.1).

Summary of Results

In training, participants who had navigation assistance spent significantly less time completing the routes, spent less time looking at the map, and travelled a shorter distance. This reduced exposure to the 3D environment, however, did not translate into reduced performance in unassisted test tasks when compared to participants who had trained with no assistance (and whose testing experience was therefore much closer to their training experience).

Navigation assistance also led to significantly better scores during training for perceived effort, frustration, performance, and mental demand. As with the performance measures, removing the assistance in testing did not lead to a worse experience than what was reported by participants who had trained with no assistance — we observed no differences in the subjective measures between the groups.

4.5 Study 2

Our first study suggested that navigation assistance helped participants during training, and did not hurt them (relative to the no-assistance group) when the aids were removed. However, the first study provided only a short training phase (eight routes in each map), and we wanted to determine whether differences might emerge if participants had more training time to learn the maps. Therefore, we conducted a follow-up study that used a much longer training period.

4.5.1 Study 2 Experimental Design

The second study was similar to the first but used three training sessions over three days. The study was again deployed on Amazon's Mechanical Turk, and we reproduced all aspects of the procedure described above. Participants completed the same navigation tasks for training and testing, with the same assistance levels used during training. We also used the same dependent measures.

The difference in Study 2 was the duration of training — three sessions on three consecutive days, totalling 48 navigation tasks instead of 16. The testing phase was identical to Study 1, and immediately followed the third training session.

4.5.2 Study 2 Participant and Recruitment

We recruited participants on MTurk using a qualification task, limited to 100 participants, in which people completed the tutorial navigation task and demographics questionnaire that we used at the start of Study 1. The qualification task took less than 5 minutes and paid $0.50 USD. The average framerate was logged during the navigation task. We excluded people from the main study if they had a framerate lower than 45 FPS (to ensure high framerates in the more graphically intense maps), or if they had moderate or greater experience with the two games and maps described above. After exclusions, 73 participants were eligible to participate in the main study.

The first day of the experiment was open to the first 50 participants who accepted the task on MTurk, and consisted of the personality trait questionnaires and a 16-route training session (as described earlier, 8 routes in each of two maps, seen in balanced order). Participants were paid $3.50 USD for completing the first day. The second day consisted only of the same 16-route training session, and participants were paid $3 USD. The final day consisted of the final 16-route training session, the same spatial-knowledge questionnaires as in Study 1, and the 8-route testing session, with the same testing routes as Study 1. Participants were paid $4.50 USD for day 3.

Of the 50 who started the multi-day study, 46 completed all three days. We excluded two participants from our analysis due to logging errors, leaving us with 44 participants (29 male, 15 female, mean age of 33.7, SD=8.68; min=20; max=59). Male and female participants were evenly distributed among the assistance groups as in Study 1. The 44 participants were randomly assigned to assistance levels: 14 had no assistance, 16 had moderate, and 14 had strong.

4.5.3 Study 2 Data Analyses

To evaluate the training sessions, we used a repeated measures MANCOVA — the same MANCOVA model as in Study 1 but with Day (one, two, and three) as an additional within-subjects factor. The testing session was analysed using the same statistical model as in Study 1.

4.5.4 Study 2 Results

Effects of Assistance Level and Day on Training

In terms of the performance measures, there were no main effects of day on the time taken to complete the training routes ($F_{2,72} = 0.73$, $p = .487$), the time spent reviewing the map ($F_{2,72} = 0.67$, $p = .51$), the distance travelled ($F_{2,72} = 0.72$, $p = .492$), or the number of routes completed ($F_{2,72} = 1.6$, $p = .218$). There were significant main effects of assistance on the time taken to complete the training routes, the time spent reviewing the map, the distance travelled, and the number of routes completed (see Table 4.1). Pairwise comparisons showed that each level

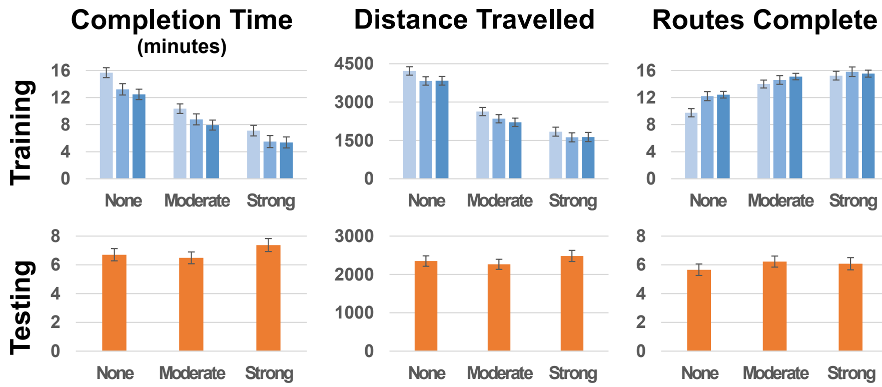


Figure 4.7: Descriptive statistic results for the performance measures of the 16 training routes and 8 testing routes for day 1, 2, 3, and testing. Values are estimated marginal means; error bars are ± s.e.

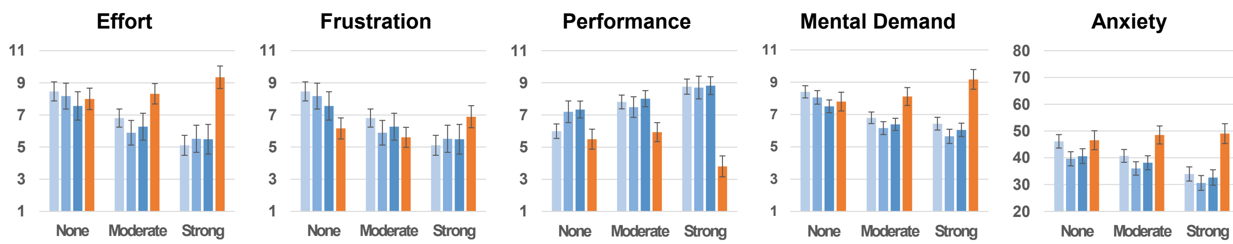


Figure 4.8: Descriptive statistics for subjective measures in Study 2 for day 1, 2, 3, and testing. Values are estimated marginal means; error bars are ± s.e.

of assistance decreased the time spent (all $p < .029$) and the distance travelled (all $p < .011$). Furthermore, both moderate and strong assistance reduced the time spent on the map (all $p < .020$) and increased the number of routes completed (all $p < .001$). There were no significant interactions between day and assistance on time, time reviewing the map, or distance; however, a significant interaction between day and routes completed ($F_{4,72} = 3.7$, $p = .009$, $\eta_p 2 = .169$) showed that the differences between the assistance techniques were less pronounced over time.

In terms of the subjective results, there were no main effects of Day on effort ($F_{2,74} = 0.36$, $p = .696$), frustration ($F_{2,74} = 0.36$, $p = .696$), perceived performance ($F_{2,74} = 1.2$, $p = .299$), mental demand ($F_{2,74} = 0.28$, $p = .759$), anxiety ($F_{2,74} = 2.4$, $p = .095$), or self-rated map knowledge ($F_{2,74} = 2.8$, $p = .066$). There were significant effects of assistance level on effort, frustration, perceived performance, mental demand, and anxiety (see Table 4.1). Pairwise comparisons revealed that participants who trained with strong assistance experienced less effort ($p = .040$), frustration ($p = .040$), perceived performance ($p = .021$), mental demand ($p < .001$), and anxiety ($p = .017$) than those who received no assistance. In addition, those with moderate assistance rated their mental demand as lower than with no assistance ($p = .001$). There were no interactions between Day and Assistance on any of the measures.

Effects of Assistance Level on Testing

There were no significant main effects of assistance level on any of our performance measures, including time taken, distance travelled, and routes completed. There were also no main effects of assistance level on any of our spatial-knowledge questions, including landmark location, scene location, and route duration estimation. There were also no differences in confidence ratings (see Table 4.1). There were no significant effects of assistance on any of the subjective measures, including effort, frustration, perceived performance, mental demand, anxiety, or self-declared map knowledge (see Table 4.1).

Summary of Results

The findings from Study 2 mirror the findings from Study 1, with the additional result that anxiety during training was higher without assistance. As Figures 4.7 and 4.8 show, the longer training period did not substantially change any of our subjective or performance measures.

4.6 Discussion

4.6.1 Summary of Results

Our expectation, based on the guidance and retrieval-effort hypotheses, was that increased navigation effort in training would result in better spatial understanding of the map, and thus better performance during testing. However, this did not occur in either study. Although it was clear that both kinds of navigation assistance helped when they were present, we found no differences in route-finding performance when assists were removed, and no difference in spatial-knowledge questions (regardless of the duration of the training period — 16 or 48 routes). The lack of differences across assistance level is even more surprising given that the overall time in the game world for the no-assistance group was approximately double that of the strong assistance group, and 1.5 times that of the moderate group. In addition, navigation assistance improved subjective experience. When assistance was taken away during the testing session, the differences between assistance groups disappeared.

4.6.2 Possible Explanations for Results

There are several possible reasons why navigation assistance did not hinder route-finding performance once the assists were removed. First, it may be that incidental learning took place during training (as suggested by previous work [11, 126]), even though the navigation tasks were made easier by the assistance. One mechanism for this incidental learning could be that the assistance reduced the cognitive effort of the task to the point where players could pay more attention to their surroundings. We believe it is important that players still experienced the entire route and participated in traversing it, even though they were assisted — if the assist had taken the player out of the route (e.g., by teleporting them to the destination), the opportunity for incidental learning would have been much reduced.

A second (and related) possibility is that spatial learning was somehow hindered by the no-assistance condition. It is possible, for example, that the tasks were so difficult for novices that they were outside Vygotsky's "zone of proximal development" where people learn best [325] (similar to the "flow state" in Kiili's learning model [162]). Players in the no-assist condition may have been unable to learn the maps effectively because they were overwhelmed by the basic actions of locating themselves on the map, recognizing landmarks, and understanding the relationships between the pop-up map and the first-person view of the game world.

4.6.3 Implications for Game Designers

The use of assistance did not have a significantly negative effect on performance in the test tasks — even despite the large differences in training time. This finding has intriguing consequences for utilizing route guidance as an assistance technique in games.

Skill Development Considerations

Some games are better candidates for navigational assistance than others. There are many games that require players to operate in the same environment many times, so players must become familiar with the maps if they want to succeed.

Skill assists have been investigated previously to improve player balancing [47, 322], but a common concern is that providing assistance will result in player reliance. Our results add to increasing evidence that some degree of assistance does not necessarily reduce learning. For example, Gutwin et al. [122] found that providing aim assistance to novices did not hinder the development of either aiming skill or overall FPS abilities. When games require many skills, providing assistance in one area (and thus reducing effort overall) can allow players to improve in other areas. For example, navigation assistance in an FPS game could free the player to work on skills such as aiming, movement, or monitoring audiovisual cues [154].

It is even possible that providing dynamic skill assistance can enhance learning. Kiili's [162] experiential gaming model proposes that a balanced game that facilitates the player reaching a flow state will result in the strongest learning. This corresponds to Anderson and Bischof's suggestion that guidance gradually be removed as a learner gains expertise [9]. Gradual removal of the assist would also reduce the sharp drop in experience measures that we observed between training and testing.

There are still likely to be cases in which providing a strong assist for a particular skill results in a dependency on the assist. Further research is required to determine when and where this effect appears.

Implementation Considerations

We chose our assistance techniques because they are already used extensively in games. The augmented (moderate-assistance) map works if the player has time to study it, and the glowing trail works if the destination is known to the system. The trail allowed participants to reach a higher level of performance in less time than the augmented map, so it may be worthwhile to consider implementing it in more games. In games where destinations are not known (or where there are many possible destinations), it is not clear how well this method will work.

Assistance could be made context-sensitive: for example, if a player has no weapons or is low on health, the game could show a trail to the nearest weapon or the nearest health pack. A player's role in a team game could also determine which routes are visualized for that player (e.g., a trail to a wounded player for a medic role). Finally, for scenarios in which navigational assistance is not possible, or where the player chooses to turn off the aids [348, 229], it appears that the use of even strong assistance early in a player's experience will not significantly affect their long-term performance.

4.7 Conclusions and Future Work

3D game navigation is difficult for novices. Games can provide visual assistance for wayfinding, but there is a risk that players will become overly reliant on these assists and fail to develop independent spatial understanding. We investigated the benefits and potential risks of navigation assistance through two online studies. We found that having assistance helped significantly when it was turned on — and when it was turned off, navigation performance did not suffer. This work provides new evidence that navigation assistance is a valuable tool to help novices deal with the complexities of 3D games, and that incidental learning of 3D game environments can occur, even with strong assistance.

In our future work, we plan to examine several issues raised by our experiments. First, we will look in more detail at whether navigation assistance acts as a scaffold that can improve learning (by putting players in a flow zone or zone of proximal development). Second, we will test other kinds of navigation assistance and other game environments to see if our results hold in different settings. Third, we will develop versions of the assist that gradually disappear, to see if this further improves spatial knowledge. Fourth, we will test our techniques in actual play settings, to see if navigation assistance can improve play experience and player balancing in real games.

		Study 1			Study 2		
		$F_{2,23}$	p	η_p^2	$F_{2,37}$	p	η_p^2
Effort	Training	**5.31**	**.013**	**.316**	**3.61**	**.037**	**.163**
	Testing	0.21	.812	n.s.	0.99	.381	n.s.
Frustration	Training	**8.10**	**.002**	**.413**	**3.61**	**.037**	**.163**
	Testing	0.43	.655	n.s.	0.89	.418	n.s.
Perceived Performance	Training	**4.66**	**.020**	**.288**	**4.12**	**.024**	**.182**
	Testing	1.53	.238	n.s.	2.95	.065	n.s.
Mental Demand	Training	**4.33**	**.025**	**.274**	**12.0**	**<.001**	**.394**
	Testing	0.09	.915	n.s.	1.37	.267	n.s.
Anxiety	Training	0.60	.556	n.s.	**4.33**	**.020**	**.190**
	Testing	0.80	.460	n.s.	0.14	.874	n.s.
Map Knowledge	Training	0.33	.725	n.s.	0.22	.800	n.s.
Time	Training	**12.4**	**<.001**	**.529**	**26.6**	**<.001**	**.596**
	Testing	0.33	.723	n.s.	1.02	.369	n.s.
Distance Travelled	Training	**25.2**	**<.001**	**.696**	**57.2**	**<.001**	**.761**
	Testing	1.34	.292	n.s.	0.57	.571	n.s.
Routes Completed	Training	**8.29**	**.002**	**.430**	**16.5**	**<.001**	**.478**
	Testing	0.16	.851	n.s.	0.57	.573	n.s.
Map Time	Training	**6.64**	**.006**	**.376**	**12.9**	**<.001**	**.418**
Landmark Identification	Testing	0.21	.815	n.s.	0.22	.801	n.s.
Landmark Confidence	Testing	0.42	.662	n.s.	0.04	.959	n.s.
Scene Identification	Testing	0.54	.591	n.s.	1.99	.152	n.s.
Scene Confidence	Testing	0.13	.883	n.s.	0.17	.841	n.s.
Route Estimation	Testing	0.32	.730	n.s.	0.48	.241	n.s.

Table 4.1: MANCOVA results for the effects of assistance on the subjective and objective measures for the training session and testing session.

5 Introduction to Manuscript B

Like Manuscript A, this manuscript also explores the effects of concurrent guidance on performance and learning. I learned from Manuscript A that in the context of navigating a virtual environment, concurrent visual guidance does not affect learning to the extent that one might expect from prior research (as summarized in Subsection 2.4.1). Manuscript A also prompted a thought: if following a trail did not negatively affect learning, then at what point might concurrent guidance be detrimental to learning? Surely if someone were simply teleported to their destination within an environment they would be unable to navigate to that destination on their own. Therefore, we hypothesized that at some point, a "stronger" amount of guidance could lead to reduced learning.

Subsection 2.4.1 discusses the idea of *concurrent mechanical guidance* as a means of forcing one to make a correct response or to limit the response that one might make. This improves performance by limiting errors. In the context of navigating within a virtual environment, I decided to implement concurrent mechanical guidance by having the player travel through the environment as if they were on rails (as in, the rails forming a railroad). Players would simply need to hold down a button and they would get taken in the correct direction. At least on the surface this approach would seem to be "stronger" than concurrent visual guidance because players can ignore the trails if they choose (no particular response is forced), yet players cannot ignore the rails; they are forced to travel along them or not engage with the navigation task at all.

This manuscript presents two studies. The first is a re-interpretation of the results from Manuscript A, and the second is a new study wherein we evaluate the effects of the rails guidance on performance and learning, compared to no guidance and the trails guidance. We wanted this new experiment to utilize only the urban environment (i.e., "Gold Rush") used in Manuscript A within this new experiment, as the environment from *Quake Live* ("Furious Heights") featured teleporters and jump pads, two game elements that may have been confusing to participants. To make potential comparisons[1] between the two studies in Manuscript B, we performed analyses using the data from Manuscript A's Study 2, but only for the "Gold Rush" environment. This actually led to slightly different results compared to Manuscript A, where we actually did find some difference on the transfer test between Position and Trails guidance (what we called "moderate" and "strong" in Manuscript A) — the position guidance was *slightly* better for learning than the other two types of guidance.

Study 2's biggest change, as mentioned, is the inclusion of concurrent mechanical guidance, but it also featured a few additional changes. Study 2 included three levels of guidance: no guidance, trail guidance, and rail guidance. We removed the map from the game entirely and so the no guidance condition does not even include a map. Our thought

[1]To clarify, no *statistical* comparisons are possible given that the experimental conditions are different and the samples are different, but it helps the story of the paper be more cohesive if there isn't a second environment used within one of the studies and we additionally don't have to try to explain how the teleporters and jump pads may have affected the results in one of the studies.

was that it was possible that players in our previous studies may have spent their time studying the map rather than looking around the environment. Removing the map eliminates this possibility, but may cause players to get even more lost. The environment used is one where players can get easily lost, and some sections of the environment include no relevant landmarks. Such sections were removed, which may have made the environment easier to learn for everyone, but it also means that the unguided players would have a better chance of stumbling upon the correct destination by chance, hopefully aiding their learning of the environment.

It should be clear by these changes that our intention was to give the no-guidance group a good chance of being able to learn the environment this time around. We wanted to reproduce the findings of Manuscript A and remove all doubt that what we found was a result of random chance. Therefore, this new study also had participants spend more time training within the environment (four days of training compared to three) and complete two separate tests (transfer and retention), each with eight routes instead of four.

This time around, we did observe that the no guidance group learned the environment better than the trails or rails groups, but still less than one might expect based on theory. There were differences present in the transfer test, where the no guidance group performed better (in terms of completion time) than the rails or trails groups. There were, however, no differences between the groups in terms of completion time on the retention test. The no-guidance group was able to handle new routes better, but they weren't significantly better at navigating the routes they trained on than the other groups. As far as how the "stronger" guidance affected learning, there were no statistically significant differences between the rails and trails groups in any of our performance measures.

5.1 Methodological Clarifications

5.1.1 Names of Conditions

Study 1 of Manuscript B uses the same data as was presented in Study 2 of Manuscript A. However, I use different names for the assistance conditions on each of the manuscripts. In Manuscript A I name the three levels of assistance "no assistance", "moderate assistance", and "strong assistance". In Manuscript B I name those three levels of assistance, "map assistance", "position assistance", and "trail assistance".

5.1.2 Sample Size and Effect Sizes

Participant counts were determined by resource constraints rather than an a-priori power analysis [170]. Therefore, here I will report the effect sizes we found for the training and testing completion times for Study 2.

In Study 2, I found a strong effect size of training with assistance (1.30) and a medium effect size of assistance on transfer (0.50) and a small effect size on retention (0.33).

5.1.3 Participant Payment

In Study 1, the payment information was reported in Section 3.1.3. In Study 2, participants were paid $4.00 USD for day 1, $2.00 USD for days 2-4, $2 USD for day 5, and $3 for day 6. Day 1 took 21.0 minutes on average to complete (SD=8.9), days 2-4 took 10.2 minutes on average (SD=6.2), day 5 took 10.0 minutes on average (SD=6.5) and day 6 took 15.8 minutes on average (SD=10.3). This works out to $11.63 USD per hour for Study 2, which is more than the United States' federal minimum wage of $7.25 an hour, where our participants were from.

5.1.4 Retention Session

For the retention session, I had participants wait one week before returning and attempting to navigate the environment again. The length of time had to be long enough that any temporary effects of training would disappear [257], but not so long that participants would have forgotten what they had learned. I chose one week and not a shorter or longer interval because it is common for players to step away from a game for a week. For example, friends might meet up once a week to play a game.

In the manuscript, I did not compare the participants' testing performance to their retention performance in my statistical tests. This was because of the differing counts of participants in training compared to testing. Furthermore, including it would have guaranteed a violation of the assumption of sphericity [102], as the differences between each pair of measurements within a group would not be equal due to the large change in performance between the last day of training and the retention session for two of the groups. However, there are no issues if only one pair of measurements are included in the RM-ANCOVA, so in the appendix (Section E.2.2) I present the within-subjects results from an RM-ANCOVA that compares the last training day and the retention session.

5.2 Additional Analyses, Results, and Figures

When writing Manuscript B, there were some results and figures that did not make it into the final paper due to space constraints. These have been included in the appendix, in Section E.2. This includes:

- Reporting the statistical effects of the covariates used.
- Reporting on possible gender differences.
- Reporting on what our participants said regarding strategies used to remember routes or landmarks.

5.3 Publication and Individual Contributions

This work was published as [152]:

Johanson, C., Gutwin, C., & Mandryk, R. L. (2023). Trails, Rails, and Over-Reliance: How Navigation Assistance Affects Route-Finding and Spatial Learning in Virtual Environments. *International Journal of Human-Computer Studies*, 178, 103097. https://doi.org/10.1016/j.ijhcs.2023.103097

My contribution to the work included: designing the experiment, implementing the digital system, performing data analyses, and writing the majority of the manuscript.

6 Manuscript B

Trails, Rails, and Over-Reliance: How Navigation Assistance Affects Route-Finding in Virtual Environments

6.1 Abstract

Many novices struggle with navigation in 3D virtual environments — they frequently get lost and are unable to find objects and locations. In some virtual environments, novices are provided with *navigation assistance* (e.g., mini-maps, directional markers, or glowing trails) that help them move around in the world. However, it is possible that providing navigation assistance could lead to over-reliance, where the novice's dependence on the assist means that they never develop a mental model of the environment that would allow them to navigate on their own. To investigate both the benefits and potential risks of navigation assistance in virtual environments, we carried out two online studies in which participants carried out route-finding tasks with different types of navigation assistance. Participants completed training trials, in which they practiced a set of routes with the assist, and transfer trials, in which they had to navigate without the assist. The studies focus on two questions: whether assistance improves performance and user experience when it is present, and whether assistance leads to over-reliance and a drop in performance when the assist is removed. For the first question, both studies found that navigation assistance substantially improved performance and subjective experience while it was present — clearly showing that assistance can improve virtual environments for novices. For the second question, we found mixed evidence regarding the problem of over-reliance: the first study showed no performance differences between the highest and lowest levels of navigation assistance when the assist was turned off; the second study showed that there was a performance reduction when the assist was removed, but that the size of the reduction was much smaller than the improvement provided during training. We found that even when the navigation assist was extreme (e.g., pressing a button to be automatically taken in the correct direction), participants were still able to navigate the trained routes, suggesting that incidental learning does successfully occur in virtual environments. Our studies suggest that designers of virtual environments should strongly consider providing navigation assistance: assists can improve a novice user's performance and experience by reducing navigation problems, and the risks of over-reliance appear to be small in comparison to the benefits for inexperienced users.

6.2 Introduction

Three-dimensional virtual environments (VEs) are now an extremely common platform for digital interaction: VEs are regularly used for game worlds (e.g., *World of Warcraft*, *Team Fortress 2*, or *Skyrim*), for training environments (e.g., aircraft or driving simulators, emergency procedures training), for exploratory educational activities (e.g., virtual museums and art galleries), for architectural previews (e.g., walkthroughs of planned buildings), and for social interaction (e.g., *Second Life*, *There*, *Metaverse*). There are many types of 3D virtual environments, but all VEs are spatial environments in which users must *navigate* from place to place. Navigation is "coordinated and goal-directed movement through the environment" [206, p. 257] and includes two main components: route finding (selecting a goal location and planning a path to that goal — using strategy, memory, and decision making) and locomotion (actually moving through the space to reach the destination) [206].

Previous studies have shown that navigation and spatial learning in 3D virtual environments are difficult for many users [e.g., 68, 156, 85] — more so than in the physical world. There are several possible reasons for this difficulty, such as the reduced kinaesthetic feedback of virtual locomotion, the reduced field of view, the relative lack of visual details and distinctiveness that can be used as landmarks, and the lack of non-visual sensory information [328]. In addition, the size of the virtual environment, the density of objects within it, and the amount of movement required can also affect navigation [66].

These difficulties have led designers of several VEs to provide navigation assists to users. Many different types of assist have been considered, such as verbal directions, superimposed arrows, compasses, maps showing the user's location, audio cues, landmarks, or even glowing trails that the user can follow to reach a destination [208, 188]. Some of these assists are directly inspired by those available in the physical world while others are only possible within virtual environments.

Studies have shown that navigation assists can be effective in improving navigation performance [e.g., 209, 77]. However, there is little understanding of the longer-term effects of navigation assists on spatial learning of the environment — in particular, whether assists can lead to users becoming overly reliant on the assist, leading to reduced spatial learning of locations, routes, and the overall layout of the environment [299, 300, 301]. That is, navigation assists may help novices in the short term, but if users come to depend on the assist, they may be unable to navigate effectively when the assist is not available.

The unavailability of navigation assists could be a frequent occurrence in many virtual environments. For example, if the system does not know the user's destination (e.g., in an open-world game), it cannot provide an assist; or, if the user decides to take a shortcut or change their destination in the middle of a route, any provided assist will be incorrect and the user will need to navigate using only their spatial memory.

The risk of over-reliance on a navigation assist is an example of the *guidance hypothesis* [257]. A common finding in previous studies of skill development and learning is that guidance improves performance for novices when it is present, but at the cost of reduced learning that causes a performance drop when the guidance is removed [257, 16, 234, 274, 330, 135, 182, 343]. The reduction in learning can arise because guidance allows users to pay less attention to feedback and to expend less effort overall — previous studies have shown that intentionality and effort strongly affect learning [257, 88].

Navigation, however, is different from many of the skills tested in past work on the guidance hypothesis, because spatial learning has also been shown to occur *incidentally* and without intentional effort from the learner [11, 320, 126]. For example, a learner can acquire landmark and route knowledge simply by following a guide through an environment [320]. This natural ability to learn about a 3D space may arise because an understanding of our surroundings was critical for the survival of early humans. There is some debate, however, around how much spatial learning occurs incidentally, and there are studies that point to real-world navigation assists such as global positioning systems (GPS) contributing to problems in spatial learning [e.g., 42, 146, 176].

These contrasting theories mean that it is difficult to predict the effect that navigation assistance will have on spatial learning of a virtual environment. To provide new empirical knowledge about this issue, we carried out two studies[1] that provided different levels of navigation assistance for novices in virtual environments, and measured both the effect on performance (when the assist was present) and the effect on learning (when the assist was taken away). We tested several conditions that provided different levels of assistance and required different levels of effort from the user: no assistance at all, a map of the environment (either with or without the user's location shown), a glowing trail that the user could follow, and an "on rails" condition where users only had to press a key to continue moving in the correct direction.

For both studies, we had two research questions: (RQ1) Will navigation assistance improve performance and user experience when it is present? and (RQ2), will navigation assistance hinder spatial learning and cause over-reliance on the assist?

The first study tested three types of assistance that provided increasing levels of guidance (a static map, a map showing the user's location, and a glowing trail to the destination). Our results showed that the two higher levels of navigational assistance substantially improved both performance and subjective experience (RQ1), and that the higher assistance levels did **not** reduce participants' spatial knowledge of the environment — in all assistance conditions, participants could navigate the environment at a similar performance level after the assistance was removed (RQ2).

In the second study, we chose assistance conditions that explored an even wider range of user effort — a baseline condition with no map (which required more effort than any of the conditions in the first study), a glowing trail, and an "on rails" condition where participants only had to press a key to keep moving in the correct direction (thus requiring very little navigational effort). Results from the second study showed that the higher levels of assistance again substantially improved performance and user experience (RQ1); but unlike the first study, both of the higher assistance levels led to reduced performance compared to the no-assistance condition, once the assists were removed (RQ2). However, in a retention test one week after the main study (also with assists removed), all of the conditions performed similarly.

Together, these two studies provide several contributions that can change the way designers of virtual environments

[1]The first study was initially published in the Proceedings of the Annual Symposium on Computer-Human Interaction in Play [**?**]; here we present a revised and re-analyzed version. The second study has not been published elsewhere.

think about and apply navigation assistance.

- We show that navigation assistance substantially improves novices' navigation performance in virtual environments, and also substantially improves subjective experience.
- We show that the amount of user effort required for navigation does not accurately predict spatial learning of the environment: Study 1 found no difference in performance once assists were removed, comparing between the most-effortful and least-effortful forms of training; Study 2 only found an advantage for the highest-effort condition, and this difference largely disappeared after a week.
- We show that although more extreme forms of navigation assistance may slightly hinder the development of spatial memory compared to a no-assist condition, the effects are not disastrous: participants who trained with higher levels of assistance in Study 2 were still able to successfully navigate without the assists, and the benefit of having the assist outweighed any detriment.

Our studies suggest that navigation assistance provides substantial benefits and relatively small drawbacks for novices in virtual environments. Our results add to our understanding of how assistance affects spatial learning, and provide useful information for designers who want to make their systems more accessible to novices.

6.3 Related Work

6.3.1 Learning with Guidance

Before discussing the skill of navigation and how guidance affects navigation, we present related work on different types of guidance as well as the general effects of guidance on skill learning.

Skill learning can generally occur without any explicit support or guidance, through a trial-and-error approach where a learner makes errors and observes the results until the correct response is acquired [135, 272, 192, 234]. The role of guidance is to aid a learner so that they can execute skills with reduced errors [272, 257]. Learning in this way has been described as "guided" learning [274, 275], "errorless" learning [234, 273, 192, 139], or "error-free" learning [273, 272, 192, 155]. However, there is debate as to whether or not making mistakes is essential to learning psychomotor skills [135, 274, 272, 258, 192, 155, 257], as identifying and correcting mistakes is often a key component of many theories of learning (such as operant conditioning [341] and experiential learning [166]).

Types of Guidance

Guidance refers to instructions or assistance given to the learner by some external source before performing an action or while an action is ongoing. The goal of guidance systems is to assist learners in forming mental representations of the task they are trying to complete [136]. Guidance can also assist learners less directly, as it is known to have motivating effects for novices [302]. Guidance can either *precede* the execution of a task or occur *concurrently*, and can be presented mechanically, visually, or verbally (though we do not look at verbal guidance in this work).

Visual guidance is guidance that is presented to the learner with the intent of helping the learner develop a mental image of the task as well as how to complete it [136]. Visual guidance that precedes task execution takes the form of videos, charts, visual assists, or demonstrations [136, 216]. Visual guidance that is presented concurrently takes the form of visual assists that the learner can leverage while carrying out the task, provided by an instructor (e.g., following a demonstration [136]) or a software system (e.g., participants drawing a pattern could be shown the desired pattern by a computer monitor if they veer off-target [139]).

Mechanical guidance is any type of guidance that introduces a mechanical restriction on the learner to minimize errors or force a particular response [216, 136]. Mechanical guidance is primarily provided during an action [216] and often guarantees that performance will be high [343].

Efficacy of Guidance

The guidance used in our work — navigation assistance — is concurrent guidance that is either visual or mechanical, presented concurrently with the task of navigation. This type of guidance, for the most part, significantly improves *performance* while it is present, but potentially at the cost of reduced *learning* when evaluated by testing the participants again without the guidance [257, 16, 234, 274, 330, 135, 182, 343].

However, it must be noted that although several different experiments have shown a similar reduction in learning, the tasks used in those experiments were relatively simple (e.g., reproducing specific patterns [16], or manipulating specific inputs with one's hands and feet [274]). Such tasks are not difficult to learn using inherent response-produced feedback and therefore the findings may not apply to more complex tasks [350]. For example, findings for a study involving a ski simulator found that providing mechanical guidance benefited performance and learning [351], and similar findings exist in the context of musical training [119] and gymnastics [127]. One example of mechanical guidance found in games is aim assistance. Vicencio-Moreira and colleagues found that aim assistance was an effective way to improve short-term performance [323, 322], and Gutwin and colleagues found that long-term use of aim assistance had no detrimental effect on learning [122].

It is also worth considering the efficacy of visual guidance given before the task — in particular, demonstrations. The mechanical ("on rails") guidance utilized in this work would allow participants to simply observe the environment around them as they have the correct path demonstrated for them. Studies have found that allowing participants to observe a demonstration of a task improves performance when compared to no demonstration [232, 196, 243]. These demonstrations are effective when they are able to reduce uncertainty on the part of the learner [216]. In one example, Pollock and Lee tested the effect of watching another person play a game (Microsoft's *Olympic Decathlon* [276]) on performance and found that participants' performance was improved by watching another player play before playing themselves, regardless of whether the other player was an expert or a novice at the game [232].

6.3.2 Navigation

A wide variety of research has been carried out to investigate the ways that humans learn and perform navigation in real-world environments — for example, researchers have looked at the development of spatial knowledge in children [e.g., 124], sex differences in navigation [e.g., 51, 172], and theoretical models for navigation [e.g., 53]. One major focus in navigation research is on wayfinding, the process by which people orient themselves to an environment and move from place to place. Early work identified three kinds of knowledge that are important for wayfinding, and that are associated with increasing spatial understanding [299, 300, 301]:

- *Landmark knowledge* involves remembering specific objects or settings in an environment — such as a statue or a building in a city centre.
- *Route knowledge* involves understanding how to navigate between specific locations, and the actions required to reproduce a specific path between them. Route knowledge often builds on landmark knowledge (e.g., by linking different landmarks together).
- *Survey knowledge* is a map-like mental representation of an environment and is the highest form of spatial understanding. Survey knowledge allows people to navigate skillfully, estimate relative distances, and choose alternate routes to objectives.

There are two ways in which people can gain this spatial understanding of an environment [68]. First, people learn through direct exposure to their surroundings — that is, simply being in an environment and moving through it. Second, external information sources such as maps provide other forms of spatial learning. When used in an actual navigation task, maps require that users identify their own location on the map, and then translate orientations, directions, and distances from the map representation to the actual environment.

Navigating Virtual Environments

Navigation in virtual environments has also been extensively studied. One main interest is in whether virtual environments can be used as training simulations for real-world navigation [328], and whether spatial knowledge and wayfinding ability transfer to real environments. Researchers have also identified that navigational difficulties are common in virtual environments [e.g., 68, 156, 85]: "Virtual world navigators may wander aimlessly when attempting to find a place for the first time. They may then have difficulty relocating places recently visited. They are often unable to grasp the overall topological structure of the space" [68, p. 166].

To combat these difficulties, previous work has also looked at a variety of navigational aids to improve navigation efficiency. The value of landmarks has led researchers to consider the idea of allowing users to place visual markers, having the system create a visual trail showing where users have been, or having a fixed marker to provide a consistent indication of north [68, 67]. Results with these forms of assistance are mixed, however: adding a simple compass did not substantially improve navigation performance [65], and trails can quickly clutter an environment. Designers of other virtual environments, such as digital games, have created a variety of navigation aids for users. These are discussed in Section 6.3.2 and the effects of such assists on spatial learning is discussed in Section 6.3.2.

Incidental versus Intentional Spatial Learning

A continuing debate concerns the relationship between spatial knowledge acquisition and intentionality. Studies indicate that at least some aspects of location learning occur automatically [11, 126]. For example, one study showed that recall of word locations was unaffected by the difficulty of a concurrent task [11]. Other work, however, shows the importance of intention; studies have shown that when people focused their attention on a route through a building, they were better able to draw a map of that path [320], and that even a long experience with an environment may still result in poor survey knowledge [50]. In particular, passive observation of the environment can allow one to acquire route knowledge, though survey knowledge seems to require more intentional effort [320, 55].

How the Design of Virtual Environments Affects Navigation Difficulty

Virtual environments are generally considered to be more difficult to navigate than physical-world environments [245, 68, 156, 328]. This is due to the reduced interface fidelity (lack of kinesthetic feedback, and a reduced field of view) and environment fidelity of virtual environments (lack of visual detail that can be used as landmarks along with a lack of non-visual sensory information) [328]. Designers can therefore make virtual environments easier to navigate by increasing the visual fidelity of the environment.

Three additional factors affect navigation difficulty. First, the most significant is the *size* of the environment: environments that are small with minimal opportunity for exploration will be easier to navigate than environments that are large and complex [208, 66]. A second factor is *density*: a sparsely populated world has fewer objects of interest to leverage in navigation [66]. Third, *activity* is also an important factor: an environment where the position of objects changes over time is more difficult to navigate than if all objects remain stationary [66].

Navigation Assists in Virtual Environments

Guidance within virtual environments is provided in a variety of ways. Researchers have identified a variety of navigation assists that can be found within virtual environments (in digital games in particular) [188, 208]. Moura and El-Nasr categorize navigation assists as being either directional signs, identification signs, or orientation signs, with some assists fitting into multiple categories [208]. First, directional signs inform players where to go and what to do — for example, compasses, maps, GPS (maps which show one's location), arrows, or markers (see Figure 6.1 for examples from games). Second, identification signs indicate to players when they have reached their destinations — for example, markers, signs, or GPS (see the quest markers from *Skyrim* [29] in Figure 6.1 and *Diablo 2 Resurrected* [32] in Figure 6.3). Third, orientation signs inform users of their relative position within the environment — for example, maps that show the user's location.

Navigation assists can also be categorized as being presented separately from the environment or being situated within the environment. For those that are presented separately from the environment, some of the most common are maps, compasses, GPS, and arrows. Maps can be brought up via a menu or hotkey and can be shown full-screen or can be continually visible in the form of mini-maps [188]. Maps are 2D representations of the 3D environment and are often augmented with other useful information, such as the user's current location (i.e., a GPS), as well as the

Figure 6.1: A variety of different types of navigation assistance found in virtual environments and games. This includes arrows, compasses, mini-maps, indicators, quest markers, and visual highlighting. From left to right, screenshots are from *Skyrim*, *Midtown Madness*, *Metal Slug*, *World of Warcraft*, and *Wolfenstein: Enemy Territory*.

Figure 6.2: Ways that navigation assistance can be included directly within a virtual environment itself. From left to right, screenshots are from *Half-Life*, *Counter-Strike: Global Offensive*, and *Morrowind*.

locations of objectives, items, or other users. In some games, the map is revealed as a player explores the environment; players then know which areas of the environment they have or have not explored [188]. Compasses are sometimes shown separately (as in *Skyrim* [29], Figure 6.1) from the map or alongside the map (as in *World of Warcraft* [31] or *Wolfenstein: Enemy Territory* [280], Figure 6.1). These compasses show the user their current direction of travel and can also be augmented with additional information such as the direction of objectives. Arrows are also commonly placed within the environment to indicate the required direction of travel. For example, as part of the game's heads-up-display, *Midtown Madness 2* [13] provides players with a yellow arrow that indicates the direction the player needs to travel to reach the objective (see Figure 6.1). As a simpler example, *Metal Slug* [212] prompts the player to move to the right at certain times to make progress within the game (see Figure 6.1).

Navigation assists that exist within the environment itself are commonly found as indicators, arrows, trails, and highlights. For example, the same indicators that appear on many maps or compasses also appear within the environment itself: e.g., the quest markers in *Skyrim* and *Diablo 2 Resurrected* (see Figures 6.1 and 6.3). Arrows and signs are often placed into the environment itself as they would appear in the physical world (see Figure 6.2). Many games

Figure 6.3: Examples of how the objects within the environment can be highlighted to indicate the direction of travel or important objects within the environment. From left to right, screenshots are from *Mirror's Edge*, *Left 4 Dead 2*, *Neverwinter Nights*, and *Diablo 2 Resurrected*.

Figure 6.4: Two examples of trail guidance, from *Fable 2* (left) and *Skyrim* (right).

also make use of visual highlighting to direct a player's attention to important navigational information. For example, *Mirror's Edge* [76] uses the colour red to indicate to the player which objects they need to climb or interact with next (Figure 6.3). Finally, some games show players a visible trail within the environment to follow to reach their destination. For example, *Fable II* [179] and *Neverwinter* [62] have particle trails which can be turned on or off through the user interface (Figure 6.4). Sometimes this is part of a gameplay mechanic and provided only temporarily; in *Skyrim*, magic users have access to a "clairvoyance" spell that temporarily reveals the exact route to a quest marker with a smoke trail (Figure 6.4).

One final type of navigation assistance, commonly used in 3D games, is the guided tour [208, 188]. Guided tours are often presented as scenes that walk a user through the environment or direct a user's attention to a specific object, providing a demonstration of how to navigate through the environment.

Effects of Navigation Assists on Spatial Learning

Navigation assists can affect spatial learning in different ways. [161] suggest that assistance that reduces the user's mental effort interferes with spatial learning [161, 59]. Therefore, reducing decisions during navigation can negatively affect spatial learning [161, 24]. Other work has found that learners were less able to remember landmarks if navigation aid was provided [109], although the authors attribute this effect to the navigation aid dividing the learner's attention. Similarly, [320] found that having a learner follow another person impaired their survey knowledge formation, although learners still acquired route knowledge. Khan and colleagues claim that maps in particular interfere with spatial learning because they place a cognitive load on the learner [161], who must perform mental rotations to make use of the map (as described by [65]). However, other researchers have pointed out that the map is a source of information that can provide a learner with survey knowledge [67]. Recent research has also looked at the effects of guidance systems such as GPS and has found that people can become overly focused on the directions provided by external guidance, hindering the development of their spatial knowledge [e.g., 42, 146, 176].

6.4 Materials and Methods for Both Studies

We conducted two online studies[2] to answer our research questions relating to the efficacy of navigation guidance for navigation performance and user experience (when assistance is present) and the potential for reduced spatial learning due to over-reliance on the assist (when the assistance is removed). We designed and implemented a system that allowed online participants to navigate virtual environments between specific start and end points (i.e., routes). The system could vary the amount of navigation assistance provided to the participant, as described below.

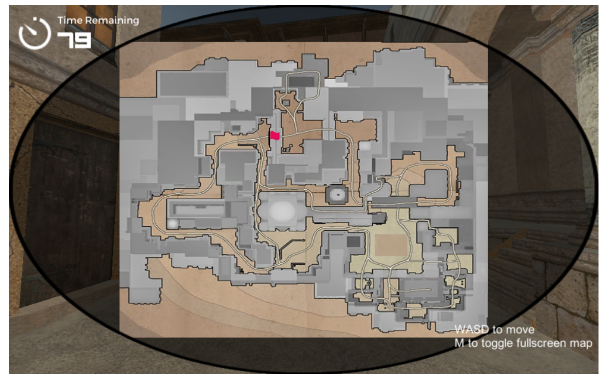
Both studies involved a set of phases: a tutorial, in which participants were introduced to the environment and the route-finding tasks; a training phase, in which participants carried out tasks with navigation assistance; a transfer test, in which participants navigated routes with the assist removed; and (in Study 2 only) a retention test, in which participants navigated routes without assistance, but one week after their final training.

6.4.1 Virtual Environment

We used a virtual environment that was extracted from the commercial game *Wolfenstein: Enemy Territory*. The game's source code and tools are freely available, so we downloaded the game level and edited it using the GtkRadiant application [145] to remove all unneeded objects relating to game logic, leaving only the level's geometry and textures. The compiled level was then converted into a standard 3D model that was imported into the Unity game engine [317]. Other 3D assets from the game were placed by hand within the environment (such as a tank and a truck) to be used as landmarks. We implemented first-person movement and view controls in Unity to match what is seen in typical games (WASD movement and mouse-based view control); we also created custom implementations of the different types of navigation assistance and added experiment infrastructure to present a set of routes and navigation tasks to the user.

[2]Portions of Study 1 were reported in a CHI Play 2017 paper [?]. The version here presents an expanded and revised analysis of the data from the virtual environment that matches the one used in Study 2, to better allow comparison across the two studies.

(a) The interface with map assistance. No mini-map is provided.



(b) The full-screen map with the map assistance version of the interface. No mini-map is provided, and the map does not show the user's current location.

(c) The interface with position assistance. At top right is the minimap that shows where the user is located (as a yellow circle) within the environment in relation to the flag. Users could also open the full-screen map, which also showed their position.

(d) The interface with trail assistance, showing the white trail that users could follow to be taken to the destination. This also the interface from the position assistance version.

Figure 6.5: Interfaces used for the three conditions of Study 1's training tasks.

Figure 6.6: The first-person view of the "Gold Rush" environment from *Wolfenstein: Enemy Territory*.

We used the "Gold Rush" level (Figures 6.6 and 6.7) which presents a fictional town in northern Africa, with most of its routes located outside. The town has a variety of streets, walls, buildings, passages, plazas, and staircases. There are several naturalistic landmarks such as palm trees in a town square, vehicles including carts and tanks, and multi-story towers. This level was chosen because it had an adequate level of complexity with multiple possible routes to each destination.

A few details of the level were changed between Study 1 and Study 2. Because Study 2 did not provide a map, it was possible for participants in the no-assist condition to become lost — therefore, we slightly modified the environment for Study 2 to make two areas without landmarks inaccessible, reducing the overall chance of participants being unable to find the target destination.

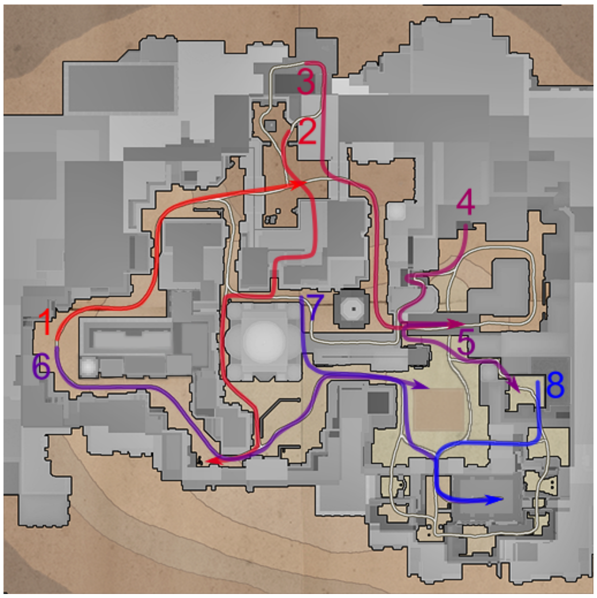
6.4.2 Types of Navigation Assistance

Both studies introduced three levels of navigation assistance that were provided to participants during the training phase of the experiment.

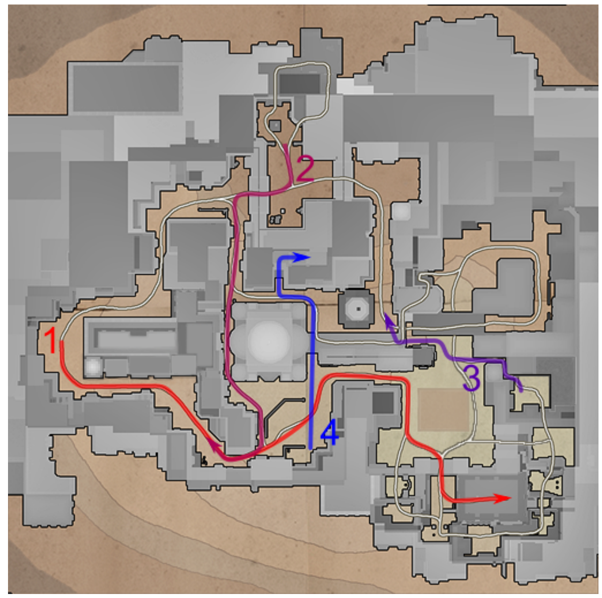
Study 1

We designed three types of navigation assistance that varied in the amount of required navigation effort (see Figure 6.5).

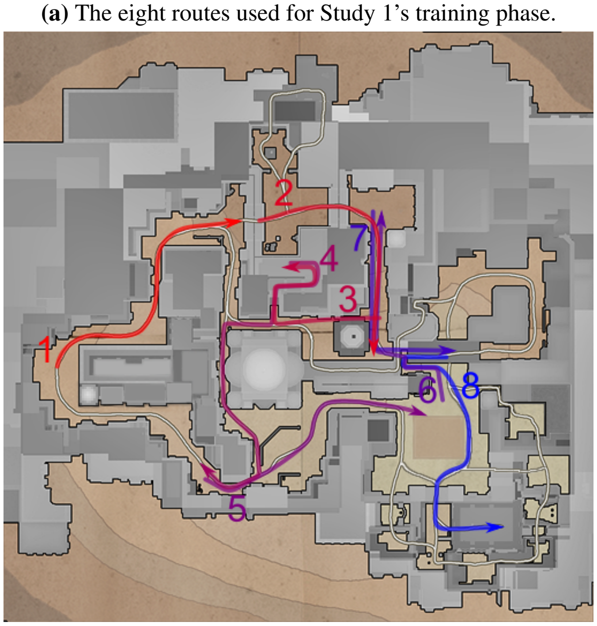
Map Assistance. With map assistance, the participant had access to a full-screen pop-up map (invoked with the M key) that showed the target destination (see Figures 6.5a and 6.5b). In this condition, participants had to identify their



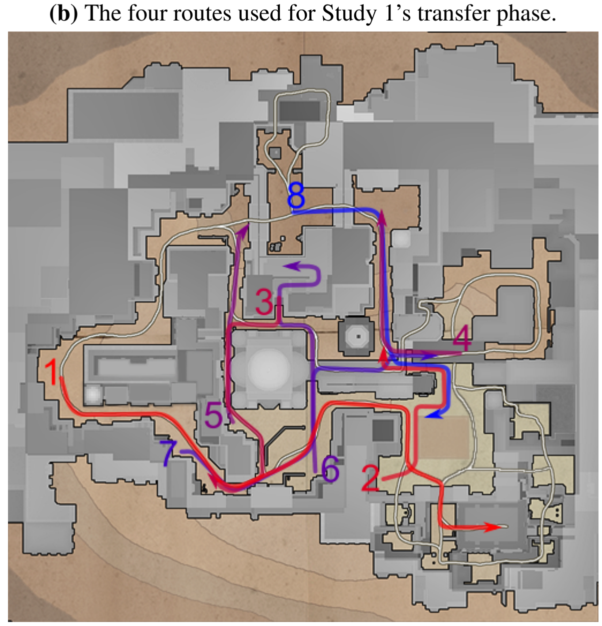
(a) The eight routes used for Study 1's training phase.



(b) The four routes used for Study 1's transfer phase.



(c) The eight routes used for Study 2's training and retention phases.



(d) The eight routes used for Study 2's transfer phase.

Figure 6.7: Overhead views of the "Gold Rush" environment, showing the routes used in Study 1 and 2.

own position on the map, plan a route to the destination, and translate directions and distances from the map view to the first-person environment.

Position Assistance. With position assistance, the same map as described above was also available, but with the participant's current position on the map now marked (see Figure 6.5c). Additionally, the interface included a mini-map in the top right corner of the screen. Similar to mini-maps found in other games [355], the mini-map was circular, player-centred, and used a north-up orthographic projection of the environment. Both maps included an icon indicating the participant's current location and direction (similar to the user icon used in Google Maps). In this condition, participants could see their dynamic progress on the map views — and if they navigated solely by focusing on the map, there was less of a requirement to translate information to the first-person view.

Trail Assistance. With trail assistance, the interface additionally showed the path to the destination as a solid white line drawn in the 3D environment (see Figure 6.5d). The two maps described above were also available: these showed the participant's location (but did not show the trail). The trail line was a guide only, and participants could take any route they wanted to the destination. The trail visual effect is similar to the navigational assists used in several commercial games, as discussed above. In this condition, participants had to expend far less effort than with the other interfaces — they did not have to identify their location or plan a route, and could simply follow the trail to the destination.

Study 2

For Study 2, we developed three different types of navigation assistance to explore a greater range of required navigational effort (see Figure 6.8).

No Assistance. In this condition, participants navigated using only the first-person view: no maps or visual guides were provided (see Figure 6.8a). Participants were given a screenshot indicating the landmark to navigate to and had to find it on their own — this condition represents the maximum navigational effort in the study, as participants had to develop a spatial understanding of the environment using only the first-person view.

Trail Assistance. With trail assistance, the route that the participant was intended to take was indicated using a glowing white within the first-person environment (see Figure 6.8b). This condition was similar to the one used in Study 1 (except that no maps were available in Study 2).

Rail Assistance. With rail assistance, when the participant held down a key, they would move forward towards the destination (i.e., the participant was "on rails"), using the same route as would be indicated with trail assistance (see Figure 6.8c). Participants could look around in any direction while moving and still move in the correct direction. No trail was shown, and no maps were available.

6.4.3 Navigation Tasks

Both studies had participants complete a series of navigation tasks. In each task, the participant was placed at a starting location in the environment and had to navigate to a target destination (either a red flag on a pole or an obvious landmark such as a tank or a truck). The target destination was communicated to the participant either using the overview map

(a) The interface used when there was no assistance.

(b) The interface showing the trail assistance. The white trail indicates the route to the destination.

(c) The interface showing the rail assistance. Nothing additional is shown in the game, but the shown controls are different (participants could only press the W key).

Figure 6.8: Interfaces for the three conditions in Study 2's training tasks.

(a) The interface used for Study 1's transfer phase. No assistance was provided to the user, not even the map.

(b) The interface used for Study 2's transfer and retention phases. No assistance was provided to the user (equivalent to the no-assist condition).

Figure 6.9: Interfaces used for transfer tasks (Study 1 and Study 2) and retention tasks (Study 2). In these tasks, users were asked to navigate to a landmark shown in the top-right corner of the screen.

(in Study 1 training tasks) or by showing the user a picture of the target landmark. Participants moved through the environment using the WASD keys to move forward, left, backward, or right (except for the rails assistance condition in Study 2, in which participants could only press the W key to move along the rail). Participants could look around the environment by moving the mouse.

Study 1

Study 1 had three different versions of the navigation task: a tutorial version, a training version, and a transfer version. In the *tutorial* version, participants had no navigation assistance and the route used a linear path — there were no decisions to be made regarding the direction of travel. The tutorial introduced participants to the controls and allowed them to ensure that their computer system would perform well enough to handle the rest of the study.

Tasks in the *training* phase (shown in Figure 6.5) took place over three days, with participants attempting a fixed set of eight routes each day, and with assistance depending on the condition that participants were assigned to (one of map, position, or trail assistance). All participants had access to the full-screen map (accessed by pressing the M key); the map showed the current target destination as a red flag icon. For each route, participants travelled between predefined start and destination locations. A 90-second time limit was given for each route to ensure that the participant could make progress in the study (although typical times to traverse a single route ranged from 5-25 seconds). The eight routes were the same on each day and were presented in the same order.

Tasks in the *transfer* phase (shown in Figure 6.9a) took place after the final training phase; participants were asked to complete an additional four routes without any assistance, not even a map. The four routes were different than those used in training, and instead of displaying the target destination on the map, an image of a target landmark (e.g., a truck) was displayed and participants were instructed to travel directly to that landmark. The landmarks were objects that participants had seen during training (e.g., they were beside the flags used in training, or they were on a required route); however, the routes differed because they started at different locations. There was no prior indication during the study that participants would be tested on their ability to navigate to these landmarks; we hypothesized that participants could acquire spatial information incidentally as they were navigating the training routes.

Study 2

Study 2 had four different versions of the navigation task: a tutorial version, a training version, a transfer version, and a retention version. The first three were similar to the equivalent versions in Study 1, so only the differences will be described here.

The *tutorial* was similar to Study 1, except that the navigation assistance for the assigned condition was also present in the tutorial because the controls for the rail assistance condition were slightly different (only the W key was used for movement instead of the WASD keys).

The *training* phase took place over four days instead of three. Three different types of assistance were used (no assist, trail assist, or rail assist) and participants were instructed to navigate to a landmark (permanently shown in the corner of their screen) rather than a point on a map.

The *transfer* phase (Figure 6.9b) took place one day after the last day of training instead of on the last training day; participants completed eight new routes instead of the four used in Study 1. The routes for the transfer phase had new starting locations, but the landmarks used as destinations were all objects that participants had encountered during training. As in Study 1, participants completed these routes with no assistance.

The *retention* phase took place one week after the transfer phase and asked participants to complete the same routes as they had trained on but without assistance. Retention tasks were only used in Study 2.

6.4.4 Procedure

Both studies were deployed on a custom website built using an existing web framework designed to aid the creation of online studies [150]. This website presented the questionnaires as HTML forms and embedded the game directly into the web browser using WebGL. Upon opening the website, participants would be asked to read a consent form and provide informed consent before being directed to the questionnaires and navigation tasks. Because both studies took place over several days, participants were invited back via an email (sent anonymously via Amazon's Mechanical Turk API) at the start of each day.

Study 1

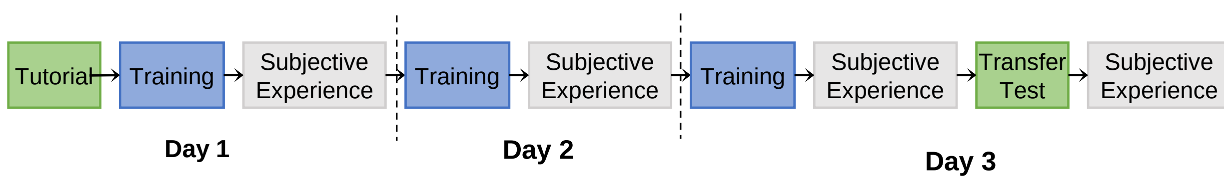


Figure 6.10: Procedure for Study 1. Green and blue boxes indicate navigation tasks (blue: assist present; green: no assist). Grey boxes indicate questionnaires.

The overall procedure for Study 1 is diagrammed in Figure 6.10. Participants first completed the tutorial version of the navigation task and then completed a questionnaire related to their gaming experience. This was presented as a separate qualification task that participants needed to complete to be eligible for participation in the study. Participants who had an adequate framerate (over 45 frames per second) and stated that they were "not at all" or only "slightly" experienced with the Gold Rush environment were invited to complete the rest of the study.

Participants who accepted this invitation were assigned to one of the types of assistance and then completed demographics and individual-differences questionnaires. They then began the training phase where they navigated eight training routes. After the eight routes, they answered questions relating to subjective experience. On the next day, participants completed the training again and responded to the same questions about subjective experience. This was repeated a third time on the final day, followed by the transfer version of the navigation task and a final round of the same subjective experience questions.

Study 2

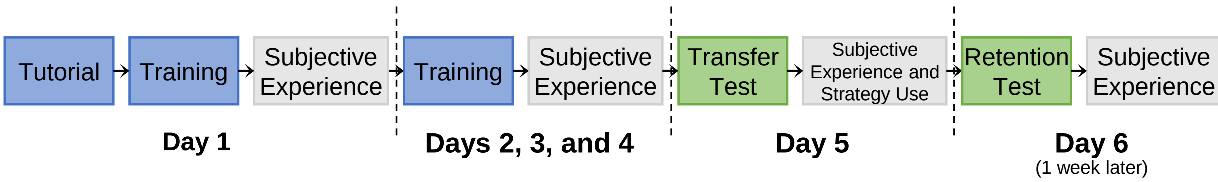


Figure 6.11: Procedure for Study 2. Green and blue boxes indicate navigation tasks (blue: assist present; green: no assist). Grey boxes indicate questionnaires.

The overall procedure for Study 2 is diagrammed in Figure 6.11. Participants completed questionnaires relating to demographics and individual differences before carrying out the tutorial task. Participants then began the training tasks; after the routes were completed, they responded to questions relating to subjective experience. The training procedure was the same on days two, three, and four. As in Study 1, participants were invited back each day using email reminders.

On the fifth day, participants completed the transfer version of the navigation task, answered the same questions relating to subjective experience, and also answered two additional questions relating to navigation strategy. One week later, participants completed the retention version of the navigation task and the same questions about subjective experience.

6.4.5 Measures

Both studies measured aspects of individual differences that might affect one's ability to navigate an environment, navigation performance outcomes, and subjective experience outcomes.

Study 1

The following measures of individual differences were collected at the start of the study:

- **Gaming Expertise.** We asked participants questions to establish their gaming expertise: how much they self-identified as a gamer, their experience with video games, their experience with keyboard-and-mouse input in games, their FPS (first-person shooter) experience, and their experience with 3D games. These questions were included because prior experience navigating virtual environments in games could affect how (and how well) participants navigated during the study [41].
- **Immersive Tendencies.** We used the Immersive Tendencies Questionnaire (ITQ) [346] to measure participants' tendency to experience presence in virtual environments. The questionnaire consists of three subscales: involvement (propensity to get involved with an activity), focus (ability to concentrate on enjoyable activities), and games (how much they play games and whether they become involved enough to feel like they are inside the game). These questions were included because the sense of presence within a virtual environment can affect task performance within that environment [346].

- **Wayfinding Anxiety.** We measured each participant's trait anxiety and tendency to use a "route-learning" strategy or an "orientation" strategy using Lawton and Kallai's [172] International Wayfinding Anxiety Scale and International Wayfinding Strategy Scale, respectively. These questions were included because wayfinding anxiety can affect navigation performance [177], and different strategy use (e.g., a reliance on landmarks or a tendency to navigate using cardinal directions) can affect wayfinding efficiency [142].

Route-finding performance was measured in two ways:

- **Completion Time**. The system recorded each participant's total time to complete the eight training routes, the four transfer routes, and the eight retention routes (Study 2 only). The maximum time per route was 90 seconds.
- **Distance Travelled**. The system recorded the total 3D Euclidean distance travelled by the participant for each route (using Unity's default measuring system). A greater distance indicates that the participant made more errors while navigating.

Subjective experience was measured after completing each day's training, and again after the transfer session. These questions relate to RQ1:

- **NASA Task-Load Index (TLX)** [125]. The NASA Task-Load Index questionnaire is a widely-used [181] questionnaire to rate perceived workload when completing a task. We used the questionnaire's mental demand, performance, effort, and frustration questions.
- **Perceived Map Knowledge**. To measure each participant's perceived map knowledge after training, we asked them to rate their knowledge of the layout of the map, on a 5-point scale from "very poor" to "very good".

Study 2

Study 2 also measured individual differences. The **Immersive Tendencies** and **Wayfinding Anxiety** questionnaires were the same as what was used in Study 1, but other questionnaires were added or revised:

- **Experience with First-Person 3D Games.** Our gaming expertise questions from Study 1 were expanded into a scale with 5 questions. The scale used a 5-point Likert scale from "Not at all" to "Extremely", and the questions included: "Are you a gamer?", "Are you experienced at playing video games?", "Are you experienced with using keyboard and mouse input simultaneously to control games?", "Are you familiar with navigating 3D virtual environments?", and "Are you experienced at playing first-person shooter games?".
- **Spatial Ability.** We also included the Spatial Ability Self-Report Scale [304]. This consists of three subscales: Object-Manipulation Spatial Ability (OMSA), Spatial Navigation Ability (SNA), and Visual Memory (VM). Object-Manipulation Spatial Ability involves the ability to mentally rotate or fold objects and the ability to visualize spatial relationships. Spatial Navigation Ability involves the ability to form a mental map of the environment and navigate within it. Visual Memory involves the ability to notice and remember differences in visual stimuli. These questions relate to participants' ability to develop spatial understanding of the environment (RQ2).

- **Intrinsic Motivation.** We used the Intrinsic Motivation Inventory (IMI) [193] to evaluate participants' intrinsic motivation toward the tasks. This inventory measures four dimensions: Interest-Enjoyment, Perceived Competence, Effort-Importance, and Tension-Pressure. These were included because intrinsic motivation toward the task may affect one's interest in continuing to engage with the task. These questions relate to RQ1.

Additionally, we prompted users to answer questions relating to any strategies they used to help them learn about their surroundings:

- "Did you make use of any intentional strategies to remember the specific routes you had trained on previously? ('Yes' or 'No')"
- "Did you make use of any intentional strategies to remember the locations of the landmarks? ('Yes' or 'No')"

6.4.6 Participants

In both studies, participants were recruited through Amazon's Mechanical Turk (MTurk) crowdsourcing platform. MTurk connects willing workers to paid Human Intelligence Tasks (HITs). Ethical approval for the studies was obtained from the behavioural ethics board of the University of Saskatchewan, and participants were asked to renew their consent at the start of each day's task. To comply with ethical guidelines, the task was only available to workers from the United States who were over 18 years old.

Study 1

Participants were first recruited through a HIT limited to 100 people that involved completing a simple navigation task presented as a tutorial and filling out a demographics questionnaire. The task took 5.5 minutes on average (SD=2.1) and paid $0.50 USD.

During the tutorial task, the participant's in-environment framerate was logged and they were asked about their prior experience with our chosen virtual environment. To be eligible for the study, participants needed to be "not at all" or only "slightly" experienced with the Gold Rush environment and our chosen game. Additionally, they needed to have had a framerate higher than 45 frames per second, otherwise, their system may not have performed adequately during the navigation tasks in the study.

We invited back 73 people to complete the full study, with only 50 spots available. On the first day, participants were randomly assigned to one of the three assistance groups, completed initial questionnaires, and a sixteen-route[3] training session. Participants were paid $3.50 USD for completing the first day, which took 21.7 minutes on average (SD=8.1). The second day consisted only of the same sixteen-route training session, and participants were paid $3 USD and it took 14.8 minutes on average (SD=10.0). The final day consisted of the final sixteen-route training session

[3] Eight of the training routes were for a game environment not presented in this work, but in a prior publication. Similarly, four of the transfer routes are not presented here.

and the eight-route transfer session. Participants were paid \$4.50 USD for day three, which took 30.0 minutes on average (SD=10.2).

Of the 50 who started the multi-day study, 46 completed all three days. We excluded two participants from our analysis due to logging errors, leaving us with 44 participants (29 male, 15 female, mean age of 33.7, SD=8.68; min=20; max=59). All participants were randomly assigned to one of the three assistance groups, balancing for self-declared gender: 14 people (5 female, 9 male) received map assistance, 16 people (6 female, 10 male) received position assistance, and 14 people (4 female, 10 male) received trail assistance.

Study 2

Participants were recruited in a slightly different way in Study 2; eligibility was determined by having participants first complete a very brief qualification HIT to confirm that they had little to no experience with the game we selected. A total of 500 participants completed this HIT, which asked three questions and paid \$0.05 USD. Two of the three questions were used to mask our intention, which was to only invite back participants who indicated that they have "no experience" with playing *Wolfenstein: Enemy Territory*, the game that our 3D environment comes from.

Based on the qualification HIT, we invited back participants to complete the full 6-day study. 136 participants completed the first day of the study, which paid \$4 USD and took 21.0 minutes on average (SD=8.9), with 88 of these also completing the next three days of training, which each paid \$2 USD and took 10.2 minutes on average (SD=6.2) and the transfer day, which paid \$2 USD and took 10.0 minutes on average (SD=6.5). Of those, 78 completed the retention day as well, which paid \$3 USD and took 15.8 minutes on average (SD=10.3). Eight participants were removed from the analyses due to a low framerate during the task (less than 30 frames per second). This left 80 participants who completed the training and transfer task and 70 participants who completed all tasks including the retention task.

Due to some participants dropping out of the experiment over the multiple days, we were left with an unequal distribution of participants in assistance groups. For the 80 participants that completed everything except for the retention task, 24 received no assistance (10 female, 14 male), 28 received trail assistance (15 female, 13 male), and 27 received rail assistance (12 female, 15 male). The average age of the participants was 38 years (Min 22, Max 70, SD 10.7). Of the 70 participants that completed every task, 22 received No assistance (9 female, 13 male), 24 received trail assistance (13 female, 11 male), and 24 received rail assistance (10 female, 14 male). The average age was 37.8 (Min=22, Max=70, SD=10.8). There were no non-binary participants.

6.4.7 Data Analyses

Study 1

To explore differences in training due to assistance, we performed separate repeated-measures analysis of covariance (RM-ANCOVA) tests for each of our outcome measures (completion time, distance travelled, task-load index, and perceived map knowledge). The day of the training session (1, 2, and 3) was used as the within-subjects factor, and assistance type was used as the between-subjects factor. For the transfer session, separate ANCOVA tests were used for the same outcome measures (no within-subject factor was used).

Individual differences between participants in terms of navigation anxiety, wayfinding strategies, gaming expertise, and immersive tendencies were considered as potential covariates for our statistical tests, based on whether those traits significantly correlated with the dependent measures. For the performance measures, the following covariates were included: ITQ's games subscale, wayfinding anxiety, wayfinding orientation strategy, and gaming expertise. For the subjective measures, the following covariates were included: involvement, wayfinding anxiety, and gaming expertise. Questions are shown in Table 6.1.

The individual differences measures were also used to verify that there were no differences between the groups as a result of random assignment. One-way ANOVAs with these measures as dependent variables showed no significant differences between the assistance groups. Alpha was set at 0.05, all covariates were mean-centred [38, 259], degrees of freedom for within-subject effects were corrected with Huynh-Feldt estimates of sphericity [102]. We performed post-hoc pairwise comparisons following each of the RM-ANCOVAs for the training session's outcome measures and following each ANCOVA for the transfer session's outcome measures. All effect sizes were estimated using partial eta squared.

Study 2

For our measures of performance (Completion Time and Distance Travelled), we used separate RM-ANCOVAs for the Training session, with Day as a within-subjects factor and Assistance as a between-subjects factor. For these measures during our Transfer and Retention sessions, we used separate ANCOVAs.

For the measures of subjective experience, measurements from each day of the Training session were included in separate RM-ANCOVAs (one for each measure) and we report results of the between-subject effects only. Additionally, we used separate ANCOVAs for each subjective measure for the Transfer and Retention sessions.

Individual differences between participants (adding object manipulation ability, spatial navigation ability, and visual memory in addition to those used in Study 1) were considered as potential covariates for our statistical tests, based on whether those traits significantly correlated with the dependent measures. For the performance measures, we used the following covariates: Gaming Experience, Games (from ITQ), and Visual Memory. For the subjective experience measures, we used Gaming Experience, Involvement, Games (from ITQ), Focus, Wayfinding Anxiety, Orientation Strategy, Visual Memory, Object Manipulation, and Spatial Navigation as covariates.

The individual differences measures were additionally used to verify that there were no differences between the groups as a result of random assignment. One-way ANOVAs with these measures as dependent variables showed no significant differences between the assistance groups. Alpha was set at 0.05, all covariates were mean-centred [38, 259], degrees of freedom for within-subject effects were corrected with Huynh-Feldt estimates of sphericity [102]. We performed post-hoc pairwise comparisons following each of the RM-ANCOVAs for the training session's outcome measures and following each ANCOVA for the transfer and retention session's outcome measures. All effect sizes were estimated using partial eta squared.

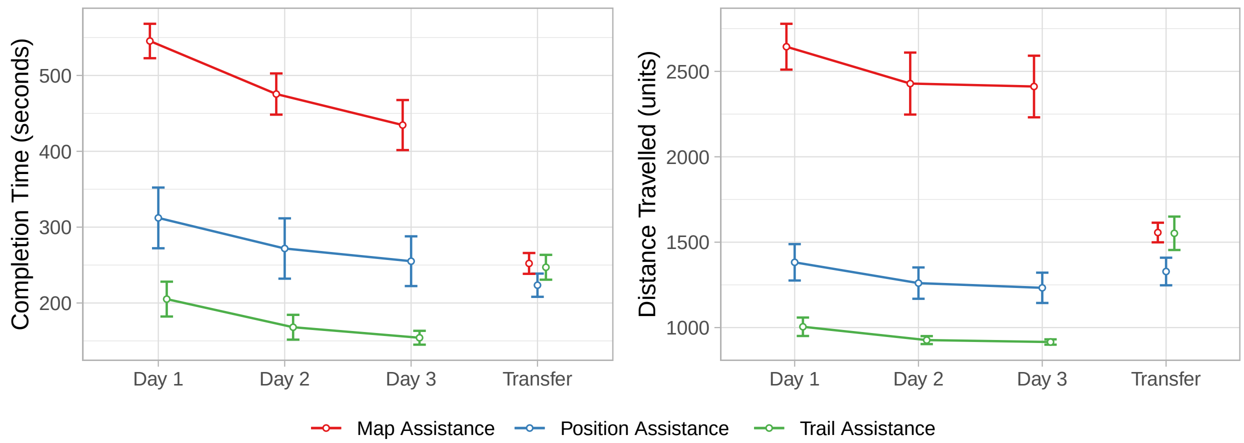


Figure 6.12: Performance during Training and Transfer sessions for Study 1. Error bars show standard error.

6.5 Results

6.5.1 Study 1

Training Tasks: Did Assistance Improve Navigation Performance? (RQ1)

We expected that navigation assistance would help participants' performance when it was present during the Training session, and this is what we found. There was a significant main effect of Assistance on Completion Time ($p < .001$; see Figure 6.12 for descriptive results and Table 6.3 for the results of our statistical analyses). Pairwise comparisons showed that the Map group performed worse than the Position and Trail groups (both $p < .001$), but the Position and Trail groups did not show a difference ($p = .149$). The differences were large: participants in the Map condition took approximately 210 seconds longer to complete the routes on average each day compared to the Position participants, and approximately 275 seconds longer than the Trail group. The results were similar for Distance Travelled: there was a significant main effect of Assistance ($p < .001$), and pairwise comparisons showed that the Map group travelled further than either the Position group (1200 units more, $p < .001$) or the Trail group (1500 units more, $p < .001$).

Training Tasks: Did Assistance Improve Subjective Experience? (RQ1)

During Training, with Assistance present, we found significant main effects of Assistance (see Table 6.3 for full the results of our statistical analyses, and Figure 6.13 for descriptive results) on Effort ($p = .006$), Frustration ($p < .001$), Perceived Performance ($p < .001$), and Mental Demand ($p < .001$), but not Perceived Map Knowledge ($p = .056$). Introducing Position or Trail assistance led to reductions in Frustration ($p < .001$) and Mental Demand ($p < .001$), and an increase in Perceived Performance ($p \geq .029$) over just Map assistance. Introducing Trail assistance led to a reduction of effort compared to just Map assistance ($p = .007$) but not compared to Position assistance ($p = .060$). There was little subjective difference between Position and Trail assistance; only a reduction in Mental Demand ($p = .009$).

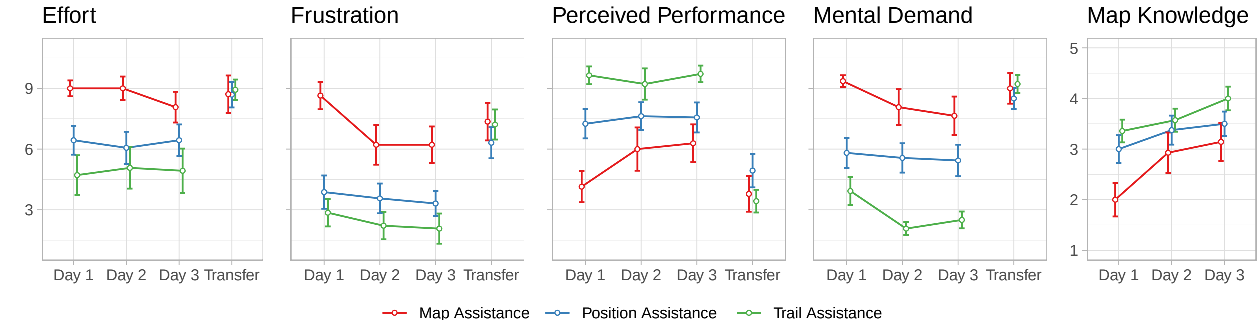


Figure 6.13: Descriptive results for the measures of subjective experience for Study 1.

Training Tasks: Did Participants Improve Over Time? (RQ2)

For some of our experimental conditions, the degree to which performance improves over time can be considered an approximate measure of spatial learning during the study. If performance did not improve across the days of the study, it may be an indication that the assist was playing a primary role in performance, rather than the participant's spatial knowledge. We note, however, that in the Trail condition, any change in performance may be reduced because navigation was tightly constrained by the assistance.

There was a main effect of Day on Completion Time for the training routes ($p < .001$; see Figure 6.12 and Table 6.2). Pairwise comparisons showed that there were significant improvements to Completion Time between Day 1 and 2 ($p < .001$), as well as Day 1 and 3 ($p < .001$), but not between Day 2 and 3 ($p = .064$). This improvement in Completion Time over the days was not affected by Assistance; there was no significant interaction between Day and Assistance.

We found similar results for Distance Travelled. There was a significant effect of Day on Completion Time ($p = .034$). No pairwise comparisons between Days were significant ($p \geq .105$). There was no significant interaction between Day and Assistance.

Transfer Phase: Was Performance or Experience Reduced with the Assist Removed? (RQ2)

Training with assistance could negatively affect the participants' ability to navigate the environment without assistance, but this was not the case in the study. Although we did find a significant main effect of Assistance on Completion Time for the Transfer session ($p = .037$; see Table 6.3), pairwise comparisons showed that the only significant difference was between Position and Trail assistance ($p = .039$). Results for Distance Travelled were similar, with a significant main effect of Assistance ($p .019$) and a significant pairwise comparison between Position and Trail ($p = .026$). We found no differences between the assistance conditions that required the most effort (Map assist) and the least effort (Trail assist), for either time or distance.

We also asked participants to complete the subjective experience questions after the Transfer tasks. We found no significant main effects of Assistance on Effort, Frustration, Perceived Performance or Mental Demand ($p \geq .162$).

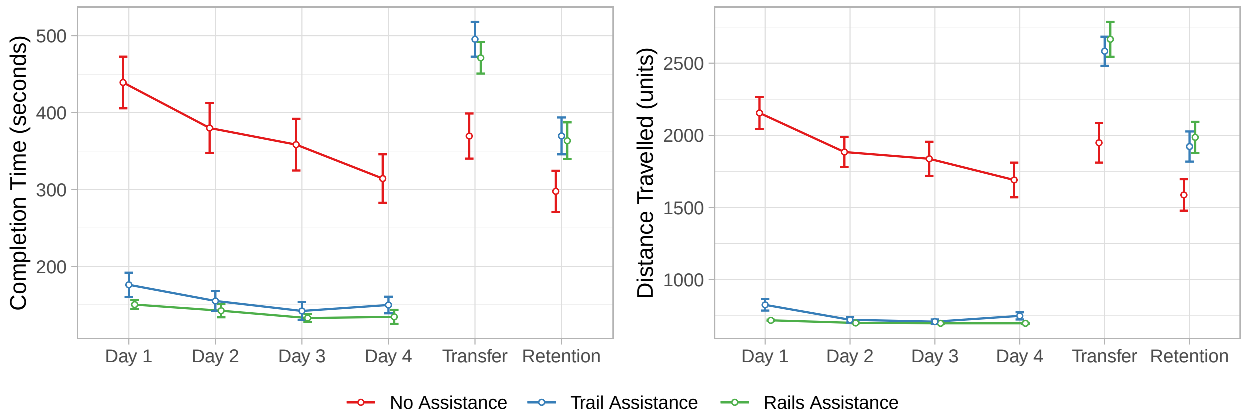


Figure 6.14: Performance during Training, Transfer and Retention sessions for Study 2. Error bars show standard error.

6.5.2 Study 2

Training Tasks: Did Assistance Improve Navigation Performance? (RQ1)

As in Study 1, we expected that navigation assistance would help participants when it was present (during Training). We found that Assistance reduced Completion Time and Distance Travelled (both $p < .001$; see Figure 6.14 and Table 6.4). Post-hoc tests showed that for both measures, there were significant reductions comparing Trail to No assistance ($p < .001$) and when comparing Rail to No assistance ($p < .001$). In both cases, the differences were large — participants in the No assistance condition took more than 220 seconds longer to complete the training routes on average each day and travelled 1100 units more, than either Trail or Rail participants. There were no differences between Trail and Rail assistance ($p > .999$).

Training Tasks: Did Assistance Improve Subjective Experience? (RQ1)

During Training, we found significant main effects of Assistance (see Table 6.4 for the results of our statistical analyses and Figure 6.15 for descriptive results) for Effort ($p = .004$), Frustration ($p = .004$), Perceived Performance ($p = .006$), Mental Demand ($p = .021$), Perceived Competence ($p = .015$), and Tension-Pressure ($p = .018$). Pairwise comparisons showed no differences between Trail and Rail assistance ($p \geq .811$). However, we find that when comparing to No Assistance, Trail assistance reduces Effort ($p = .012$), Frustration ($p = .015$), and Mental Demand ($p = .034$), while increasing Perceived Performance ($p = .021$). Comparing No assistance to Rail assistance, we find reductions in Effort ($p = .009$), Frustration ($p = .007$), Tension-Pressure ($p = .021$), and increases in Perceived Performance ($p == .009$) and Competence ($p = .021$).

Training Tasks: Did Participants Improve Over Time? (RQ2)

As in Study 1, we considered whether participants were learning the environment and improving over their task time during training. In Study 2, both the Trail and Rail conditions strongly dictate completion time (in Rails, for example,

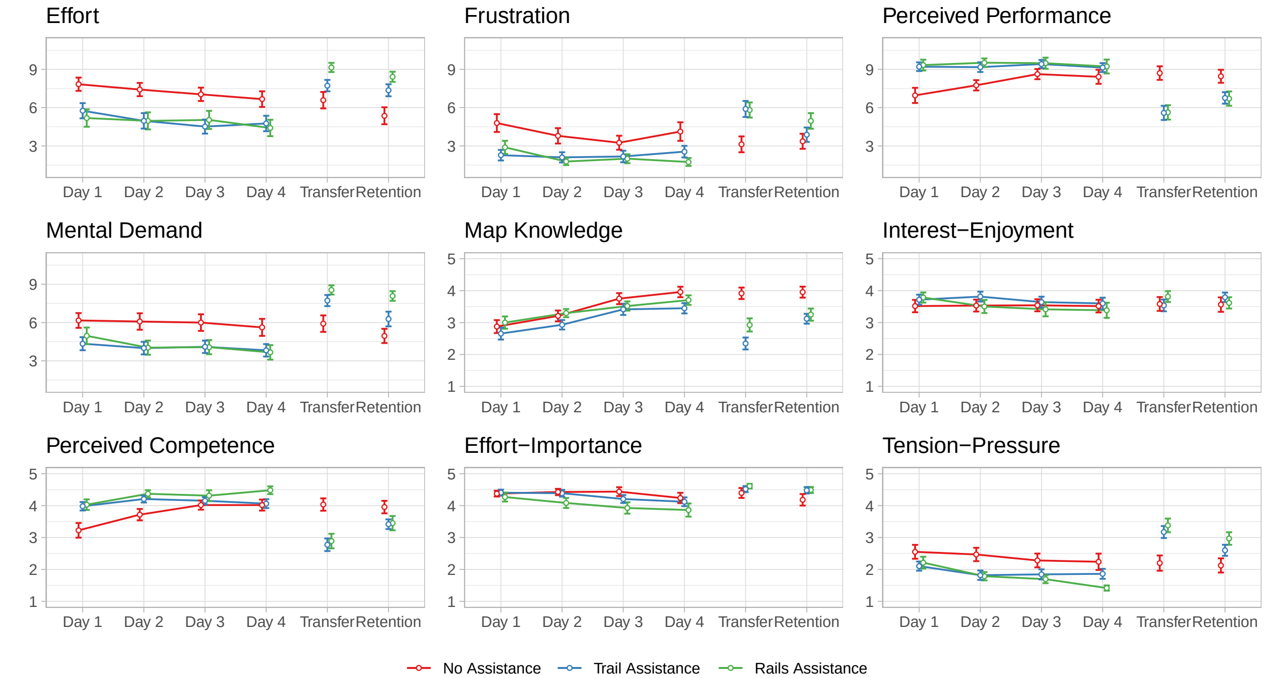


Figure 6.15: Descriptive results for the measures of subjective experience for Study 2.

task times will always be the same if the participant presses and holds the W key), and so we only expected to see learning effects in the No assistance condition.

We found that participants did improve over time, but that the amount of improvement depended on the assistance. There was a significant interaction between Day and Assistance for Completion Time and Distance Travelled ($p < .001$ for both; see Table 6.2).

To determine which of the groups were improving during training, we examined the pairwise comparisons to compare Day 1 and Day 4 performance for each Assistance group. For Completion Time, we found that the No assistance group improved over the training phase ($p < .001$), but Trails and Rails did not (both $p > .999$). For Distance travelled, results were similar: the No assistance group improved ($p < .001$), but the Trails and Rails groups did not ($p > .999$).

Transfer and Retention Tasks: Was Performance or Experience Reduced with the Assist Removed? (RQ2)

The effect of assistance on spatial learning was inferred by measuring performance in the Transfer and Retention tasks. On the Transfer task, where assistance was removed and participants were asked to navigate new routes, we found a significant main effect of Assistance on Completion Time and Distance Travelled ($p < .001$ for both; see Table 6.4). Post-hoc tests show that there were significant increases in Completion Time and Distance Travelled when comparing Trail to No assistance ($p \leq .002$), as well as when comparing Rail to No assistance ($p \leq .013$), but not when comparing Trail and Rail assistance ($p \geq .672$).

On the Retention task (also without assistance but on the routes participants had previously practiced during training), we did not find a significant main effect of Assistance on Completion Time ($p = .073$), but we did find an effect of Distance Travelled ($p = .042$; see Table 6.4). Post-hoc tests for Distance Travelled show only a difference between No assistance and Rail assistance ($p = .041$, $p \geq .258$ for the others), with the No assistance group travelling a shorter distance.

For subjective experience measures taken after the Transfer session, we found significant main effects of Assistance (see Table 6.4) for Effort ($p = .008$), Frustration ($p = .003$), Perceived Performance ($p < .001$), Mental Demand ($p = .003$), Perceived Map Knowledge ($p = .008$), Perceived Competence ($p < .001$), and Tension-Pressure ($p = .001$). As in Training, pairwise comparisons showed no differences between having trained with Rail assistance to having trained with Trail assistance ($p \geq .100$). When comparing No assistance to Trail assistance, we find increased Frustration ($p = .005$), Mental Demand ($p = .023$), Tension-Pressure ($p = .003$), along with decreased Perceived Performance ($p < .001$), Map Knowledge ($p < .001$), and Competence ($p < .001$). Comparing No assistance to Rail assistance, we find increased Effort ($p = .006$), Frustration ($p = .016$), Mental Demand ($p = .004$), and Tension-Pressure ($p = .001$), as well as decreased Perceived Performance ($p < .001$), Map Knowledge ($p = .002$), and Competence ($p < .001$).

After the Retention session, we found significant main effects of Assistance (see Table 6.4) for Effort (p), Perceived Performance ($p = .029$), Mental Demand ($p = .002$), Perceived Map Knowledge ($p = .008$), and Tension-Pressure ($p = .033$). Pairwise comparisons showed no differences between Rail and Trail assistance ($p \geq .147$). Comparing No assistance to Trail assistance, we find increased Effort ($p = .029$) and decreased Perceived Map Knowledge ($p = .014$). Comparing No assistance to Rail assistance, we find increased Effort ($p = .002$) and Mental Demand ($p = .001$) and decreased Perceived Map Knowledge ($p = .029$).

Did Participants Make Use of Intentional Strategies? (RQ2)

The differing level of effort for the three assistance techniques may have prompted users to employ different strategies for navigation. Based on questionnaire responses, we found that 35% of participants made use of intentional strategies to remember the locations of landmarks, and 22.5% used strategies to remember the routes taken. Comparing between the groups, participants in the No-assistance group were less likely to report using strategies for routes (16.7% for No assistance, 27.6% for Trail, and 22.2% for Rail). This pattern was reversed, however, for remembering landmarks (41.7% for No assistance, 34.5% for Trail, and 30% for Rail).

6.6 Discussion

6.6.1 Summary of Results

Our studies explored two research questions: first, will navigation assistance improve performance and user experience when it is present? and second, will navigation assistance hinder spatial learning and cause over-reliance on the assist? We investigated the first question by having participants undergo training with one of three levels of assistance (different levels were used in each study). In Study 1, we found that higher levels of navigation assistance (marking the user's position on a map or displaying a glowing trail to the destination) led to significant performance improvements compared to a map alone, both in terms of completion time and distance travelled — although there was little difference between the low-effort Trails condition and the medium-effort Position condition. In Study 2, which used assistance techniques with a wider range of requirements for navigation effort, we found a similar result. In training, the No assistance group performed significantly worse than either Trail or Rail assistance in terms of completion time and distance travelled (the two lowest-effort conditions, Trail and Rail assistance, performed similarly). Both studies also showed the benefit of navigation assistance on user experience. Study 1 showed reduced effort, frustration, and mental demand, as well as an increase in perceived performance, for both of the lower-effort conditions (Position and Trail) compared to the Map condition. Study 2 showed reduced frustration, mental demand, effort, and tension, as well as an increase in perceived competence and performance, for the lower-effort conditions (Trail and Rail) compared to the No-assist condition.

We investigated the second research question by having participants navigate the same environment again, but without assistance — completing transfer tasks (with new routes) and retention tasks (with the same routes as in training). In the transfer test of Study 1, we found that the performance of all three groups was similar once the assistance was removed. Although there was a difference between Position and Trail assistance, there were no significant differences between the least-effortful condition (Trail assistance) and the most-effortful condition (Map assistance). In Study 2 we carried out both a transfer test and a retention test. In transfer tasks (and unlike Study 1) we found that the No-assist group performed significantly better in terms of both completion time and distance travelled, with the two lower-effort techniques performing similarly. Experience measures were also reversed compared to Study 1, with the No-assist condition rated better than either the Trail or Rail conditions. In retention tasks, we found no significant difference in completion time regardless of which assistance participants had seen in training, but we did find an effect on distance travelled (participants who trained with no assistance travelled less during the retention tasks). Removing the assist in the transfer tasks also led to different effects on user experience, depending on the study. In Study 1, there were no differences between the conditions in the transfer task. In Study 2, removing the Trail and Rail assistance in the transfer tasks led to significantly lower scores on experiential measures compared to the No-assist condition (which was the same between training and transfer phases); similar results were found for the retention tests.

6.6.2 Explanation of Results

Why did Map assistance not result in better learning?

In Study 1's transfer test, where navigation assistance was taken away, we observed that the least-assisted group (Map assistance) performed similarly to the most-assisted group (Trail assistance), even though users had to expend considerably more effort when they only had a map. We propose two possibilities for why this occurred.

One possibility is that incidental learning took place during training, allowing players to learn the environment regardless of the amount of effort they invested (as suggested by previous researchers [11, 126]). Incidental learning may have occurred because participants were still observing their surroundings during training, and this may have helped them remember enough to navigate later on. In both the Position and Trail conditions, we believe that it is important that participants actively participated in traversing the route — if they had not had this experience (e.g., if the assist had teleported them directly to their destination), their ability to learn the environment incidentally would be greatly reduced.

Another possibility is that the Map condition did not do a good job of scaffolding spatial learning, even though it required more effort than the other conditions. For example, if translating back and forth from the map to the first-person view was confusing for participants, it may have induced an incorrect mental model that led to participants becoming so lost that they were unable to learn the environment, despite the extra effort. However, we believe that this explanation is unlikely given that most participants in the Map-assist condition were able to complete the routes, and found at least some of the landmarks during the transfer task. Considering how much additional time the Map-assist participants spent in training (a total of 23.8 minutes across the three days, compared to 10 minutes for Trail assistance), it is surprising that they did not perform better than the other conditions on the transfer test.

To avoid the potential problem of users becoming so lost that they are unable to learn anything, Study 2 slightly modified the 3D environment to block off two sections of the map that did not have clear landmarks. Study 2 also removed the Map-assist condition, and instead used a No-assist setup as the lowest level of assistance (and the highest level of effort). The No-assist group showed the largest difference compared to the lower-effort techniques once assistance was removed, and it may be that training with no assist at all provides some benefit to learning. Participants in the No-assist condition of Study 2 took about the same time to complete training routes as people in the Map-assist condition of Study 1 — and although the routes in Study 2 were easier due to the modified environment, it is possible that the single consistent representation of the No-assist condition was better than two competing representations of Map-assist. Future work should consider whether an assist such as a basic map could be detrimental to learning, perhaps because of the added cost of translating back and forth between representations of the environment [161].

Why did Position assistance result in greater learning than Map and Trail assistance?

In Study 1, we observed that the Position assistance group performed the best in testing, significantly outperforming the group that trained with Trail assistance, and also having a lower mean than the Map-assist group (although this difference was not significant). Based on the positive results from the No-assist condition in Study 2 (see discussion above), it is possible that participants in the Map assistance group were unable to properly make use of the map to navigate through the environment. The Position assist provided the participant's current location on the map, which freed the user from having to identify and track their own location — and this likely made the map more beneficial (or otherwise allowed participants to better parse it). Having a clearer understanding of their own location in the environment may have helped participants in the formation of survey knowledge (as suggested by [300] and [290]) that could be leveraged when finding new routes to landmarks.

Why did Rail assistance not limit learning compared to Trail assistance?

Given the ability of participants in the Trail assistance condition from Study 1 to successfully complete the transfer tasks, we believe that incidental learning can be effective even when navigation effort is low. That is, by passively observing the environment (and without any intentional effort), users can develop spatial knowledge of an environment.

Study 2's results provide additional support for this idea — our results showed that participants who simply held down a button to be taken in the correct direction (Rails assistance) were equally capable of navigating the environment as participants who actively engaged with the environment to follow the glowing trail of the Trails condition.

How could this be? First, we felt it was important that participants maintained their attention on the task. Therefore, instead of allowing them to simply watch what went by on the screen, we asked them to hold down a button — if they had not watched what was on screen, there would have been no way for them to learn about the environment. Second, the Rails-style form of assistance has similarities to watching a demonstration, something that has been shown to be beneficial for early learning when considering perceptual-motor skills [232, 216].

Why weren't the observed differences more pronounced?

Although we did find statistically significant differences between our different assistance groups, when the results are considered from a descriptive perspective, it is rather surprising how small the differences between the groups are in transfer tasks. In Study 1, the Map and Trail assistance groups completed the transfer test within 10 seconds of each other, and the Map group was only 45 seconds faster than the slowest group (in a set of tasks that took about 240 seconds overall). And yet, the Map group spent considerably more total time training within the environment — about 1.8 times that of the Position assistance group and 2.4 times that of the Trail assistance group.

In Study 2's retention test, the Rail and Trail assistance groups performed similarly to the No assistance group's third day of training. This is despite the No-assist group having spent more time practicing within the environment at that point (a total of 13.7 minutes for No assistance after two days, compared to 10.4 and 9.3 minutes after all four days of training for the Trail and Rail assistance groups). Therefore, while training without any assistance does appear to result in better learning, it does not appear to be a time-efficient way for users to learn how to navigate the environment. Overall, No-assist participants spent a total of 24.9 minutes in training across all four days, more than double the training time of Trail and Rail assistance. Even with less than half the training time, assisted participants were still able to navigate the environment competently, even if they weren't able to match the performance of the unassisted participants in the transfer tasks.

Why, then, are the differences between the groups not more pronounced? Previous work on assistance and skill learning (presented in Section 6.3.1) suggests that learners given assistance will become reliant on the assist and be unable to complete the task if that assistance is taken away. However, this applies primarily to tasks which are relatively simple, such as tasks where the goal is to learn specific stimulus-response pairings [e.g., 274] or reproduce specific movements [e.g., 16]. However, when considering more complex skills, such as reproducing slalom-type movements on a ski simulator, guidance can benefit learning [351].

The learning of perceptual-motor skills often occurs through trial and error [272], where a learner attempts a task and observes the response-produced feedback to evaluate how well they performed that task [288, 257]. For simple skills, response-produced feedback is easy to parse. A learner notices their mistakes and adjusts how they are executing the skill [155]. Navigation differs from simple perceptual-motor skills in that it is not always apparent when a mistake has been made (such as taking a wrong turn). The user might simply keep moving forward until they conclude that their destination is not in sight and will not be in sight any time soon — and only then do they realize they have made a mistake sometime in the past. Navigation learning is therefore different from the learning of perceptual-motor skills that rely on trial-and-error learning, and it makes sense that the findings of work looking at those skills may not generalize to navigation learning.

This raises the question of how participants are learning the environment, if not through trial and error. As discussed previously, a likely reason that participants are able to learn how to navigate through the environment is due to incidental learning — the ability to passively acquire spatial knowledge about the environment. We believe it important that navigation assistance systems allow participants to keep their full attention on the game rather than dividing their attention, as might have occurred using other forms of assistance. Trail, Rail, and No assists all allowed participants to give the environment their full attention, while the Map and Position assists may have allowed participants to gain survey knowledge from the map.

6.6.3 Applying Navigation Assistance to Virtual Environments

In our two studies, we found that navigation assistance provided immediate benefits to participants. In Study 1, navigation assistance also did not affect learning; all participants navigated the environment equally well once assistance was removed. In Study 2, there were some detrimental effects to learning, however, performance gains made by including assistance were substantial, and the detrimental effects on learning were slight, especially when considering just how much time is saved during training by having assistance present. Furthermore, it may be possible to mitigate the detrimental learning effects of assistance by adjusting the presentation and the type of assistance provided to users. The assistance used in our studies covered an extreme range (particularly in Study 2) and there may be ways to get similar benefits while limiting any negative effects on learning.

One might first consider, however, whether there even will be a time when assistance gets removed. It might be that assistance can always be provided to users and that the detrimental effects of removing the assistance will never be experienced. Nonetheless, if there is the possibility that assistance will be removed, it should be possible to mitigate the detrimental effects. One useful approach may be to “fade out” the assistance — an approach that has been shown to be beneficial in the context of assistance for perceptual-motor skills [274, 343]. For the glowing-trail assistance, for example, this can be done by literally fading out the path as the participant spends more time in the game. For other types of assistance, similar approaches could be achieved by removing the assistance for some attempts and therefore being presented less often over time. For the rail assistance in particular, we note that a user who is taken along a route and then asked to navigate it on their own is a scenario that closely resembles the rail assistance used in Study 2, and is analogous to taking away one’s assistance.

Second, the potentially detrimental effects of assistance are dependent on the assistance, so consideration should be given to the type of assistance presented to users. For making this choice, both studies offer some insights. In Study 2, we saw that Trail assistance and Rail assistance resulted in equal levels of performance. It therefore may make more sense to give the user the autonomy to navigate the world on their own and provide Trail rather than Rail assistance. The Position assistance of Study 1 resulted in similar performance to Trail assistance, so if the system allows the user enough time to pause and review a map, this type of assistance could be used.

Applying Navigation Assistance to Game Navigation

One setting where navigation assistance would be particularly helpful is digital games. In selecting the type of navigation assistance to give a player, game designers should consider the pace of the game and whether knowing which direction to travel will greatly affect a player's performance. For example, Trail assistance is well suited to fast-paced games where navigation errors are detrimental (e.g., in a multiplayer first-person shooter). However, if navigation errors are less of a problem, and if the game affords the player enough time to pause during navigation, then providing the player with a map with Position assistance would provide the player with additional survey knowledge that they could apply later on in the game, while also providing benefits to immediate performance. This approach would be better suited for single-player games with a slower pace (e.g., open-world role-playing games). Aside from the pace of the game, there may be other reasons to include navigation assistance. In particular, navigation assistance may help facilitate social play by allowing a novice player to play with a more skilled friend.

The Rail method was our most extreme form of assistance — are there realistic scenarios in which providing players with this type of guidance makes sense? On the surface, it may seem unlikely that a virtual environment would ever make use of rail-based assistance as it is quite invasive. However, there are many scenarios in games that resemble aspects of our Rail assistance. In particular, rail-like navigation assistance can be found in cut scenes of single-player games; and in multi-player games the concept of watching others play and navigate through the environment is common. This situation comes up fairly often, for example, when watching others play games on platforms such as Twitch or YouTube, or when a player is spectating in-game (e.g., when waiting to respawn in an online team-based first-person shooter). That players can learn the environment through passive observation has interesting implications in terms of designing the experience of learning the game for new players. For example, a new player of a first-person shooter game who dies frequently will be given opportunities to learn the game simply because they will watch others play the game. This time that might have previously been seen as simply waiting to play again actually might be beneficial to the player.

Navigation assistance can also be applied by game designers to aid a player's learning of the game. New players have a lot to learn about a game before they can be successful at it, and providing navigation assistance will aid them in two ways. First, it encourages them to engage in part-task practice [185], in which the player can direct their attention towards learning only specific skills within the game. For example, navigation assistance in an FPS game could free the player to work on the skills of aiming, movement, or monitoring audiovisual cues to detect enemies [154]. Second, it decreases the difficulty of the game, potentially aiding learning by providing them with challenges just at the edge of their capabilities [325], where players feel they are able to overcome them [158] and are motivated to do so [114].

An implementation of navigation assistance in a commercial game could also consider the game state and direct the player's attention toward important objectives. For example, if a player has no weapons or is low on health, the game could show a trail to the nearest weapon or the nearest health pack. A player's role in a team game could also determine which routes are visualized for that player (e.g., a trail to a wounded player for a medic role). Finally, for scenarios in which navigational assistance is not possible, or where the player chooses to turn off the assists [311, 156], it appears that the use of even strong assistance early in a player's experience will not significantly affect their long-term performance.

6.6.4 Limitations and Future Work

A limitation of this work was that the environment we used was from an older (2003) game, and therefore the textures used were of low fidelity and the different parts of the environment were not as visually distinct from one another as they may be in virtual environments within newer games. However, this environment was used because of its ecological validity (it is an environment from a commercially produced game), the availability of the source code from the game, and because it kept the system requirements of the experiment relatively low so that more participants could successfully complete it. Further, it contains multiple alternative routes to reach each landmark. We must acknowledge that if the environment was more visually varied, it is possible that some of our participants' strategies would have been more effective (e.g., trying to remember what the area surrounding each landmark looked like). Therefore, future work should investigate whether our results will hold in different environments with different levels of detail.

An additional limitation is that this experiment was conducted using our online participants' desktop or laptop setups. Therefore, participants completed the task with a variety of displays of different sizes and types. Different displays or devices (such as head-mounted virtual-reality devices) may provide different levels of immersion or require different interfaces to help people navigate virtual environments, limiting the generalizability of our findings.

Our future work will also examine several issues raised by the studies. First, we will examine whether navigation assistance does in fact enable part-task practice that allows players to focus on other skills. Second, we will test navigation assistance in virtual environments with different styles and contents (e.g., a forest, the interior of a large building, or a more dense urban environment). Third, we will explore versions of navigation assistance that gradually disappear, to see if the downsides of abruptly removing assistance can be mitigated. Fourth, we will implement navigation assistance in actual play settings, to see if navigation assistance can improve play experience and game balance. Fifth, we will look further at whether a secondary representation of the environment (such as a map) can actually be a hindrance due to the costs of translating between representations. Sixth, we will further investigate the idea that the strength of the guidance hypothesis may be dependent on the complexity and temporal sequencing of the task. Seventh, we will revise our studies for testing with head-mounted VR displays — this will involve some changes to the different assistance techniques, but the same research questions can be explored.

6.7 Conclusion

Navigation in 3D virtual environments can be difficult for novices, and this difficulty is something that designers of these environments often want to minimize. We carried out two studies in which we tested the effects of navigation assistance on performance and spatial learning.

Results from Study 1 showed that during training, increasing the level of navigation assistance significantly and substantially improved performance: the Position-assist and Trail-assist conditions were significantly faster than the Map-assist condition, and by a large margin (e.g., for Position assist, a mean of 265 seconds totalled across the 8 routes, and 200 seconds for Trail assist, versus 480 seconds for Map assist). There was no statistical difference between Position-assist and Trail-assist. Results were similar for distance travelled.

In the final tasks with only Map assist, we found no evidence that increased assistance during training led to reduced performance with the assistance removed: in particular, there was no difference in either completion time or distance travelled between the groups who had trained with Position assist or Trail assist and the group who had trained with Map assist. In addition, the performance differences between the three conditions were much smaller in the final task than in training, with the greatest difference between our conditions being only 50 seconds compared to 275 seconds in training.

Overall, Study 1 shows that navigational assistance substantially improves performance for novice users, and suggests that early assistance does not substantially reduce performance when the assist is removed. In addition, even though the Map-assist group spent considerably more time in the virtual environment during training, this additional time did not appear to make them more familiar with the environment than the groups who had navigation assistance.

Study 2's results showed again that navigation assistance significantly and substantially improved navigation during training: mean completion time for the No-assist group was about 375 seconds, but was 155 seconds for Trail and 140 seconds for Rail assists. Similar results were found for distance travelled (mean of 1900 units for No assist vs. 760 units for Trail and 690 units for Rail).

Our wider range of assist conditions in Study 2, however, did show an effect of the navigation assist used in training when considering performance in the transfer tasks (with no assistance). Groups who had trained with either Trail or Rail assist were significantly slower on the transfer tasks (by about 130 seconds for Trail and 110 for Rail) than groups who had trained with No assist. Results were similar for distance travelled (a difference of 650 units for Trail, 770 units for Rail). Unlike Study 1, where there appeared to be no negative effect of providing increased navigation assistance, Study 2 showed that there can be performance detriments if people have an early navigation assist that is subsequently removed. However, it is important to consider the overall balance of benefits to drawbacks — and in Study 2, the increased performance associated with the navigation assist was twice as large as the reduction when the assist was removed (+220 seconds vs -110 seconds for Trail +235 seconds vs -130 seconds for Rail).

Furthermore, the performance detriment appears to diminish over time. On the retention task one week later, we found that those who trained with Trail or Rail were not significantly slower than those who trained without (only by about 60 seconds). Although they were still travelling greater distances, the size of the difference was now smaller (270 seconds and 390 units). If users who had early navigation assistance can quickly catch up with users who had to learn the environment without assistance, then there is a stronger argument for using navigation assistance.

Overall, our studies provide new empirical evidence about how navigation assistance affects performance and spatial learning — and our results imply that designers of 3D environments should strongly consider adding navigation assistance. Assistance provides major benefits both in terms of performance and user experience, and the limited negative effects on spatial learning appear to be considerably outweighed by the benefits for novices.

Measure	Question
Map Knowledge	How would you rate your knowledge of the layout of the map? ("Very poor" to "Very good")
Mental Demand	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, forgiving or exacting?
Temporal Demand	How much time pressure did you feel due to the rate at which the task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Peformance	How successful do you think you were in accomplishing the goals of the task set by the experiment (or yourself)? How satisfied were you with your performance in accomplishing these goals?
Effort	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?
Interest-Enjoyment	I enjoyed this game very much.
Interest-Enjoyment	Playing the game was fun.
Interest-Enjoyment	I would describe this game as very interesting.
Interest-Enjoyment	While playing the game, I was thinking about how much I enjoyed it.
Interest-Enjoyment	This game did not hold my attention.
Perceived Competence	I think I am pretty good at this game.
Perceived Competence	I am satisfied with my performance at this game.
Perceived Competence	After playing the game for a while, I felt pretty competent.
Perceived Competence	I am pretty skilled at the game.
Perceived Competence	I couldn't play this game very well.
Effort-Importance	I put a lot of effort into this game.
Effort-Importance	It was important to me to do well at this game.
Effort-Importance	I tried very hard while playing the game.
Effort-Importance	I didn't try very hard at playing the game.
Tension-Pressure	I felt tense while playing the game.
Tension-Pressure	I felt pressured while playing the game.
Tension-Pressure	I was anxious while playing the game.
Tension-Pressure	I was very relaxed while playing the game.

Table 6.1: All questions relating to subjective experience used within the two studies. Questions relating to mental demand, temporal demand, performance, effort, and frustration come from the NASA-TLX [125]. Questions relating to interest-enjoyment, perceived competence, effort-importance, and tension-pressure come from the IMI [193].

Within-Subject Effect	Measure	Study 1 *df*	F	p	η_p^2	Study 2 *df*	F	p	η_p^2
Day	Completion Time	**1.70, 62.9**	**18.4**	**<.001**	**.332**	**2.88, 209.9**	**14.5**	**<.001**	**.166**
	Distance Travelled	**1.73, 63.9**	**3.76**	**.034**	**.092**	**2.43, 157.9**	**7.455**	**<.001**	**.103**
Day * Assistance	Completion Time	3.40, 62.9	1.11	.355	.057	**5.75, 209.9**	**5.09**	**<.001**	**.112**
	Distance Travelled	3.45, 63.9	0.23	.901	.012	**4.86, 157.9**	**5.15**	**<.001**	**.137**

Table 6.2: Within-subjects effects for the RM-ANCOVAs for performance measures for the Training sessions from Study 1 and 2.

		Main Effect of Assistance				Pairwise Comparisons (p)		
Session	Study 1 Measure	*df*	F	p	η_p^2	Map - Position	Map - Trail	Position - Trail
Training	Completion Time	**2, 37**	**39.3**	**<.001**	**.680**	**<.001**	**<.001**	.149
	Distance Travelled	**2, 37**	**75.4**	**<.001**	**.803**	**<.001**	**<.001**	.113
	Effort	**2, 38**	**5.80**	**.006**	**.234**	.060	**.007**	.996
	Frustration	**2, 38**	**14.2**	**<.001**	**.428**	**<.001**	**<.001**	>.999
	Perceived Performance	**2, 38**	**9.75**	**<.001**	**.339**	**.029**	**<.001**	.232
	Mental Demand	**2, 38**	**24.6**	**<.001**	**.565**	**<.001**	**<.001**	**.009**
	Perceived Map Knowledge	2, 38	3.12	.056	.141	n/a	n/a	n/a
Transfer	Completion Time	**2, 37**	**3.61**	**.037**	**.163**	.248	>.999	**.039**
	Distance Travelled	**2, 37**	**4.40**	**.019**	**.192**	.100	>.999	**.026**
	Effort	2, 38	0.05	.954	.002	n/a	n/a	n/a
	Frustration	2, 38	0.84	.439	.042	n/a	n/a	n/a
	Perceived Performance	2, 38	1.91	.162	.091	n/a	n/a	n/a
	Mental Demand	2, 38	0.66	.524	.033	n/a	n/a	n/a

Table 6.3: Between-subjects effects for the RM-ANCOVAs for Study 1's Training session and ANCOVAs for the Transfer session. Each line represents a separate RM-ANCOVA (for the training session) or ANCOVA (for the transfer session).

Session	Study 2 Measure	Main Effect of Assistance				Pairwise Comparisons (p)		
		df	F	p	η_p^2	No - Trail	No - Rail	Trail - Rail
Training	Completion Time	**2, 73**	**62.7**	**<.001**	**.632**	**<.001**	**<.001**	>.999
	Distance Travelled	**2, 65**	**209**	**<.001**	**.866**	**<.001**	**<.001**	>.999
	Effort	**2, 77**	**5.95**	**.004**	**.134**	**.012**	**.009**	>.999
	Frustration	**2, 77**	**5.92**	**.004**	**.133**	**.015**	**.007**	>.999
	Perceived Performance	**2, 77**	**5.56**	**.006**	**.126**	**.021**	**.009**	>.999
	Mental Demand	**2, 77**	**4.05**	**.021**	**.096**	**.034**	.058	>.999
	Perceived Map Knowledge	2, 77	1.62	.205	.040	n/a	n/a	n/a
	Interest-Enjoyment	2, 77	0.32	.729	.008	n/a	n/a	n/a
	Perceived Competence	**2, 77**	**4.47**	**.015**	**.104**	.170	**.012**	.811
	Effort-Importance	2, 77	1.93	.152	.048	n/a	n/a	n/a
	Tension-Pressure	**2, 77**	**4.21**	**.018**	**.098**	.089	**.021**	>.999
Transfer	Completion Time	**2, 74**	**9.01**	**<.001**	**.196**	**<.001**	**.013**	.672
	Distance Travelled	**2, 72**	**9.12**	**<.001**	**.202**	**.002**	**<.001**	>.999
	Effort	**2, 68**	**5.21**	**.008**	**.133**	.207	**.006**	.502
	Frustration	**2, 68**	**6.31**	**.003**	**.157**	**.005**	**.016**	>.999
	Perceived Performance	**2, 68**	**12.0**	**<.001**	**.261**	**<.001**	**<.001**	>.999
	Mental Demand	**2, 68**	**6.38**	**.003**	**.158**	**.023**	**.004**	>.999
	Perceived Map Knowledge	**2, 68**	**17.3**	**.008**	**.338**	**<.001**	**.002**	.100
	Interest-Enjoyment	2, 68	0.18	.836	.005	n/a	n/a	n/a
	Perceived Competence	**2, 68**	**14.5**	**<.001**	**.299**	**<.001**	**<.001**	>.999
	Effort-Importance	2, 68	1.33	.270	.038	n/a	n/a	n/a
	Tension-Pressure	**2, 68**	**8.35**	**.001**	**.197**	**.003**	**.001**	>.999
Retention	Completion Time	2, 65	2.72	.073	.077	n/a	n/a	n/a
	Distance Travelled	**2, 63**	**3.32**	**.042**	**.095**	.258	**.041**	>.999
	Effort	**2, 59**	**7.12**	**.002**	**.194**	**.029**	**.002**	>.999
	Frustration	2, 59	1.73	.186	.055	n/a	n/a	n/a
	Perceived Performance	**2, 59**	**3.75**	**.029**	**.113**	.077	**.053**	>.999
	Mental Demand	**2, 59**	**6.86**	**.002**	**.189**	.300	**.001**	.147
	Perceived Map Knowledge	**2, 59**	**5.32**	**.008**	**.153**	**.014**	**.029**	>.999
	Interest-Enjoyment	2, 59	2.02	.142	.064	n/a	n/a	n/a
	Perceived Competence	2, 59	2.17	.124	.068	n/a	n/a	n/a
	Effort-Importance	2, 59	2.24	.115	.071	n/a	n/a	n/a
	Tension-Pressure	**2, 59**	**3.63**	**.033**	**.109**	.454	**.028**	.673

Table 6.4: Between-subjects effects for the RM-ANCOVAs for Study 1's Training session and ANCOVAs for the Transfer and Retention sessions. Each line represents a separate RM-ANCOVA (for the training session) or ANCOVA (for the transfer and retention sessions).

7 Introduction to Manuscript C

Manuscripts A and B focused on aiding a player with the early parts of learning a new skill by providing guidance that a novice could leverage to carry out the task of navigation. This idea of helping a player with the early parts of learning a game is rather important as it often predicts continued engagement with a game [54], however, it isn't the only time that players struggle within a game. Often a player knows what they need to do to succeed within a game (they are past the first stage of learning), yet they still struggle. In this scenario, the issue isn't a lack of knowledge or information about what to do, but an inability to properly leverage the knowledge they already have. In other words, a player can know about a skill and the inputs required to execute the skill but still be unable to succeed. What this missing is a player's ability to execute the inputs fluidly and at the correct timings. Players in this scenario may repeatedly engage with the game, making multiple attempts with the aim to improve each time. This repetition is engaging and players are strongly motivated to succeed — consider, for example, the success of games such as *Super Meat Boy* [291], *Cuphead* [287], or *Dark Souls* [108]. In these games, repeated attempts serve as *practice* that allow a player to improve at the game.

What might one do if these repeated attempts eventually no longer lead to improvements in performance? How might one overcome the challenges they face in a game?

In Chapter 2 I introduced a variety of ways that have been studied previously (though sometimes not in the context of digital gaming) that might be able to help players in this situation. An approach such as introducing deliberate practice (Subsection 2.5.2) might work, but I would argue that in a lot of cases the conditions for deliberate practice are already present. Consider a player trying to defeat a boss in *Dark Souls*. They have a well-defined and very specific goal (dodge this boss's attacks and strike them when it is safe), there is feedback present (a hit directly affects their or the boss's hit points), the task is just outside of the learner's comfort zone (presumably they should be able to defeat the boss if they were able to defeat the enemies leading to it), and the task is being given the learner's complete attention.

One other strategy that could help a player in this scenario is to make use of *spaced practice* — the idea of scheduling periods of rest to break up periods of activity. Theory suggests that these rest breaks allow the brain to generalize and compile the feedback that has been gathered, leading to improvement once the rest is finished [256, 60, 10]. Studies have shown spaced practice to be effective at improving performance across many tasks, when compared to continuous practice (i.e., no rests) [257, 72].

However, the research that existed before this manuscript was published provided little guidance on how to incorporate spaced practice within a digital game or even if it would be as effective in this setting as it has been for other perceptual-motor skills. In general, research into the theory has been criticized for rarely testing it outside of laboratory settings [48, 73], so we were interested in testing it in an environment with slightly more external validity than past work. In particular, we wanted to test a variety of different rest intervals, with a game that was commercially successful, and with players engaging with the game from their own computers and being permitted to do whatever they like during their break.

In summary, we found that spaced practice can be an effective tool for helping players get better at a game. Compared to playing continuously, spaced practice helped players perform better in the game regardless of how long they rested.

7.1 Preregistration

The design of the experiment in this manuscript was preregistered on the OSF Registries. A reproduction of the preregistration is in the appendix, Section B.1.

> Colby Johanson, Carl Gutwin, and Regan Mandryk. 2018. Spaced Practice in Video Games. https://doi.org/10.17605/OSF.IO/SK2W9

7.2 Methodological Clarifications

7.2.1 Participant Payment

The study in this manuscript paid $5.00 USD for the training session, which took 35 minutes on average to complete, excluding breaks. This works out to $8.60 USD per hour, which is more than the United States' federal minimum wage of $7.25 an hour, where our participants were from. For the retention session, participants were paid $1.50 USD and it took 10 minutes on average, so participants were paid $9 USD per hour.

7.2.2 Retention Session

For the retention session, I had participants wait one day before returning and attempting to navigate the environment again. This time was chosen because it was also the length of the longest inter-session rest interval used within the study. Additionally, for the one-day interval group, they would have started their sessions on Monday and a one-day interval between their last training session and the retention session ensured that they were completing the study within the same work week, on Friday.

7.3 Additional Analyses, Results, and Figures

When writing Manuscript C, there were some results and figures that did not make it into the final paper due to space constraints. These have been included in the appendix, in Section E.3. This includes:

- Reporting the statistical effects of the covariates used.
- Reporting on possible gender differences.

7.4 Publication and Individual Contribution

This manuscript was published as [151]:

> **Johanson, C.**, Gutwin, C., Bowey, J. T., & Mandryk, R. L. (2019). Press pause when you play: Comparing spaced practice intervals for skill development in games. *CHI PLAY 2019 - Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 169-184. https://doi.org/10.1145/3311350.3347195

This work involved contributions from my supervisors, as well as my colleague, Jason Bowey. My contribution to the work included: designing the experiment, implementing the digital system, performing data analyses, and writing the majority of the manuscript.

8 Manuscript C

Press Pause when you Play: Comparing Spaced Practice Intervals for Skill Development in Games

8.1 Abstract

Games allow players to fulfill the need for competence by providing well-designed, increasingly difficult challenges. To meet these challenges, players repeatedly attempt to achieve objectives—and through this repetition, they improve their game skills. Players are keenly aware of whether they are making progress during these attempts, and they want to get better as quickly as possible. Previous research suggests that one way of improving skill development is by taking breaks between periods of activity (called "spaced practice"). However, there is little knowledge about whether this idea works in games, what the optimal break length is, and whether the effects last. We carried out a study comparing spaced and continuous practice in a *Super Hexagon* clone, using five-minute play intervals and five break lengths (no break, two minutes, five minutes, ten minutes, one day). We found that spaced practice led to significant gains in performance, particularly for novices. This result shows that players can achieve an immediate improvement in skill development, simply by scheduling short breaks in their play session; designers can also make use of this result by building rest periods into the structure of their games. Our study also indicated that breaks are valuable both in the short and the longer term—in a retention test after one day, all of the groups performed similarly, suggesting that even if a player does not use spaced practice initially, taking a break after the play session can still lead to improvements. Our study provides new information that can aid in the design of practice schedules for perceptual-motor tasks in games.

8.2 Introduction

Repetition of a task is an essential part of many digital games. Repetition in games may inspire memories of classic arcade games on which players willingly spent coin after coin, all for the chance to beat their high score. However, even in modern games, players show a desire to repeat—and improve their performance on—in-game tasks [237, 247]. Consider, for example, the success of games like *Super Meat Boy* [291] or *Cuphead* [287], which require players to repeat, and improve, sequences of game actions; or the appeal of "speedruns", where players practice a level over and over to set a time record.

In most games, these repeated attempts are not just thrashing; instead, players get better with practice. When a player attempts a difficult challenge and fails, they can apply that experience to future attempts [162]. Repetition therefore serves as *practice* that can lead to the development of in-game skills—a desirable outcome for the player. Furthermore, because games are typically designed to become more difficult as the player makes progress [237, 342], players must continuously develop their skills to keep progressing. In-game skill development is a key reason why gamers are intrinsically motivated to play, because developing and improving skills satisfies their need to experience competence [237, 167, 230, 180]. In addition, many games also have mechanisms that provide extrinsic motivation to improve one's performance, such as in-game rewards for success [342], out-of game achievements [75], or leaderboards that allow players to compare themselves to their peers [35]. Because players are so interested in improving their skills, and in doing so as quickly as possible, it is important that we better understand how in-game skills develop and how this development can be accelerated.

The skills that are developed in many games are perceptual-motor skills (requiring coordination of physical actions such as button presses with on-screen events), and in this domain, several strategies based on theories of skill development have been shown to improve performance—for example, in aiming [103, 25], pursuit tracking [7, 34], and mirror tracing [277]. However, although game developers have made use of ideas about skill development in other areas (e.g., introducing mechanics one at time [123, 160], providing clear feedback [297, 158], and allowing players to immediately practice skills they have just learned [342, 143]), there has been little application of theories of motor learning in games.

One theory of skill development that has been widely studied is *spaced practice*—the idea of scheduling periods of rest to break up periods of activity. Theory suggests that these rest breaks allow the brain to generalize and compile the feedback that has been gathered, leading to improvement once the rest is finished [256, 60]. Studies have shown spaced practice to be effective across many tasks, when compared to continuous practice (i.e., no rests) [257, 72]. However, there have been very few investigations of spaced practice in digital games—and in general, research into the theory has been criticized for rarely testing it outside of laboratory settings [48, 73]. Furthermore, the motivational pull of games changes the idea of "practice" in that it is not the chore that it is in other contexts where this effect has been found, and players may want to keep playing without a break.

This is evident by the strong reactions of players at the mere suggestion of taking a break. Many *Wii Sports* [221] players mock the game's suggestion that they should take a break after playing for an hour [313]. Even games that take a more implicit approach to breaks spark debate in the gaming community. For example, players who die fighting a boss in *Dark Souls* [108] are made to traverse the game world to make another attempt and there are players who question this design choice. These players would rather attempt the boss again immediately (e.g., [315, 316, 40]), and they describe the walk back as simply a waste of time; others argue that the relatively easy task of walking back to the boss gives them an opportunity to improve by reconsidering their strategy [315]. This debate shows that a typical behaviour for many people is to play continuously with no breaks—and if pausing could be shown to actually help these players overcome challenges, it could have a broad effect on how people play.

To better understand how spacing practice affects skill development in digital games, we carried out a study using a clone of a commercial game (*Super Hexagon* [45]) that is almost entirely based on perceptual-motor skills. We divided a total play time of 25 minutes into five sessions: four training sessions and one retention session. Participants received different amounts of rest between training sessions: three seconds (i.e., continuous practice), two minutes, five minutes, ten minutes, or one day. The retention session took place one day after the participant's last training session. Our study was designed to answer three questions:

- Are there benefits of spaced practice on performance in digital games?
- Do performance benefits last beyond the training session?
- Is there an optimal rest period for this game?

Our results showed that there were clear and *immediate* benefits of spaced practice. There were significant differences in the rate of improvement in the game, with the continuous-practice group performing worse during training than any of the spaced-practice groups. These results suggest that scheduling short rest breaks can provide an immediate benefit in skill development, particularly for novice players (who may be the group most interested in accelerating their skills). In addition, we found that a longer break (a one-day rest between training and the retention test) was also effective in improving performance, even for the group that played continuously during training – by time of the retention test, all of the groups performed similarly. Among our rest intervals, there was no one interval that worked better than any other. The ten-minute group improved more quickly early on, but the differences between groups tapered off in later sessions, suggesting that the effects of spaced practice could interact with player experience.

Our results show that designers can use spacing to help players improve in the moment—an important factor in perceptions of competence and player experience—and also show that longer breaks can assist players regardless of practice schedule. Our study provides designers with valuable new understanding of how skill development works in games, and suggests new strategies for helping gamers improve as they play.

8.3 Related Work

8.3.1 Spaced Practice

Spaced practice means scheduling periods of rest to break up periods of work within a training session [256]. This approach has been shown to improve performance during training, and to improve retention as well [174, 81, 256]. There is no fixed timing for the rest periods relative to work periods [321], and any amount of rest compared to a continuous-practice condition is typically considered spaced practice [256].

The effect of spaced practice has a long history, going back to Ebbinghaus's work on learning lists of nonsense syllables [86] and Snoddy's work on mirror tracing [277]. Some of this past work differentiated between performance during training (temporary performance) and retention after training (permanent performance) [7, 34]. In some studies, improvements due to spaced practice affected only temporary performance (because continuous practice conditions could be affected by increased fatigue, boredom, or failure to sustain attention [321]). Retention tests are therefore seen as an important part of assessing the value of different practice schedules—and studies have shown positive effects of spaced practice on retention after a delay of a day or longer [7, 23].

Meta-reviews of spaced practice studies have showed strong overall effects for the technique [174, 81]. For example, Lee and Genovese found a large mean weighted effect size of 0.96 for training (temporary performance), and a medium effect size of 0.53 for retention [174]. However, Verhoeven and Newell suggest that the meta-reviews do not necessarily provide unequivocal support for the idea that spacing practice enhances learning compared to continuous practice, as there are aspects of practice that moderate the effectiveness of spacing (e.g., differences in the task or the learner) [321]. Additionally, there is little agreement as to what length of rest optimizes the effect [321]—some suggest that longer breaks are more effective than shorter breaks [256], and others suggest that performance follows an inverted U function [48, 81].

Task Factors

Spaced practice has been shown to work for several different tasks, but there are also examples of spaced practice not producing improvements. For example, more complex tasks such as math problems [207] and learning a musical sequence on the piano [345] did not show a benefit for spaced practice. In general, spaced practice appears to be more effective for simple perceptual-motor tasks than for complex tasks [81]. Additionally, Lee [175] found that discrete skills with a definite beginning and end did not benefit from spaced practice.

Individual Differences

In addition to the complexity of the task being performed, the individual's skill level in the task has also been shown to influence the effectiveness of spaced practice. As early as 1926 it was thought that spacing was most effective in the early stages of learning [277], and other work has also shown that spacing is more important in early stages, with continuous practice being better in later stages [164]. Recent work has also showed that an individual's performance after a break was better predicted by the stage of skill development rather than the break itself [292].

Spaced Practice in Digital Games

The amount of research investigating spaced practice in digital games is limited, with the majority focusing on serious games in the context of education and verbal learning (e.g., [87, 251, 252]) rather than on in-game perceptual-motor skills. When limiting our search to this latter topic, we found only two relevant experiments: a 1985 experiment by Metalis [200], and a 1999 experiment by Shebilske et al. [268].

The 1985 experiment [200] used the Apple II+ game *Little Brick-Out*—a *Pong*-inspired single-player game in which the player controls a paddle to hit on-screen bricks with a bouncing ball. The experiment consisted of 10 rounds of the game, with either a 2-minute break or no break between rounds. Each round lasted anywhere from 57 seconds to 232 seconds, depending on how well the participant performed. It was found that spacing practice resulted in better performance at the 10th training sessions than continuous practice. No retention test was performed. However, this study has a number of issues. The participant group in the spaced practice condition had higher initial performance than the continuous group, and because the game's rounds lasted longer if the player performed well, this meant that the spaced-practice group actually trained with the game for far longer than the continuous-practice group (27 minutes compared to 15 minutes). In addition, the 18 minutes of rest made up a large amount of the total training time.

The 1999 experiment [268] looked at spaced practice in the context of a more complex game called Space Fortress [80]. Participants played for much longer (10 hours over either 2 days or 10 days), and also performed a retention test one week after the last day of training. The group that trained over 10 days outperformed the two-day group both at the end of training and on the retention test. The game used in this study involved complex tasks with strong strategic components, and so did not focus primarily on perceptual-motor tasks. In addition, the experiment included substantial coaching of the participants (e.g., participants watched videotaped instructions, were encouraged to try different strategies, were encouraged to try their best, and were instructed on strategies that other players had found effective). Even with these two studies, there is still little information about the effects of spaced practice on motor skill development in digital games.

8.3.2 Esports: Skill improvement as a profession

Whereas many players repeat game objectives simply to make progress in an entertaining game, the recent rise of esports shows that skill improvement and practice effects have now become a job as well as a leisure activity. Several definitions of esports focus on the importance of training (e.g., "sport activities in which people develop and train mental or physical abilities in the use of information and communication technologies" [326]). The esports industry is now large: a 2019 market report estimated the esport economy to be at $1.1 billion with 26.7% year-on-year growth; further, the global esport audience is estimated to be 443 million people [217].

The need for skill development in esports is a clear example of why it is critical to better understand how techniques such as spaced practice affect training. Practice has long been an important part of traditional sports, and many parallels are found between athletes in esports and more traditional athletic pursuits, particularly in the areas of perceptual-motor expertise and the drive for improvement of those skills [263, 131]. Among these similarities we see common patterns involving training and practicing fundamental skills. Development of skills is crucial for esport players—but is also important for any player who wants to perform at a high level. Understanding how skills in a digital game are developed is important for creating effective training programs both for professionals and amateur players.

8.4 A Framework of Skill Development for Games

There are several issues that need to be considered when designing or evaluating a practice scheme for digital games. In the following sections we review game skill types, the stages of skill development, and how skill development is evaluated. This framework was used to inform the design of our game and our experiment; and although our research focuses primarily on perceptual-motor skill learning, this framework describes skill learning in general.

8.4.1 Game Skill Types

Games contain many different skills that players must master in order to succeed [225, 224, 154, 149, 322]. A skill refers to the ability to carry out a specific task to achieve a specific goal [98, 302, 256, 89]. Skills must be learned—learning refers to a relatively permanent change in behaviour that occurs as a result of practice, expertise, or experience [298, 341, 249, 302].

Skills can be broadly classified as either *cognitive* or *motor* [302]. Cognitive skills include problem solving, memory, language, and emotional skills, while motor skills include anything that requires body or limb movements in order to make a physical response [98, 302, 184]. Despite these separate classifications, the reality is that many tasks, including digital games [225], require components of both types of skills [302, 184]. Motor skills are usually referred to as *perceptual-motor* or *psychomotor* because of the importance of *perception* and *decision making* in the process [89, 134]. The learner must learn to process stimuli so as to recognize features that require a response [89, 341, 113, 236, 302]. In other words, the learner develops a stimulus-response *coding* that lets them quickly select correct responses [333, 236].

8.4.2 Games That Facilitate Skill Development

Games are considered to be effective learning environments [162, 114]. This is because they have clear goals with strong feedback [297], and are a task that players are willing to give their complete attention to [52]. These factors have been shown to facilitate continued skill development [95].

Feedback acts as a source of motivation for players [158, 226, 237]. It provides them with information regarding how well they are doing [297]—if they are doing well then it can enhance their needs satisfaction of competence, which will in turn make the player more motivated to keep playing [237]. Conversely, if the player is doing poorly, then a motivated player will leverage this information to modify their strategy or response [110, 237].

This feedback is most effective when players are presented with tasks just outside of their comfort zone [95, 162]. Without this, the player may adopt a strategy or response that is acceptable for the difficulty of the task, but is sub-optimal compared to other approaches or when the task becomes more difficult [260, 95]—a phenomenon known as "satisficing" [271]. Games use different approaches to achieve this. When a game's challenge primarily comes from its mechanics or design, the game can be implemented to consistently increase in difficulty as the player makes progress [342, 237]. If the game's challenge primarily comes from competing against other players, a matchmaking system can be applied so that similarly skilled players compete against one another [226].

8.4.3 The Stages of Skill Development

In digital games, players can continue to make performance improvements over dozens or hundreds of hours [33, 173, 140]. For some games (as with esports) the high skill ceiling makes it possible to play a game as a full-time job and continue to see performance improvements [17, 84]. Changes in performance are described primarily by two theories: the power law of practice [277, 214, 257, 302], and the three stages of skill development [257, 104, 103, 236, 235, 164,

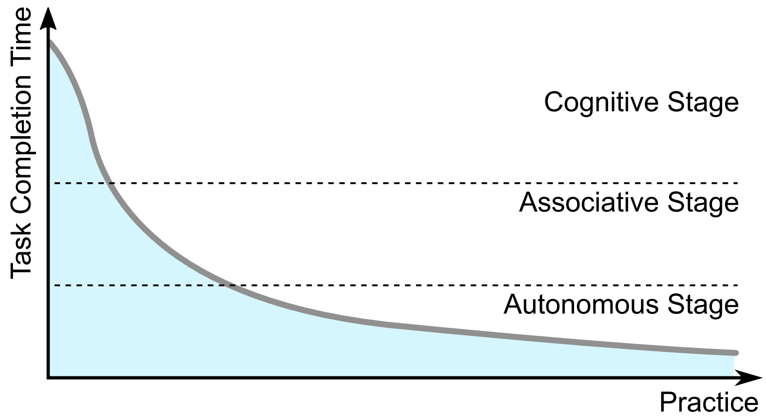


Figure 8.1: The three stages of learning approximately mapped to the power law of practice. Adapted from [164].

163].

The acquisition of skill follows a predictable pattern, described by the *power law of practice* [277, 214, 257, 302] (shown in Figure 8.1). Simply put, skill improvement occurs as a function of the number of repetitions, with performance improving dramatically at the start, but eventually slowing down. In the case of digital games, special consideration should be given to the fact that many games are designed to increase in difficulty as the player also improves [342, 237].

The *cognitive* stage of skill development is characterized by initial poor performance with many errors, but also by rapid improvement [214, 257, 104, 103, 236, 235, 164]. Learners must give the task their full attention [56, 261] as they use their declarative (verbalizable) knowledge of the task [164] to learn how the task is performed [104, 256, 302]. As they attempt the task, they begin to refine their attempts based on any feedback that is provided [344, 248, 272, 166, 341, 302, 113]. At the end of the cognitive stage, learners start to form the stimulus-response codings [236, 136] and the procedural knowledge [164, 339] that they will use in higher stages; these structures are more robust to decay [164] and allow quicker response times and improved performance over declarative knowledge alone [191, 27].

The *associative* stage is characterized by a dramatic reduction in errors, greatly improved performance, and slower improvements [214, 104, 302, 257]. The learner's attention is on optimizing their performance [302, 256]. There are three main reasons for performance improvements in this stage: first, the learner develops a consistent response in the form of learned patterns [104]; second, their ability to identify relevant stimuli improves [302]; third, the learner no longer needs to use the declarative knowledge of the task [164] and can instead leverage procedural knowledge and "direct stimulus-response associations" [236].

The *autonomous* stage is characterized by few errors, stable expert performance, and little further improvement in the motor domain [256, 104, 214]. The learner performs the skill with coordination, smoothness, and accuracy [302]. The learner can respond to stimuli with automaticity [104, 302, 236, 261], which allows them to direct attention towards improvement through changes in high-level strategies [95]. While continued improvement at this stage is difficult, it can be accomplished by making use of deliberate practice [95] — a type of practice where the learner avoids acting with automaticity and focuses on specific aspects of the task that can be improved.

8.4.4 Evaluating Skill Development

Skill learning cannot be observed directly, and so is inferred by examining the learner's performance [257, 302, 134, 278]. However, the learner's performance during or immediately after training is often insufficient because transient factors such as fatigue may be at play during the training session [257, 278]. Longer-term learning must therefore be evaluated by having the learner perform an additional test with the experimental variable removed, either in a transfer test (where the learner completes a task different from the training task) or a retention test (where the learner completes the same task but after a delay) [257, 278]. It is important to note, however, that in the game domain, we are interested in both immediate improvement (i.e., as the player is repeatedly attempting the objective) as well as longer-term retention of a skill.

8.5 A Digital Game for the Study of Perceptual-Motor Skill Development

Given that a typical experiment lasts less than an hour and that games can be played for dozens or hundreds of hours, designing a game that can be used to evaluate skill development in the context of this tight time-frame poses some challenges. We wanted a game that players could begin playing immediately, with little need for instructions, demonstrations, or guidance to introduce the game mechanics. The game should also present the player with clear goals, and only require minimal learning of declarative knowledge. In addition, players should be able to continuously make performance improvements over several play sessions, and should be motivated to continue to play and improve [80]; the game should be designed to avoid "at-game frustration" [204] and should provide clear feedback [167, 158]. Finally, performance in the game should be straightforward and easily quantifiable [80], and ideally not based on aggregated measures [33].

8.5.1 Super Hexagon

Based on these guidelines, we produced a clone of the game *Super Hexagon* using the Unity game engine. *Super Hexagon* is described by its creator as a "minimal action game" [46]. The player controls a triangle that can rotate radially around a central hexagon, using the left and right arrow keys. In each of the six regions on the screen, obstacles appear at the outside of the screen (white bars) and move inwards towards the center—requiring that the player rotate to avoid the obstacles (see Figure 8.2). The game's goal is simple—last as long as possible while avoiding the incoming obstacles. The camera rotates continuously to provide some perceptual confusion. As long as the player stays alive, the game gets progressively harder by increasing the camera rotation speed and the rate at which the obstacles spawn and move inwards. The difficulty is reset every time the player fails.

Super Hexagon fits our guidelines because it has minimal controls (only two keys) and a clear goal (avoid the obstacles), meaning that players can begin playing immediately. Performance improvements can continue for a long time (the average completion time for the entire game is around 15 hours [140]), and performance is easily measured as time until failure. Additionally, its commercial success [147, 284] and critical praise [199] suggest that many gamers

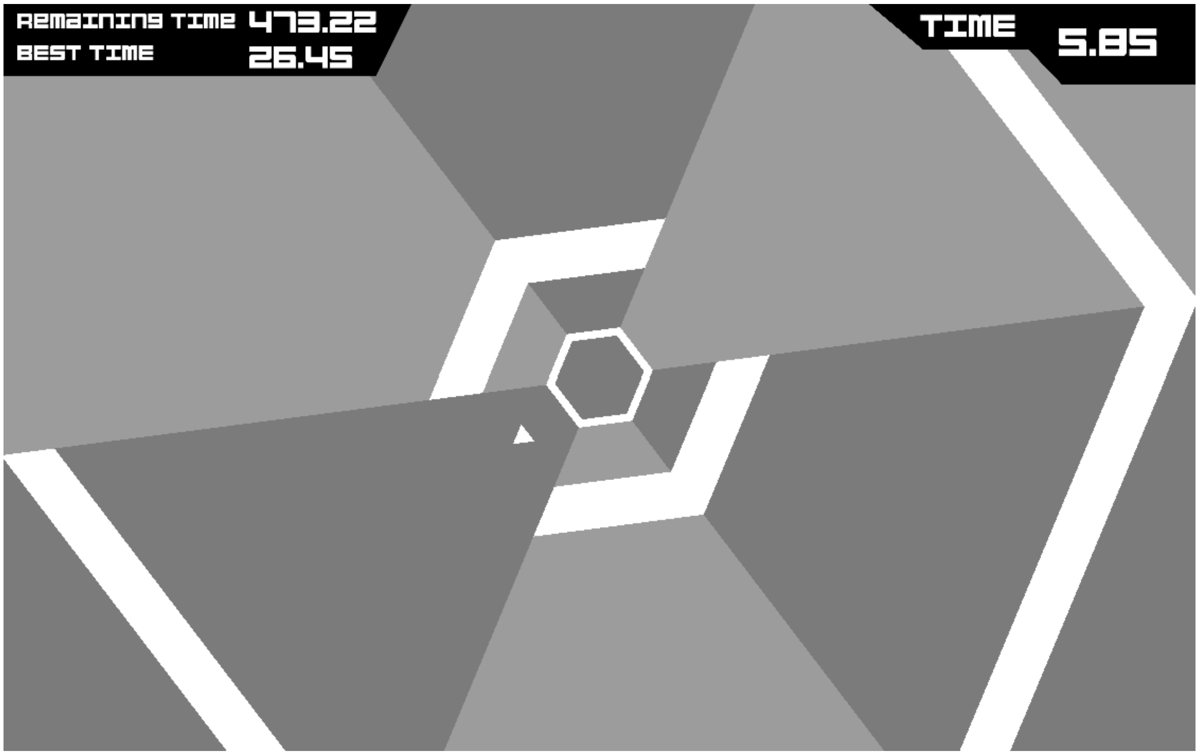


Figure 8.2: A screenshot of our *Super Hexagon* clone.

find the game enjoyable.

The game's challenge comes primarily from interpreting and processing the on-screen stimuli and choosing a correct response: a simple motor movement of pressing one of two keys. The stimuli are the on-screen obstacles that players must observe and process, and the player must decide whether to move or stay still. The human visual perception system takes in far more information than can actually be processed at any point in time [333], so performance improvements in this task are due to the player developing a "coding" [236] of the patterns of the obstacles that they can use to identify what they are seeing. If the player correctly identifies that a response is required, then they have to choose the correct response in terms of the right or left arrow keys and the amount to move. This requires players to develop a mapping between the key and the action, something that also results in improved performance. While this skill is relatively straightforward, it can still be considered to be a high-performance skill similar to those found in many sports and esports, in that there are very clear differences between experts and novices [133, 95] and performance gains continue for a long time [260, 95].

8.6 Experiment

We designed an online experiment to test the effects of spaced practice in a high-performance game environment[1]. We had participants play a total of 25 minutes of our *Super Hexagon* clone, broken into five-minute sessions — four training sessions and one retention session.

The first four sessions served as the training period, and between these sessions participants were given inter-session rests, with those rests varying depending on the group they were randomly assigned to. There were five inter-session rest interval groups: A Continuous Practice Group—operationalized as an inter-session rest interval of 3 seconds—and four Spaced Practice Groups, with inter-session rest intervals of 2 minutes, 5 minutes, 10 minutes, or 1 day. These rest intervals were chosen based on the Donovan and Radosevich [81] meta-review, which found that tasks similar to ours benefited most from a rest interval between 1 and 10 minutes, although we included a 1-day interval to make further comparisons. To evaluate how these different rest periods affected the participants' learning, a final 5-minute retention session was completed one day after the final training session. The delay was chosen to ensure that any temporary effects of spacing practice would not be present [81].

To accurately reflect what would happen "in the wild" when players are given rest periods, we instructed the participants assigned to the Spaced Practice Groups that they were allowed to use their computer in any way they wished during the periods of rest, but that they should return to the task promptly when the rest period ended. We notified participants via an audio cue to return to the game, except for the *1-Day* interval group, who were invited back to the task via an email. One day after their last training session, all participants were invited back via email to complete the final five-minute retention session.

8.6.1 Measures

Questionnaires

We used several questionnaires to get a sense of each participant's interest in the task and ability to complete the task.

Gaming Experience. We expected that individuals who enjoy playing video games regularly might enjoy our game more than those who do not, and so may try harder or improve more quickly. We therefore asked participants a number of questions to evaluate their level of experience with video games, including: "How much do you self-identify as a gamer?", "How many years have you been playing video games?", "How often (on average) do you play video games?", and "If you have played games more often in the past, how often were you playing at peak times?".

Attentional Control. We anticipated that participants who are better at maintaining attention and focus on the task might perform better. We therefore used Derryberry and Reed's Attentional Control Scale (ACS) [74] to measure each participant's attentional control.

Immersive Tendencies. We used Witmer and Singer's Immersive Tendencies Questionnaire (ITQ) [346] to measure participants' tendency to experience presence in virtual environments. The questionnaire consists of three sub-

[1]The design of this experiment was pre-registered on OSF.io at: `https://osf.io/sk2w9`

scales: involvement (propensity to get involved with an activity), focus (ability to concentrate on enjoyable activities), and games (how much they play games and whether they become involved enough to feel like they are inside the game).

Current Motivation. Participants may have different levels of interest in completing our task, so we used Guay et al.'s Situational Motivation Scale (SIMS) [121] to measure the participant's intrinsic motivation, identified regulation, external regulation, and amotivation towards our experiment. We also used Rheinberg et al.'s Questionnaire on Current Motivation (QCM) [240, 324] to measure the participant's task-related anxiety, probability of success, interest, and challenge.

Achievement Orientation. Participants with a competitive nature may invest more effort into the task, so we measured their competitiveness, win orientation, and goal orientation using Gill and Deeter's Sport Orientation Questionnaire [116].

Dependent Measures

We used four dependent measure for the experiment (one for performance, and three for subjective experience). Performance was measured for every session while subjective experience was measured twice: after the last training session and after the retention test.

Average Life Time. For each session, the participant played multiple rounds. Looking at the sessions individually, we took the average of the scores (time) among each session's rounds. If the round was ended prematurely by the timer running out, that round was excluded from the average.

Intrinsic Motivation. We evaluated the participants' intrinsic motivation towards the game by using the Intrinsic Motivation Inventory (IMI) [193]. The IMI measures the participant's interest-enjoyment, effort-importance, and tension-pressure.

Flow. We used Engeser and Rheinberg's Flow Short Scale (FSS) [90] to measure the participant's experience of flow while playing the game. We used the fluency of performance and absorption by activity subscales.

Immersion. We used Jennett et al.'s questionnaire on in-game immersion [148]. The questionnaire measures the participant's attention, temporal dissociation, transportation, challenge, emotional involvement, and enjoyment.

8.6.2 Recruitment and Participants

Our online experiment was conducted on Amazon's Mechanical Turk—a system that acts as a broker between willing workers and requesters, who provide paid human intelligence tasks (HITs). Mechanical Turk has been used for human-computer interaction research in the past (e.g., [153, 35, 269, 165]) and has been proven to be effective, as long as some attention is given to verifying the quality of the data [83, 198, 190, 228, 227].

We were interested in recruiting exclusively *Super Hexagon* novices, so we first posted a screening task, which asked three questions: "What type of input device are you currently using?", "Which hand are you currently using for that input device?", and "Do you have any experience with the game *Super Hexagon*, or its clones?", on a five-point scale from "no experience" to "extremely experienced". The first two questions served to obscure our intent. We had

Session Comparison	2–1	3–2		4–3		Retention–4		4–1	Retention–1
Measure	Δ Mean	Δ Mean	p	Δ Mean	p	Δ Mean	p	Δ Mean	Δ Mean
3-Second (actual: 3.01 ± 0.06 sec.; $n = 20$)	1.799	0.674	>.999	-0.088	>.999	**2.513**	**.004**	2.386	4.899
2-Minute (actual: 2.20 ± 0.21 min.; $n = 22$)	2.507	**1.834**	**.009**	**1.671**	**.026**	0.501	>.999	6.012	6.513
5-Minute (actual: 6.49 ± 1.69 min.; $n = 21$)	3.253	0.998	.531	**2.155**	**.003**	-0.432	>.999	6.407	5.975
10-Minute (actual: 11.08 ± 1.26 min.; $n = 21$)	3.561	**2.498**	**<.001**	0.095	>.999	1.310	.382	6.154	7.464
1-Day (actual: 24.64 ± 2.29 hr.; $n = 21$)	1.884	**2.200**	**.002**	0.429	>.999	**3.480**	**<.001**	4.513	7.993

Table 8.1: The Performance difference between the Sessions for each Interval Group. Bold text indicates a significant difference between the two Sessions' Performance. Δ Mean is the difference between the estimated marginal means of the two sessions. Session 1's normalized Performance (average life time) was evaluated at 6.562 seconds, as computed by the RM-ANCOVA. $\pm$ on actual rest time is the standard deviation.

535 workers complete the HIT. Of those, we gave 397 that indicated they had no experience with *Super Hexagon* a "qualification" that would permit them to participate in our experiment.

The immediate time commitment of the training phase could vary considerably between conditions (from 5 minutes to 50 minutes), so we were transparent in outlining how the the immediate time requirements could vary based on random assignment to condition. Once randomly assigned, the participants MTurk ID became associated with a particular condition and so attempts to be re-assigned to different conditions would be unsuccessful. These steps were taken to avoid self-selection into groups, however, we observed differential dropout in that those assigned the *1-Day* group seemed more likely to not return for subsequent days of practice.

Ethical approval for this study was obtained from the behavioral ethics board of the University of Saskatchewan, and participants were asked to renew their consent at the start of each day's task. To comply with ethical guidelines, the task was only available to workers from the United States who were over 18 years old. Participants were paid \$5.50 USD for the training sessions, and an additional \$1.50 USD for the retention test (approximately \$10 an hour). We had 138 participants complete both the training and the retention test.

Data collected from Mechanical Turk can include low-quality responses [83, 198]; therefore, we filtered out non-compliant participants based on a variety of criteria. We excluded 2 participants due to entering an invalid age (less than 18 or greater than 99) and 13 participants due to taking a rest that was more than 1 standard deviation longer than their peers within their rest group (actual rest periods for each group are reported in Table 8.1). We also filtered out participants based on their in-game logs indicating low framerate (11 participants) or evidence suggesting that they stopped playing the game (9 participants). This was determined by counting the number of sessions in which they performed worse than their initial session, and excluding them if they performed worse in two or more sessions. In total, 33 participants were excluded (some met more than one criteria for exclusion), leaving 105 participants—52 female and 53 male, with an average age of 37.8 (min 19, max 68). Table 8.1 lists how participants were distributed among the conditions.

8.6.3 Results

To investigate the Performance differences between groups, we used SPSS to perform a repeated measures analysis of covariance (RM-ANCOVA). We used Session (2, 3, 4, and the Retention Session) as the repeated-measure factor, and inter-session rest interval (3 seconds for the Continuous Practice Group and 2 minutes, 5 minutes, 10 minutes, or 1 day for the Spaced Practice Groups) as the between-subjects factor (with all groups being subsequently being referred to as the "Interval Groups"). In-game Performance (average life time for Session) was used as a dependent measure. Instead of using Session 1 as one of the repeated measures, we used it as a covariate to compensate for individual differences in initial performance. This approach means that we cannot make comparisons to Session 1; however, it has been shown to provide more statistical power under similar circumstances [37], and we wanted to be certain that we accounted for individual differences in *initial* performance. To further acknowledge individual differences between participants, we included further covariates—the results of the subscales from our trait questionnaires of gaming experience, attentional control, immersive tendencies, achievement orientation, and current motivation—in our RM-ANCOVA. All pairwise comparisons were made using Bonferroni corrections. This approach allowed us to answer a number of questions.

Did spaced practice result in differences?

Yes. We found a significant effect of Interval Group on Performance ($F_{4,85} = 3.649$, $p = .009$, $\eta_p^2 = .760$), indicating that there were differences between Interval Groups. We also found a significant interaction between Session and Interval Group ($F_{12,255} = 3.078$, $p < .001$, $\eta_p^2 = .127$), indicating that there were differences in Performance over the Sessions that were due to the Rest Group.

Examining the Sessions individually, there were no significant differences between the Interval Groups for Session 2 ($F_{4,85} = 2.173$, $p = .079$, $\eta_p^2 = .093$), but there were for Session 3 ($F_{4,85} = 3.309$, $p = .014$, $\eta_p^2 = .135$), Session 4 ($F_{4,85} = 5.310$, $p = .001$, $\eta_p^2 = .200$), and the Retention Session ($F_{4,85} = 2.565$, $p = .044$, $\eta_p^2 = .108$).

How did the Interval Groups compare?

Pairwise comparisons for Session 3 showed only that the Continuous Practice Group performed significantly worse than the *10-Minute* Interval Group ($p = .005$), with all other comparisons not being significant ($p \geq .501$). For Session 4 (the final training session), pairwise comparisons showed that the Continuous Practice Group performed significantly worse than the *2-Minute*, *5-Minute*, and *10-Minute* Spaced Practice Groups (all $p \leq .004$), but not the *1-Day* Interval Group ($p = .339$), with all other comparisons being not significant ($p \geq .648$). There were no significant differences between the Interval Groups for the Retention Session ($p = .056$ for the comparison between *1-Day* and *3-Second*, $p \geq .254$ for the others).

How did Performance change over time?

There was a main effect of Session when controlling for Session 1's Performance ($F_{3,255} = 8.189$, $p \leq .001$, $\eta_p^2 = .088$). Pairwise comparisons between the Sessions revealed that every Session was different from the others (all $p < .001$).

It was not possible to determine significance for comparisons to Session 1 Performance due to its use as a covariate instead of a repeated measure. Therefore, to evaluate whether significant improvements were made for each spacing interval, we make our comparisons to Session 2: Session 4 to Session 2 and the Retention Session to Session 2.

After the final training session (Session 4), the Spaced Practice Groups (*2-Minute*, *5-Minute*, *10-Minute*, and *1-Day*) had made significant improvements to their Performance (all $p \leq .001$), whereas the Continuous Practice Group (*3-Second*) had not ($p > .999$). After the Retention Session, every Interval Group had made significant improvements in Performance (all $p \leq .001$).

Table 8.1 shows the Performance gains for each Interval Group over each Session. We observe that the Continuous Practice Group made only marginal improvements during training (Sessions 1 to 4), but were able to make up for it with a significant improvement in the Retention Session. In comparison, the Spaced Practice Groups made significant improvements in Performance on some of the training sessions, but not all, and the only Spaced Practice Group that continued to improve on the Retention Session was the *1-Day* Interval Group.

How did spacing affect subjective experience?

To evaluate differences in subjective experience, we used a separate RM-MANCOVA. We collected subjective measures after the final training session and Retention Session; the results of the subscales for Intrinsic Motivation (IMI), Flow (FSS), and Immersion were used as the dependent measures. We used the same covariates and between-subjects factor as in the previously described RM-ANCOVA.

Only the measures of Enjoyment differed between the Interval Groups, for both the Immersion questionnaire's Enjoyment subscale ($F_{4,85} = 2.733$, $p = .034$, $\eta_p^2 = .114$), and the Intrinsic Motivation Inventory's (IMI) Interest-Enjoyment subscale ($F_{4,85} = 2.490$, $p = .049$, $\eta_p^2 = .105$). All other subscales had no significant differences due to Interval Group ($p \geq .202$). Examining the Sessions individually revealed that the differences in enjoyment were present after the training sessions ($F_{4,85} = 3.473$, $p = .011$, $\eta_p^2 = .140$ for Immersion, and $F_{4,85} = 3.001$, $p = .023$, $\eta_p^2 = .124$ for IMI), but not after the Retention Session (both $p \geq .168$).

We used pairwise comparisons (Bonferroni) to examine the between-group differences after the final training session and found no significant differences between the groups for Enjoyment (Immersion) ($p = .051$ between *3-Second* and *2-Minute*, $p = .060$ between *3-Second* and *1-Day*, and $p \geq .283$ for the others). The only difference between the Interval Groups for Interest-Enjoyment (IMI) was between the Continuous Practice Group and *2-Minute* Interval Group ($p = .022$), with all other comparisons being not significant ($p \geq .143$).

In summary, there were marginal differences in subjective experience, but only for Enjoyment. The Continuous Practice Group enjoyed the experience less than the *2-Minute* Interval Group (mirroring the performance results for these groups).

8.6.4 Summary of Results

Our results indicated that there were differences in Performance as a result of varying the inter-session rest interval. After training (Session 4), the Continuous Practice Group (*3-Second*) performed significantly worse than every Spaced Practice Group, except for the *1-Day* Interval Group. After a one-day break, on the Retention Session, these two groups effectively "caught up" to the performance of the other Interval Groups. In terms of our framework, this indicates that in our scenario, spaced practice had a strong effect on immediate performance, but a relatively weak effect on long-term performance. Even though there are no significant differences between groups after the Retention Session, the Continuous Practice Group performed the worst over all, looking at the changes in performance (see Table 8.1).

Although the performance essentially equalized by the Retention Session, there were differences in how the Interval Groups arrived at that performance. The *10-Minute* Interval Group's Performance increased faster than any other between Sessions 1 and 2 and 3, but their Performance improvements slowed down between Session 4 and 3 and the Retention Session and Session 4. The *5-Minute* Interval Group's Performance gain between Session 1 and 2 was the second highest, but they failed to make significant improvements between Session 3 and 2, then made up for it with improvements between Session 4 and 3. They then did not improve at all over the Retention Session and were the only Interval Group to perform worse on that Session compared to their Session 4 Performance. The *2-Minute* Interval Group was the only Spaced Practice Group that made gains in Performance consistently over the sessions. They improved third most between Session 1 and 2, and then improved significantly between Sessions 2, 3, and 4. However, they did not improve on their Session 4 Performance in the Retention Session. The *1-Day* Interval Group made the least gains in Performance between Sessions 1 and 2 of all the Spaced Practice Groups, significantly improved their performance between Sessions 2 and 3, but then did not between Sessions 3 and 4. They made up for it with large Performance improvements between Session 4 and the Retention Session. The Continuous Practice Group made the fewest gains in Performance over the training sessions, improving the least between Session 1 and 2, and with no significant improvements in Performance between Sessions 2, 3, and 4. They did improve between Session 4 and the Retention Session, allowing them to "catch up" to the other group's performance.

Our measures of subjective experience indicated that introducing spaced practice made only minor changes to the players' experience. We found differences on only one measure—enjoyment—and only when comparing the Continuous Practice Group to the *2-Minute* Interval Group after the training sessions. The Continuous Practice Group experienced less enjoyment than the *2-Minute* Interval Group.

8.7 Discussion

In the following sections we look back to our three main research questions, consider possible explanations for certain results, outline design implications for game designers, and discuss limitations and future work.

8.7.1 Returning to the Research Questions

Is spaced practice beneficial for game skill development?

Our first research question asked if there are benefits of spaced practice on performance in digital games and our third asked whether performance benefits last beyond the training session. Our results clearly show that taking breaks between practice sessions does improve the development of game skills (in our *Super Hexagon* clone) from session to session—all of the interval groups improved more than the continuous-play group (who did not make significant progress through the entire twenty minutes of practice). Furthermore, we found this result despite the possibility that participants were not using these breaks to actually rest—they could have simply switched to a different task—just as actual gamers might use the break.

However, despite these differences during practice, the results of our retention test suggest that the continuous group's lack of improvement did not mean that they failed to learn how to play the game. After one day, the continuous group saw a large (and significant) jump in performance, and the performance differences between all of the groups largely disappeared. As discussed further below, both of these results can be usefully exploited by designers who want to support improvement in either the short term or the long term.

Is there an optimal rest period for our game?

The idea that a particular task has an optimal rest period is fairly common [81, 48, 194, 39, 186]; however, there was no obvious best interval in our results. The 10-minute interval group made large initial gains, but their rate of improvement decreased at the end of the training period, and their retention-test performance was not significantly better than the other groups. The 2-minute interval is a second candidate for an optimal rest period: this group made consistent improvements, and rated their enjoyment higher; the other advantage of the 2-minute interval is that it may be easier to impose on players (either explicitly, or implicitly through game design). The 1-day interval group saw the largest overall gain from the first session to the retention test—but this is likely too long a break to be feasibly implemented. It is unlikely that players will be willing to stop playing after five minutes and then wait an entire day before returning to the game.

An interesting idea proposed in previous work is that the rest interval should change as the learner gains experience. For example, Snoddy [277] thought that the rest interval in general became less effective with continued practice; and Kim et al. [164] suggest that the interval should be tuned to the stage of learning. In this approach, the learner needs longer breaks during early stages when they are primarily dealing with declarative knowledge (in the cognitive stage), but shorter breaks, or none at all, when they are starting to compile their knowledge into procedures (in the associative stage) [164, 10, 213].

8.7.2 Explanation of Results

Although our results are focused on outcomes rather than explanatory power, there are several issues that can help researchers understand some of our findings:

- *Why did the continuous group perform poorly?* Learning theories that promote rest intervals suggest that the main problem with continuous practice is that there is no time for the brain to generalize the feedback that has been gathered. This idea seems particularly relevant to games—many players have experienced the feeling of being "stuck in a rut" when trying to solve a particular problem.
- *Why did the 10-minute interval group stop improving?* As described above, there is debate about whether the rest interval should remain the same at all levels of experience. It is possible that the 10-minute rest group was at a different learning stage by Session 3, and would have continued to improve if their rests had been shorter.

More research is needed to explore this issue further.

- *Why did performance equalize after the one-day break?* We believe that this improvement may simply be an example of the effectiveness of a different kind of break. That is, the one-day break may have allowed players from all the interval groups to generalize their practice experiences and compile declarative knowledge into procedures, leading to improvements on our retention test. Further study should examine the effects of different retention delays.

Finally, our experimental setting—and in particular, the motivation of our players—may have played a role in some of our results. For example, a player's willingness to take breaks (and potentially, the benefit that they receive from a break) may be strongly affected by their motivation to keep playing the game. For example, if one of our interval groups had more highly-motivated gamers who resented the breaks, it could have changed the performance relative to the other groups. (We note, however, that we included some of these traits as covariates in our analyses in order to try and account for their effects—but further analysis of these factors is needed).

8.7.3 Implications for Design

Our results have important implications for game designers and players. First, the during-training improvements seen are valuable because players are highly interested in short-term improvement (i.e., during their current game session). Our results suggest that players can make significantly more progress in learning a game by taking breaks as small as two minutes—and can see these results in a very short time (after one or two five-minute play periods). Second, the delayed improvement shown by the continuous-practice group means that players who dislike pausing can still get the benefits of spacing by taking a longer break after their play session (although the benefits will not be seen right away). This is valuable because some players wish to remain immersed in the game (e.g., as shown by the disagreement over the spawn locations in *Dark Souls*). These results, combined with little to no difference in subjective experience, mean that designers are not constrained to one or the other approach, and can even make alternate paths available for players with different interests.

Although our results show clear benefits of pausing during practice, some players may argue that it is more efficient to practice continuously, because the time spent resting could be used for additional attempts. We did not investigate this comparison directly; however, our data suggests that breaks may be as useful as additional practice, at least for initial learning. In our study, a single break for the 10-minute interval group allowed them to reach the same level of performance as the group that played continuously for 20 minutes. This issue needs to be explored in greater detail, but these early results underscore the suggestion that there are valuable processes going on in the brain during a break.

It is also important to note that despite the differences we saw in training, our retention test suggested that the continuous practice group's lack of improvement did not result in them failing to learn how to play the game compared to the spaced practice groups (i.e., performance for all groups was similar at the retention test). Therefore, the decision whether to play or pause may depend on the player's immediate goals. If the player is making steady performance improvements, then there may be no need to take a break. If, however, the player has reached a point in the game where their performance improvements have stalled, or they are up against a challenge that they cannot overcome, then a short break may be just what they need to continue to improve or overcome that challenge.

Finally, we allowed participants to do whatever they wished to during breaks; however, a game could be designed in such a way as to help a player to take a break from the psycho-motor task, but not leave the game environment. Inventory management, avatar upgrades, side quests, world exploration, loading screens, and cut scenes are all examples of activities that would allow for a break from the challenging task, but allow players to stay in-game. Designers already utilize these activities in various ways and there are opportunities for helping players improve performance through clever pacing switching activities to force breaks and optimize skill development.

8.7.4 Generalization and Future Work

One aspect of this work that could limit generalizablity to other games was our choice to use *Super Hexagon*. In contrast to some digital games with dedicated tutorial sections and complex gameplay mechanics, *Super Hexagon* requires new players to learn only two buttons and understand that the obstacles should be avoided. This difference in required declarative knowledge could affect how well spacing practice works or the underlying mechanisms that result in improved performance (e.g., allowing time to compile declarative knowledge into procedural knowledge [10, 164] or if fatigue is a factor [5]). Previous research on spaced practice suggests that the type of task moderates the effect (e.g., whether it is complex and requires cognitive skill [81], is discrete rather than continuous [175], or requires declarative knowledge [164]). These task factors likely apply to different types and genres of games as well. Therefore, future work could investigate this explicitly, testing games from different genres.

If this effect is robust and occurs in a variety of games, then further research could explore *why* taking a break works to improve performance. For example, a cognitive neuroscientist might be able to identify how neural systems respond differently to spaced versus continuous practice. Another possible approach is to vary what activity participants are doing during the break to understand what it is about the break that results in improved performance. These could easily be informed by game design. Consider, for example, how *Dark Souls* gives players an implicit break after each failure. Instead of forcing the player to step away from or exit the game, the player can be made to engage with the game in a less demanding way — in a role-playing game like *Dark Souls*, the activities could include looting, storytelling, or inventory management.

8.8 Conclusions

Practice to develop skill is an important part of playing a game, and the best way to practice is of interest to both game designers and players. Taking rest breaks during practice has been widely studied in learning research, but there is little knowledge of whether spacing aids the development of game skills. We carried out a study in which players completed five-minute practice blocks with either continuous play or variable spacing between the blocks (from two minutes to 24 hours). Our work provides valuable contributions that can change the way that game designers and game plays consider training:

- We show how spaced practice affects perceptual-motor tasks in a real game (a domain that has seen little research in practice) and in a naturalistic setting.
- We show that players can improve their short-term performance by taking breaks, something that many players essentially never consider, and that game designers can assist player performance simply by introducing a break.
- We show that "just playing through" is a bad approach (something many players do), especially when immediate performance gains would help them make progress.

Our results provide useful information for designers who want to encourage skill development, and for players who simply want to get better at their favourite game.

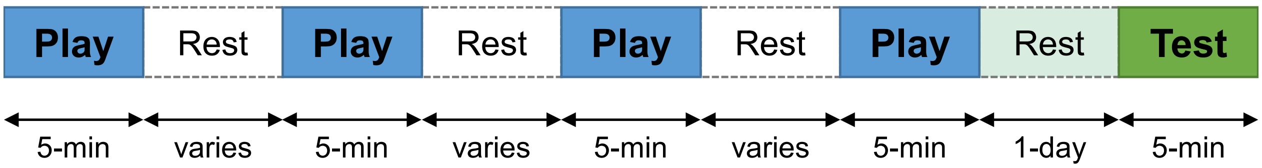


Figure 8.3: The procedure of the experiment.

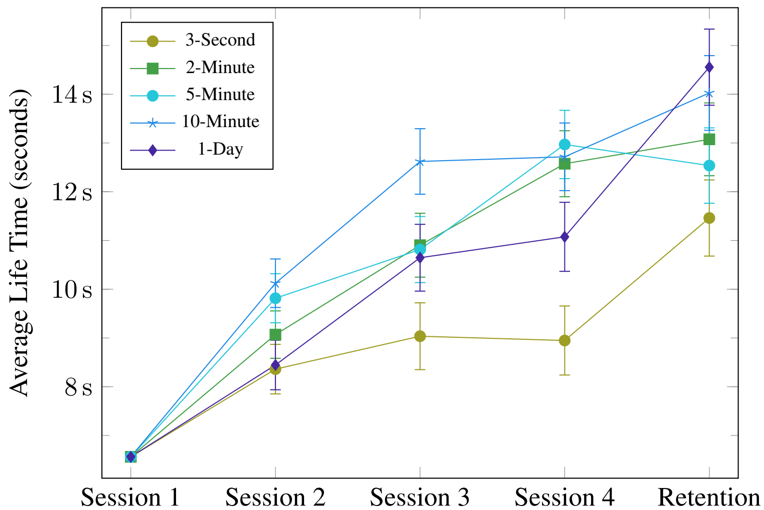


Figure 8.4: Improvements in average life time across all Sessions and Interval Groups, presented as the estimated marginal means from the RM-ANCOVA. Error bars show standard error.

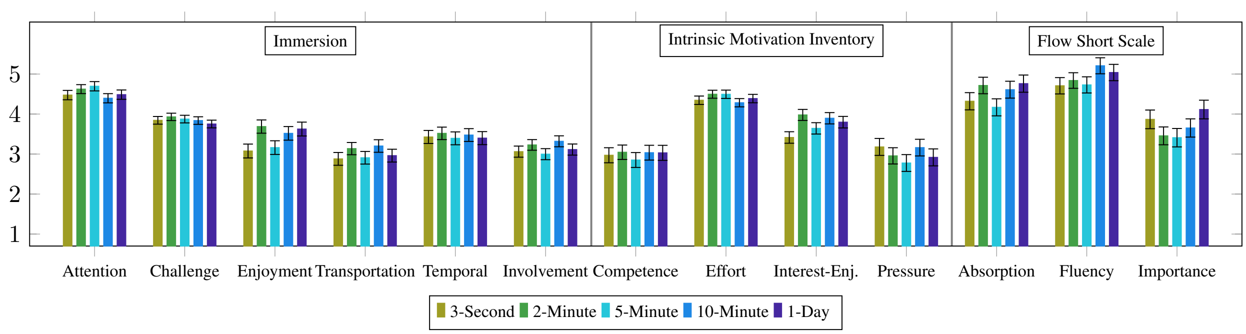


Figure 8.5: Subjective measures from the subscales of Immersion, Intrinsic Motivation, and Flow. Presented as the estimated marginal means of the combination of both Sessions from the RM-ANCOVA. Error bars show standard error.

9 Introduction to Manuscript D

Manuscript C showed that spaced practice is an effective method for improving one's immediate performance compared to playing continuously. In that manuscript, we proposed taking breaks as a means of overcoming in-game challenges. Even though spaced practice *seems* like it could be a good approach, there are still several issues to resolve to overcome before it can be applied within a game. In particular, at what times within the game should a break be presented to a player, and how long should that break be? Sadly, aside from my contributions in Manuscript C, the literature provides little guidance on this.

We therefore decided to try integrating spaced practice into a digital game in a novel way. Instead of introducing a break after a fixed number of attempts or after a certain amount of time has passed, we integrated a short (up to ten seconds) break after each failed attempt into the game. The reasoning for this was simple: If a break can improve immediate performance, then maybe it can also improve one's next attempt at overcoming an in-game obstacle. However, a two-minute long break (or longer) as used in Manuscript C makes little sense if each attempt takes only seconds rather than minutes. Therefore, we limited breaks to a maximum of ten seconds.

An issue that Manuscript C raised is the notion that taking breaks is not exactly a player's first thought and players are highly motivated to continue playing [247] so would rather make repeated attempts without delay. Manuscript C did not actually explore this, so this time we explicitly asked participants questions relating to how they felt about the breaks.

Because of this Manuscript's focus on helping players make *progress* within a game, we also used this as an opportunity to explore a game mechanic that is commonly found in games and directly affects how players make progress within a game. In older games, it was common for players to need to restart the game or level from the very beginning when they fail. For example, in *Super Mario Bros.* players had a fixed number of lives in the game and once they died enough times they would have to start the game over from the very beginning. Today's games commonly make use of checkpoints [115] or save systems that prevent players from losing progress. Like spacing practice, saving progress in this way has the potential to affect how a player learns and improves at a game.

For spacing practice, the prior manuscript showed that it positively affects performance, yet the work did not propose how to integrate spaced practice into the game apart from showing that a variety of break lengths were effective. This manuscript seeks to demonstrate that spaced practice is a robust effect and can be beneficial even when applied in novel ways; in this case, by introducing short breaks after each failed attempt. Saving progress also has the potential to modify the effectiveness of practice. It is likely that it allows players to make progress within a game more rapidly, improving the player's perception of performance, but players may not actually be learning the task — they may only be able to complete it inconsistently, but the saving of progress allows that inconsistent progress to be sufficient to make progress through the level. Furthermore, by aiding progress directly rather than aiding performance and learning, players may inadvertently be encountering parts of the game that are more difficult than they are prepared for. However, if these suspicions are false, then there is little reason not to systems within a game that automatically save progress.

The results of this manuscript are good news for players. We found that our dynamically integrated spaced practice with breaks no longer than ten seconds can improve performance over playing continuously, so players do not need to disengage from the game for very long to benefit from the effects of spaced practice. Secondly, checkpoints did positively affect performance while it was present and did not negatively affect learning — there was no apparent downside to making use of checkpoints.

9.1 Preregistration

The design of the experiment in this manuscript was preregistered on the OSF Registries. A reproduction of the preregistration is in the appendix, Section B.2.

Colby Johanson and Brandon Piller. 2020. PlayPause Checkpoints. https://doi.org/10.17605/OSF.IO/S6ZDT

9.2 Methodological Clarifications

9.2.1 Participant Payment

The study in this manuscript paid \$10.25 USD for completing the study which took 48 minutes on average to complete. This works out to \$12.8 USD per hour, which is more than the United States' federal minimum wage of \$7.25 an hour, where our participants were from.

9.2.2 Retention Session

For the retention session, I had participants wait one week before returning and attempting to play the game again. The reasons for this were the same as for Manuscript B — one week between sessions is a similar length of time between sessions to what may happen in the wild. I also wanted to test a longer interval than what was used within Manuscript C, to test spaced practice under different conditions.

In the manuscript, I did not compare the participants' testing performance to their retention performance in my statistical tests. This was because of the differing counts of participants in training compared to testing. However, unlike in Manuscript B, Study 2, there are no concerns about violating the assumption of sphericity due to only being one pair of measurements per condition. For this reason, in Section E.4, I present the within-subjects results from an RM-ANCOVA that compares between the training session and retention session.

9.3 Additional Analyses, Results, and Figures

When writing Manuscript D, there were some results and figures that did not make it into the final paper due to space constraints. These have been included in the appendix, in Section E.4. This includes:

- Charts showing the raw means for the performance measures.
- Statistical analysis of possible gender differences.

9.4 Individual Contribution

This work has been submitted to the Proceedings of the ACM on Human-Computer Interaction and is currently under review. This work was completed in collaboration with my supervisors, as well as with my colleague, Brandon Piller. My contribution to the work included: designing the experiment, modifying the digital game (that had been made previously for [231]), performing data analyses, and writing the majority of the manuscript.

10 Manuscript D
If at First You Don't Succeed: Helping Players Make Progress In Games With Breaks and Checkpoints

10.1 Abstract

Developing skill and overcoming in-game challenges is of great interest to both players and game designers. Players can improve through repetition, but sometimes practice does not lead to improvement and progress stalls. It would be useful if designers could help players make progress without compromising their long-term skill development. We carried out a study to investigate how two techniques—checkpoints and breaks—affect in-game progress and player skill. Checkpoints allow multiple attempts at a challenge without having to repeat earlier sections; this aids progress, but could potentially hinder skill development. Second, breaks in gameplay have been shown to accelerate skill development, but their effectiveness is unknown when the breaks are integrated into the game's design. Our study evaluated the effects of game-integrated breaks and checkpoints on players' in-game progress (when the techniques were present) as well as two test sessions (with all techniques removed). Our results showed that both checkpoints and breaks aid progress (combining both had the largest effect) and that neither technique reduced performance in the transfer task, suggesting that skill development was not hindered. Our work provides evidence that checkpoints and breaks are valuable techniques that can assist both player progress and skill.

10.2 Introduction

Many players want to play games that challenge them [157] — overcoming the challenges within a game is satisfying [167] and motivates players to continue playing [237]. Players and game designers therefore have a strong interest in understanding how players can improve their skills to overcome in-game challenges. Simple repetition (i.e., practice) is the main way this happens: repeatedly attempting a challenge allows players to learn from their mistakes and improve at the game [167, 114, 342, 143]. Although often a viable strategy, there are situations in which simple repetition is ineffective. In particular, players may feel that their rate of improvement is not fast enough, or they may reach a plateau and be unable to make progress, sometimes causing them to abandon the game [54].

Game designers can try and address this problem in two ways. First, they could focus on skill development, under the assumption that as skill improves, the player will also be able to make progress in the game. There are many potential strategies for improving skill development, many of which have been explored in the domain of perceptual-motor learning (a research area that encompasses many of the types of skills used in digital games) (e.g., [257, 81, 265, 105]).

Second, designers could focus on the player's in-game progress, and add mechanisms or assists that move the player along in the game. This strategy can ensure continued progress — but if the assist is too extreme or too artificial, it could compromise the player's longer-term skill development. For example, a powerup that boosts a struggling player over a challenge could prevent them ever learning the skills needed to overcome the challenge on their own.

In the first approach, there are several techniques that have been shown to improve skill development in previous work [257], including adjusting the variety of what the learner practices (e.g., practicing variations of the task to better handle novel scenarios [254]), introducing part-task practice (e.g., focusing on specific components of a skill by breaking a complex skill into components that can be practiced separately [107]), or adjusting the spacing of the practice sessions. Spaced practice adds short ***breaks*** to the practice session — that is, forced delays that interrupt practice. Taking breaks in games has been shown to improve performance over playing continuously [151, 231], but previous studies presented breaks on a fixed schedule rather than integrating them into the existing mechanics and presentation style of the game; this artificial presentation would not be acceptable in a real-world game. Breaks and pauses do occur in many games — e.g., as spawn timers, cut scenes, or mini-games — so it is possible that breaks intended for skill development could be better integrated into the game design. It is currently unknown, however, whether game-integrated breaks will continue to be effective in improving skill development.

In the second approach, there are many game mechanisms that could be used simply to assist progress, such as powerups, dynamic difficulty adjustment, or ***checkpoints***. Checkpoints create new restart points within a game that allow a player to save their progress — when players fail at a challenge, they restart at the most recent checkpoint [115] and can immediately attempt the challenge again. Checkpoints can aid player progress, but their effect on skill development is unclear. Checkpoints could hinder skill development by allowing players to avoid practicing skills in less challenging contexts, or by artificially moving players forward to a part of the game that is too difficult. However, checkpoints could potentially also help with skill development by allowing players to focus on the part of the game they struggle with (i.e., providing part-task practice), or by exposing players to a wider variety of game scenarios and skill demands (i.e., adjusting the variety of practice).

It is possible that combining the two approaches — game-integrated breaks and checkpoints — will result in even more progress being made in the game or affect skill development more than either modification on its own. Checkpoints on their own might prompt players to take risks, but the break could reintroduce a consequence by enforcing a pause after a failure. This might give players time to reconsider what they are doing.

To investigate these questions, we carried out a between-participants study in which people played a side-scrolling platform game. Participants were divided into groups who saw different versions of the game with different combinations of breaks and checkpoints, in a 2x2 design. Game-integrated breaks were implemented as variable pauses of up to 10 seconds after dying, and checkpoints were implemented as an automatic save mechanism using several fixed points in each game level. Participants played their version of the game for 20 minutes and then completed an additional

10-minute transfer session in which the game had neither checkpoints nor breaks. We additionally invited participants back after seven days to complete a retention session that had players play the game for an additional 10 minutes, using the same levels as in training but without breaks or checkpoints. The study assessed in-game progress by measuring how many levels the player completed (with the technique present); and because skill development can only be inferred by seeing changes to performance over time [257, 278], we assessed development through the player's performance in the transfer and retention sessions.

For this study we had four research questions:

- **RQ1:** For breaks, would pauses that are tightly integrated into the game design still improve a player's skill (i.e., translate into improved progress within the game)?
- **RQ2:** For checkpoints, would saving a player's progress help them make progress within the game?
- **RQ3 and RQ4:** Would having played with breaks (RQ3) or checkpoints (RQ4) have any effect on a player's skill development?
- **RQ5:** Would using both techniques have a larger effect on progress or skill development than using each individually?

For both techniques, the study also considered whether the technique detracted from or changed the play experience: for example, it is possible that players dislike forced breaks that prevent them from immediately re-attempting a challenge, and that players will view checkpoints as "false progress" that takes away from their sense of accomplishment.

Our study showed positive results for both game-integrated breaks and checkpoints as techniques for supporting player progress without hindering skill development. First, participants who had breaks in their game completed about three more levels than the baseline condition (which had no breaks and no checkpoints), and there were no reductions in performance in the transfer and retention tasks. This result provides evidence that breaks are still effective even when integrated into the game design. Second, checkpoints also allowed players to complete about three more levels than the baseline group, and performance in retention and transfer tests was again unaffected. This result indicates that a checkpoint's positive effects on progress do not come at the cost of reduced skill development. Third, players who had both checkpoints and breaks completed nearly eight more levels than the baseline, and were also no worse during transfer and retention, suggesting that combining the two techniques may provide added value.

In addition to our primary measure of progress, we also checked the effects of checkpoints and breaks on player deaths, and analysed the game logs to look for two patterns of inhibited progress: stalls (in which players get stuck at one part of the game), and regressions (in which players do worse than in past attempts). We found that player deaths were affected by the presence of breaks (likely due to players being more cautious because of the after-death pause). We also found that both checkpoints and breaks helped players overcome both types of inhibited progress: regressions occurred fewer times, and players were stalled for less time.

Finally, our player experience measures found mixed results: there were no differences in player experience for people who had checkpoints, but we did find differences in subjective flow, curiosity, and meaning for groups who had breaks. Players who started with breaks reported improvements in these measures once they moved to the transfer and retention sessions (which had no breaks). Subjective preference questions showed that none of the participants liked the breaks — but several people noted that they used the time to reconsider what they were doing in the game, suggesting that the breaks may have been having the intended effect.

Overall, our findings provide strong evidence that both techniques — breaks and checkpoints — can be successfully used to support players' progress in digital games without hindering skill development. Even when integrated into the game mechanics and presentation, breaks successfully improved progress; and even though players stated that they disliked the breaks, there were no negative effects on game-experience measures. Checkpoints allowed players to progress further without any negative effect on skill development, and were liked by participants. Our results give game designers a new understanding of how they can support a player's progress and help them improve at the game using techniques that can be woven into existing game mechanics and presentation without substantially compromising the play experience.

10.3 Related Work

10.3.1 Spaced Practice in Games

Spaced (or distributed) practice is the concept of scheduling periods of rest to break up periods of work within a training session [257]. Compared to continuous (or massed practice), adding breaks generally results in strong gains to short-term performance and slightly weaker gains to long-term performance (i.e., learning), as measured via transfer tests (testing participants on a very similar but different task) or retention tests (re-testing participants after some period of time, such as a day or a week) [174, 81, 257].

Spaced practice has been found to benefit player performance in games over four experiments [200, 268, 151, 231] and two analyses of data-sets [282, 281]. This work demonstrates that benefits of spaced practice can be had within experiments lasting less than an hour [200, 151, 231], to ten hours [268], or over periods of weeks or longer [282, 281]. It has been observed in games as simple as *Breakout* [200] or as complex as *Destiny* [281].

Because breaks improve performance, introducing breaks could potentially help players make progress and overcome a game's challenges. However, no past work has provided guidance on how to integrate spaced practice into a game. Past work introduces breaks on a fixed schedule and involves rests of two minutes or longer (e.g., [151, 231]). However, spaced practice is defined in relative terms rather than absolute terms [257] — there is no one accepted schedule for the timing of the rest periods relative to the work periods, the times for the rest and work period need not be fixed, and trials might be used to determine spacing instead of fixed time periods [257, 321]. Additionally, the rest period does not need to be as long as two minutes; breaks as short as 15 seconds have been found to be beneficial [303, 132]

10.3.2 Checkpoints

Very little prior work in digital games has studied the effects of checkpoints or save states on performance, learning, or progress. One paper examined the relationship between the frequency of saving progress (with one implementation using checkpoints) and several player experience measures, and found that the frequency of deaths was related to perceived challenge [64]. Because of the relative lack of prior work, however, the remainder of this subsection considers ways that checkpoints could affect a player's experience of the game and how this changed experience might affect performance and learning.

Skill-Challenge Balance

Practice is often an effective method of skill development because a game's difficulty increases in line with their progress [167, 114, 342, 143, 296]. To facilitate continued learning, challenges should be challenging but within or at the edge [325] of a player's abilities. Players encountering challenges of this difficulty feel that they can overcome them [158] and are motivated to do so [114]. If checkpoints help players progress through the game they will be more quickly taking on tough challenges — challenges they might not be prepared for. This could negatively affect learning.

Practice Conditions

Checkpoints may alter practice conditions in two ways. First, if checkpoints help players make progress then they may affect the *variability of practice* — they become exposed to more levels and more variations of the game's mechanics. Second, checkpoints may introduce *part-task practice* by allowing players to attempt a challenging obstacle without needing to consider the ones that come before — they can focus on overcoming an obstacle they struggle with.

Increased variability in practice is thought to promote motor learning by strengthening a learner's ability to cope with novel conditions [257, 265]. These benefits are subject to factors such as the nature of the task (e.g., benefits of variable practice are found most commonly with simple tasks) and the expertise of the learner (e.g., increased variability may be most beneficial for early learning) [350]. Increased variety typically increases errors [185], leading to a short-term drop in performance which would be undesirable in games. In games, players would experience this loss of performance alongside making progress through the game, so they may struggle more when taking on harder obstacles. The previously mentioned data analysis [281] of the online game, *Destiny*, looked at variability in terms of the propensity for "social play" and playing style, but did not find that increased variability enhanced skill acquisition.

Part-task practice is thought to be beneficial for complex skills [185, 105], as it allows a learner to focus on aspects that need improvement [211]. When players are repeatedly going up against a challenging obstacle, checkpoints may let them focus on and refine a subset of skills. Past work has found part-task practice to be effective for skill learning in digital gaming contexts [107, 118].

10.4 Modifying Game Practice with Breaks and Checkpoints

Modifying in-game practice by introducing breaks and checkpoints is something that some games already do. Sometimes this is done explicitly, while other times it is an implicit consequence of the design of the game or its systems. In this section, we discuss how these systems are presented within commercial games.

10.4.1 Breaks

In digital games, spaced practice (i.e., breaks) is often included incidentally. For example, breaks may exist in games due to technical limitations — loading screens, in particular, are commonly found in many games because of the delays encountered when loading content [242] (although these breaks may become shorter or less common as hardware improves). Other examples include players waiting for others to connect in a multiplayer game, breaks from the game's core mechanics in the form of mini-games [231], or menu systems for activities such as inventory management. Breaks also may be implicitly introduced in games simply by having enemies spaced apart from one another, with the travel time between battles acting as a break.

Although previous studies provide evidence for the benefits of breaks, there is little guidance from past work in terms of integrating them into the game. Past research presented breaks to players on fixed schedules, which could interrupt a player in the middle of play. This could be difficult to integrate into the design of many games. Less of an interruption would be preferred and could be accomplished if the break was integrated into the events of the game. Death is a common game mechanic and one where, incidentally, breaks can already be found (e.g., when waiting to respawn in an online game). Considering that breaks have been found to improve performance, this may be an ideal time to provide a break — players already doing well will receive fewer breaks whereas players who are struggling receive more breaks, giving them more opportunities to benefit from spaced practice.

If breaks *are* integrated into the game explicitly, then it is usually as a reminder to take a break. Many games prompt players to take a break after some time has passed. While this is not done for the benefits of spaced practice (it is to reduce the likelihood of seizures and repetitive strain injuries [218]), it is a common way that players are presented with the idea of taking a break. A glance at online discussions quickly reveals players who mock the suggestion or go out of their way to keep playing in response [313]. In general, the idea of taking breaks to get better goes against many players' desires or intuitions — those who might believe that if they just keep trying, they will eventually succeed. However, players do accept breaks in certain contexts, for example, the need to wait is common within multiplayer games as players wait for a new round to start, or wait to be connected to other players.

10.4.2 Checkpoints

In contrast to breaks, checkpoints support a player's desire to get back to a challenge right away. Checkpoints save a player's progress so that players can start playing again from the checkpoint when they fail [115]. These saves are typically strategically timed, such as before difficult segments or only when specific conditions are met. Checkpoints

Figure 10.1: The game used in the experiment. Pictured is the player using the grapple hook to swing toward the checkpoint (the yellow flag).

are generally accepted by players, although they occasionally aggravate players when saving occurs at an inconvenient time [168].

An alternative approach to checkpoints is allowing players to choose when to save the game via save states. When taken to the extreme, some players consider it "cheating" (e.g., [312, 314]), as players can continually save the game and prevent any amount of progress from being lost. Players opposed to this argue that players doing this will not learn the game and will rely on the aid [329, 312]. Other players point out that strategic use of saving can help with learning because it can allow for quick trials of different approaches [314] and can help one learn difficult segments of the game [312], or specific enemy patterns [329].

Mechanisms that allow players to save their progress have become more prevalent over time. In the original *Super Mario Bros.* (released in 1985), players must start over from the beginning of the level if they die. With *Super Mario World* (released in 1990), each level had a checkpoint at the midpoint. Many modern titles save progress very frequently or design their levels to be small enough to not require checkpoints. In *Super Meat Boy* (released in 2010) the levels are designed to be short enough to not require checkpoints and in *Celeste* (released in 2018) each level fits on-screen without scrolling and the game is saved at each screen transition.

10.5 Methods

10.5.1 The Game

We created a bespoke 2D side-scrolling platform game for our study (Figure 10.1), similar to a game used in past work on spaced practice [231]. The game was inspired by games such as *Super Mario World* [220], *SpeedRunners* [82], and *Super Meat Boy* [291]. Our aim was to create a set of mechanics that were easy to understand, but with enough nuance that they would take time to master. While the platforming genre would be familiar to players, our specific implementation of the mechanics and the design of our game levels would not.

The player controls a lumberjack character, moving horizontally with the arrow keys, jumping with the space bar, and pressing and holding "E" to swing with a grapple hook. Horizontal movement includes acceleration and deceleration, and players must anticipate this to avoid overshooting. Pressing the jump button allowed variable upward acceleration depending on the duration of the press, so players must time both the start of a jump as well as the duration. The jump button also allows wall-jumping. Pressing and holding "E" when a grapple location is in range initiates a swing action; the player holds the button down until they want to let go of the rope, and the speed and direction of the swing can be adjusted mid-swing with the arrow keys. Players can also exit the swing with a jump action.

The game mechanics were introduced to players through a 47-second video demonstration and through three in-game tutorial levels; these included in-game sign posts that indicated what keys needed to be pressed and were designed so that players were required to perform each skill multiple times to continue. After the tutorials, the game levels increased in difficulty as the player made progress by gradually presenting the player with more frequent or more complex obstacles. If players ever became stuck and were unable to finish a level after 3 minutes, they were given a button that they could click to skip the level.

The game was developed in Unity and presented to participants on a website via WebGL.

Experimental Factors

Participants of our experiment were assigned to one of four treatment groups, formed by crossing two factors (checkpoints and breaks) in a 2x2 design.

Checkpoints: If participants experienced the game with checkpoints, flags were placed throughout the level that saved the player's progress whenever the flag was passed. Flags were positioned between groups of obstacles within the levels. Then, if they died within the level, they would start over at the checkpoint rather than at the beginning of the level. A screenshot showing a checkpoint flag is shown in Figure 10.2.

Game-Integrated Breaks: If participants experienced the game with game-integrated breaks, then upon death they were made to wait up to ten seconds before another attempt, with an onscreen countdown timer showing them the time remaining. The length of the break was equal to the time they were alive before their death (up to a maximum break of ten seconds) — more time alive meaning a longer break. Ten seconds was chosen as a maximum based on feedback from pilot testing (and to avoid situations where a player spends as much time pausing as actually playing). If players did not receive breaks, they still were made to wait for one second before their next attempt, to prevent accidental deaths from unintentional inputs. A screenshot showing the countdown timer during a break is shown in Figure 10.3.

10.5.2 Procedure

After giving consent, participants were assigned to one of the four treatment groups. Participants then watched a video demonstrating how to play the game before responding to a questionnaire relating to their current motivation toward

Figure 10.2: Screenshot of the game showing a checkpoint flag.

Figure 10.3: Screenshot of the game showing the countdown timer for a game-integrated break.

the task and how they viewed their ability to complete it. Participants then played the game, starting with the tutorial levels. They played the game for a 20-minute *training* session, in which they played the game with the techniques that were specified by their group in the 2x2 design.

After this training session, participants responded to several questionnaires relating to their subjective experience of the task, and a questionnaire relating to demographics. They then played the game again for a 10-minute *transfer* session, in which they played a new set of levels without checkpoints or game-integrated breaks. After, they responded to further questionnaires relating to their subjective experience and two more questionnaires relating to individual differences.

Participants who completed the training session fully were then invited back one week later to complete a *retention* task. Each participant renewed consent, and played the game for 10 minutes, with the same levels and no checkpoints or game-integrated breaks. They then responded to questionnaires assessing their experience and could provide feedback once again.

The game and all questionnaires were presented to participants via a website built using an existing web framework designed to aid the creation of online studies [150].

10.5.3 Measures

We used a combination of questionnaires and data generated from in-game actions to measure differences between participants and their experience with the game.

Individual Differences

To account for individual differences that could alter a participant's in-game performance or subjective experience of the game, we used the following measures:

Gaming and Platforming Familiarity: We expected that prior experience playing platforming games might affect performance and subjective experience of the game. Therefore, we asked six questions, each presented as a slider from

1 ("Not at all") to 100 ("Very familiar", or "Gamer"). These questions were: *"How much do you self-identify as a gamer on the following scale?"*, *"How familiar are you with side-scrolling platform games?"*, *"How familiar are you with Super Mario games?"*, *"How familiar are you with the game 'Super Meat Boy'?"*, *"How familiar are you with the game 'Speedrunners'?"*, and *"How familiar are you with the game 'Celeste'?"*. The measure was the mean of responses to all six questions. These games were chosen because they are popular and because they feature one or more of the mechanics used within our game. In particular, the grapple mechanic in our game was directly inspired by and modelled on *Speedrunner*'s grappling hook, and during development, our game's movement controls were compared against *Super Meat Boy* in an attempt to replicate the feel of that game's controls for greater external validity.

Attentional Control: We expected that a participant's ability to give our game their complete attention might affect their performance and subjective experience of the game. We therefore used Derryberry and Reed's [74] Attentional Control Scale (ACS), which provides a self-report measure of *attentional control* — which relates to an individual's ability to focus on a specific task and shift their attention away from potential threats.

Current Motivation: We thought that a participant's initial motivation upon beginning the game could affect their ability to learn and improve at the game. Therefore, we used Rheinberg et al.'s [240] Questionnaire on Current Motivation (QCM) as a self-report measure of *task-related anxiety*, *probability of success* at the task, *interest* in the task, and *perceived challenge* of the task.

Achievement Orientation: We thought participants who were more competitive or interested in winning might put more effort into learning the game. Therefore, we used Gill and Deeter's [116] Sport Orientation Questionnaire (SOQ) as a self-report measure of *competitiveness* (i.e., the overall desire to meet a standard of excellence or compare favourably to competitors), *win orientation* (i.e., the importance of outperforming the competition), and *goal orientation* (i.e., the importance of achieving specific performance goals).

Initial Performance: Because participants who played our game would have different levels of prior experience playing platform games, we measured their initial performance in the three tutorial levels (where no Checkpoints or Game-Integrated Breaks were included) to further account for individual differences in prior platform game experience.

Total Training Time: Participants would end up actively engaged with the game for different lengths of time in training due to how long they ended up waiting. Therefore, we calculated the total time spent training by subtracting the wait time.

Outcome Measures

Our dependent outcome measures were chosen based on the goal of measuring in-game progress and exploring how players' subjective experience of the game changed due to training with checkpoints or game-integrated breaks.

Progress: As an objective measure of player progress within the game, we logged the *number of levels completed* in each session (training, transfer, and retention). To understand how players' in-game behaviour changed in the presence of checkpoints or game-integrated breaks, we logged and used the *number of deaths* in each session as a dependent measure.

We used two published questionnaires to measure players' subjective experience of the game.

Flow: We used Vollmeyer and Rheinberg's [324] Flow Scale Short (FKS), which measures all aspects of *flow* (i.e., challenge-skill balance, merging of action and awareness, unambiguous feedback, concentration on the task, time transformation, and fluency of action), combined into a single measure. Situations that lead to the joyful experience of flow can also lead to *worry*, and so the questionnaire measures this as well.

Player Experience: We used Vanden Abeele et al.'s [1] Player Experience Inventory (PXI) to assess several different aspects of player experience. The PXI measures player experience in terms of *functional consequences* (i.e., arising directly due to the game's design), as well as *psychosocial consequences* (i.e., second-order emotional experiences). Functional consequences consist of *ease of control*, *progress feedback*, *audiovisual appeal*, *goals and rules*, and *challenge*. Psychosocial consequences consist of *mastery*, *curiosity*, *immersion*, *autonomy*, and *meaning*. We measured the psychosocial consequences as well as the functional consequence of *challenge* after each session.

Written Responses: We also asked participants to respond to several open-ended questions, once after training and again after the test session. The questions presented after training changed depending on which version of the game they played.

If the participant played the version of the game in which checkpoints were present, then we asked them the following questions:

- "*How did you feel about the checkpoints in the game? Were there too many or too few?*"
- "*Do you think that the checkpoints made the game easier?*"
- "*Did knowing that you would re-start at a checkpoint affect how you played the game?*"

If the participant played the version of the game without checkpoints, then they were asked:

- "*When you died you had to walk back to that point in the level. Did you find this process easier than the part of the level where you died?*"
- "*Was there anything in particular you focused on or thought about while you were walking back?*"
- "*Did knowing you would need to walk back if you died affect how you approached the game?*"
- "*How did you generally feel about starting each level from the beginning after each death? Do you have any further comments on this?*"

If they played the version of the game where they had to wait to respawn, we asked them:

- "*How did you feel about waiting to play the game after each death? Was the time too short or too long?*"
- "*When waiting to play, what did you do with your time?*"
- "*Did the possibility that you would need to wait to try again change how you approached the game?*"

If they played the version of the game where they restarted instantly, then we asked:

- "*Did you take any breaks while playing? (Either intentional or unintentional.)*"
- "*Do you think you would have benefited from taking a break while you played?*"

- "*Do you think the ability to attempt the level again immediately affected how you approached it?*"

After the transfer session, in which checkpoints and game-integrated breaks were not present, we asked participants to compare the two versions of the game: "*Of the two versions of the game you played, which did you prefer?*", and "*Of the two versions of the game you played, did you find one of them to be more difficult?*". Finally, we gave them one last opportunity to provide general comments.

10.5.4 Recruitment

Because our game was designed to be somewhat challenging and we wanted participants to be able to make progress in the game, we were interested in recruiting participants who were already experienced gamers. All participants were recruited on Amazon's Mechanical Turk, so we first recruited participants to complete a task that simply had them respond to the question "How much experience do you have with playing side-scrolling platform games?" by dragging a slider to a value between 1 ("None") and 100 ("A great deal"). Participants who specified a value of 60 or higher were then assigned a "qualification" that would allow them to see the invite to the experiment.

Because we expected that self-report measures of experience might not be the best predictor of performance in our game, we advertised our task as involving as little as 5 minutes of gameplay or as much as 30. Participants would then be given a maximum of 5 minutes to complete the first three levels which served as an in-game tutorial to introduce the game's mechanics. If they could not complete the levels in that time, they were redirected to the end of the experiment and paid $1.75 USD. If they completed the levels within 5 minutes, the game would continue uninterrupted for the full 20 minutes, and would further go on to complete the full study, at which point participants were compensated with an additional $8.50 USD.

Participants were assigned to whichever group had the fewest participants at the time they gave consent. All participants were at least 18 years of age and had an average approval rating of at least 97% and more than 500 HITs (human intelligence tasks) completed. The design of this study received ethical approval from the behavioural research ethics board of the first author's university.

10.5.5 Participants

We estimated the participant count by using an a-priori power analysis in G*Power [100] with the following parameters: .25 effect size, alpha of .05, power of .80, numerator df of 1, 4 groups, and 6 covariates. The power analysis estimated that 128 participants would be required. A total of 190 participants completed our study. Of these, we filtered out a total of 40 participants: 8 due to attempting the tutorial levels more than once, 15 due to having too low of a framerate to properly play the game (<30 frames per second), 4 due to the age they entered (<18 or >90 years), 5 due to not completing any levels beyond the three tutorial levels, and 8 due to not completing any of the levels on the transfer test. This left us with 150 participants, with an average age of 33.5 years (min=18, max=64, SD=7.79). 101 identified as men, 48 identified as women, and 1 identified as non-binary. 38 played without checkpoints or breaks, 37 played with checkpoints but no breaks, 38 played with checkpoints and breaks, and 37 played with breaks but no checkpoints.

After one week, we invited back the 150 participants who were not filtered out. Of these, 117 completed the retention task: 28 had played without checkpoints or breaks, 29 had played with checkpoints but no breaks, 30 had played with checkpoints and breaks, and 30 had played with breaks but no checkpoints.

Our participants were experienced gamers and identified as such (81.1 out of 100; SD=21.1). They had quite a bit of familiarity with side-scrolling platform games (89.2 out of 100; SD=13.7), as well as with *Super Mario* games (93.1 out of 100; SD=11.2). However, they were less familiar with *Super Meat Boy* (47.5 out of 100; SD=38.6), *Speedrunners* (37.1 out of 100; SD=35.8), and *Celeste* (36.3 out of 100; SD=36.4).

10.5.6 Data Analyses

The data and associated results include confirmatory analyses and three different kinds of separate additional analyses.

Confirmatory Analyses

To verify that spaced practice positively affected performance in our platforming game, we computed individual ANCOVAs (analysis of covariance) for each outcome measure. Because of our 2x2 design, we had two between-subject factors: Breaks and Checkpoints.

Attentional Control was used as a covariate for every ANCOVA due to there being a significant difference between the groups ($F_{3,146} = 3.39$, $p = .020$), as evaluated by a one-way ANOVA with the trait measure as the dependent variable (all other trait measures were evaluated in the same way, but were not significant). For the progress measures (levels completed and total deaths), Total Training Time was used as a covariate to account for the differing amount of time actively playing the game, and Tutorial Completion Time was used as a covariate to account for individual differences in initial performance at the game. Further covariate selection was made based on whether each correlated with the specific outcome measure being tested and are reported alongside the results of the ANCOVAs. These covariates are presented in Section 10.6.4.

Jamovi [294] was used for all quantitative analyses. Alpha was set at .05. All pairwise comparisons used the estimated marginal means and Bonferroni corrections. An ANCOVA was chosen because when group sizes are equal, it is robust to violations of normality and homogeneity of variance [101]. Our group sizes were similar (see Section 10.5.5); however, we also inspected the data to ensure normal distributions and homogeneity of variance before proceeding. An ANCOVA is not robust to violations of independence [101]; however, our between-subjects experiment design ensures that observations across groups are independent.

Additional Analyses

Our second set of results are are not based on specific hypotheses. In addition to the quantitative results for our subjective measures, we also analyzed written responses to open-ended questions, and visualizations generated by processing game logs.

For the qualitative measures of subjective experience, we asked participants to complete the Flow Scale Short (FSS) [324] and the Player Experience Inventory (PXI) [1] after each session (Training, Transfer, and Retention).

Because the PXI measures functional consequences (i.e., consequences that result from the design of the game) as well as psychosocial consequences (i.e., emotional experiences that result from playing the game), we chose to measure the psychosocial consequences after every session while also including the functional consequence of challenge, as this could conceivably change due to Checkpoints, Breaks, or the slightly different levels played during the transfer session. For all repeated measures (Challenge, Mastery, Curiosity, Immersion, Autonomy, Meaning, Flow, and Worry), we used separate repeated-measure ANCOVAs, with Session as a within-subjects factor, and Breaks and Checkpoints as between-subjects factors. Attentional Control was included as a covariate (for the same reason as the confirmatory analyses). All post-hoc tests used Bonferroni corrections. Because our analyses are exploratory, we focus on significant effects only.

With the open-ended responses, we explored players' perceptions of game-integrated breaks and checkpoints. The responses were analyzed by a thematic analysis [36] performed by two of the authors. The two authors worked together to generate initial codes and group them into themes and then each author worked independently, coding all of the responses. After, the two authors met up to discuss and resolve all discrepancies. For brevity, instead of reporting the results of a full qualitative analysis, we highlight key findings that help to explain our results as well as highlight how players feel about these two mechanics. Any reported percentages are the percentage of responses that reflects the given theme unless otherwise stated.

Participants were always asked for written responses after training and again after testing. It is possible that biases related to recall or peak-end experiences may have affected their responses; however, these biases would have influenced each experimental condition similarly, and thus we do not expect to see systematic differences in qualitative responses based on which version of the game was played.

To better understand the ways that players made progress as they played the game, we generated visualizations from logs of their progress. Throughout the levels, we placed invisible triggers that we used to track progress throughout the level (shown in Figure 10.6). These were placed before and after obstacles where players would likely die. We also used the checkpoints to trigger a log of the player's progress; if the player was playing the version of the game without checkpoints, this trigger would still occur, but the checkpoint itself would be invisible and deactivated.

10.6 Confirmatory Results

10.6.1 Were Game-Integrated Breaks Beneficial? (RQ1 and RQ3)

We hypothesized that the short game-integrated breaks would positively affect the progress made in training and that skill development — progress in retention and transfer — would be unaffected. During training, we found a significant main effect of Breaks on the adjusted number of levels that players completed (see Table 10.1), after accounting for individual differences — players were able to complete more levels with Breaks. We similarly found a significant main effect of Breaks on the adjusted number of deaths.

These breaks did provide players with an opportunity to rest during training. Participants who were assigned to train with the version of the game that included breaks waited an average of 274 seconds throughout training (SD=68.4,

Measure	Session	df	Checkpoints				Breaks				Checkpoints*Breaks			
			F	p	η_p^2	Δ	F	p	η_p^2	Δ	F	p	η_p^2	Δ
Levels Completed	Training	1, 139	**38.8**	**<.001**	**.218**	**3.98 ± 0.64**	**5.5**	**.020**	**.038**	**3.74 ± 1.59**	3.0	.084	.021	7.72 ± 2.09
	Transfer Test	1, 141	0.4	.533	.003	0.36 ± 0.57	3.6	.060	.025	2.69 ± 1.42	2.9	.092	.020	3.05 ± 1.84
	Retention Test	1, 110	0.4	.535	.004	-0.39 ± 0.63	1.5	.224	.014	1.88 ± 1.54	1.07	.303	.010	1.49 ± 2.00
Death Count	Training	1, 142	1.1	.293	.008	2.46 ± 2.33	**78.3**	**<.001**	**.209**	**-51.4 ± 5.81**	**84.2**	**<.001**	**.372**	**-48.9 ± 7.55**
	Transfer Test	1, 143	**4.9**	**.028**	**.033**	**-3.57 ± 1.61**	**11.8**	**<.001**	**.076**	**-13.7 ± 3.98**	2.7	.105	.018	-17.3 ± 5.19
	Retention Test	1, 109	**6.2**	**.014**	**.053**	**-4.54 ± 1.82**	2.0	.157	.018	-6.52 ± 4.50	0.4	.542	.003	-11.0 ± 5.82

Table 10.1: Results of statistical analyses for our measures of in-game progress and deaths, showing the main effects of Checkpoints and Breaks, as well as interaction effects between Checkpoints and Breaks. Each row contains the results of a separate ANCOVA. The Δ columns show the mean difference between having Checkpoints or Breaks (or the combination) and not having them, with ± indicating standard error.

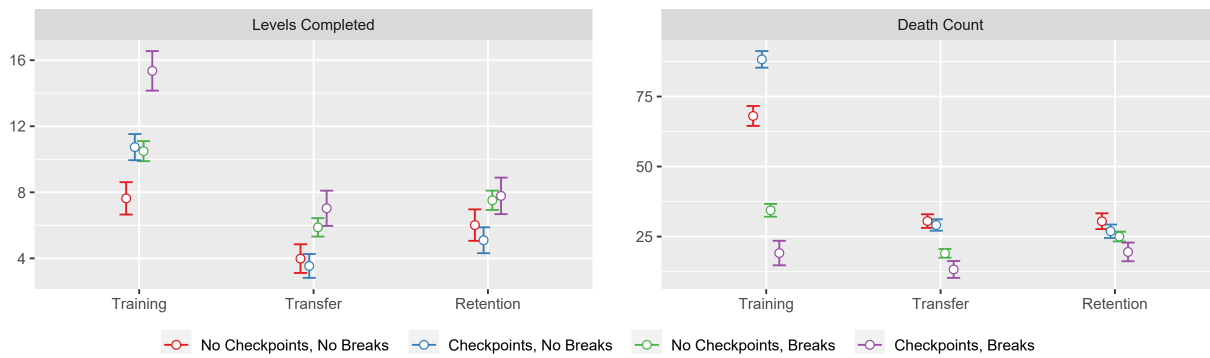


Figure 10.4: Estimated marginal means of levels completed and death count, from the ANCOVAs. Error bars are standard error.

Min=124, Max=451), compared to 61.2 seconds for those who played the game with a fixed 1-second wait after each death (SD=18.8, Min=29, Max=108).

During the immediate transfer test, in which we had all participants play the game without Breaks, with a new set of levels, we did not find a main effect of Breaks on the adjusted levels completed (see Table 10.1). We did, however, find a main effect of Breaks on the adjusted number of deaths — training with breaks resulted in fewer deaths.

During the delayed retention test, we again had all participants play the game without Breaks, but this time with the same levels as in training (except the early tutorial levels). We found no main effect of Breaks on adjusted levels completed (see Table 10.1). We also did not find a main effect of Breaks on the adjusted number of deaths (see Table 10.1).

10.6.2 Were There Any Drawbacks to Adding Checkpoints? (RQ2 and RQ4)

We hypothesized that checkpoints would help players make progress in training, and not affect skill development — progress in transfer and retention. In training, we found a significant main effect of Checkpoints on the adjusted number of levels completed, accounting for individual differences (see Table 10.1) — players completed significantly more levels with Checkpoints. There was no main effect of Checkpoints on the adjusted number of times players died.

During the transfer test, in which all participants played the game without Checkpoints on a new set of levels, we did not find a main effect of Checkpoints on levels completed, but we did find a main effect of Checkpoints on death count — they died fewer times if they trained with Checkpoints. For the retention test, we found the same; there was no significant main effect of Checkpoints on levels completed, but there was a significant main effect on death count.

10.6.3 Were There Any Interactions Between Breaks and Checkpoints? (RQ5)

We hypothesized that the combination of checkpoints and breaks would positively affect progress beyond what either checkpoints or breaks would on their own in training, but not affect transfer or retention. This was explored by looking for significant interactions between Checkpoints and Breaks. There was just one significant interaction effect between Checkpoints and Breaks, affecting the number of deaths during training (see Table 10.1). All post-hoc pairwise comparisons for the interaction were significant ($p < .001$). Examining the estimated marginal means (Figure 10.4), the interaction indicates that the introduction of Checkpoints affects the death count differently depending on whether Breaks were also present. When Breaks were present, including Checkpoints resulted in an *increase* to the death count. However, without Breaks, the inclusion of Checkpoints results in fewer deaths.

No other interactions were significant.

10.6.4 Individual Differences and Covariates

We corrected for individual differences between participants using covariates. The effects of these covariates are reported in Table 10.2.

Our measures of individual differences of attentional control, perceived probability of success, and perceived task challenge were not significantly related to levels completed or death count, however, task-related anxiety was significantly related to levels completed in Training and Transfer, as well as death count in Training. Platforming game familiarity was significantly related to all measures for which it was included.

We found that total time spent training was significantly related to levels completed during Training only, as well as death count for all sessions. Time spent on the tutorial levels was significantly related to levels completed in Training as well as both levels completed and death count in Retention.

10.7 Additional Results

10.7.1 Quantitative Subjective Experience

We found a significant interaction between Session and Breaks for Flow ($F_{2,228} = 7.5$, $p < .001$, $\eta_p^2 = .062$). From post-hoc tests, we found that Flow increased between the Training session and Transfer session ($p < .001$) as well as between the Training session and Retention session ($p < .001$) for participants who trained with Breaks.

For Worry, we found a significant main effect of Session ($F_{2,228} = 3.5$, $p = .033$, $\eta_p^2 = .030$). Post-hoc tests show that there was a significant increase in Worry between the Training session and the Transfer session ($p < .001$), as well

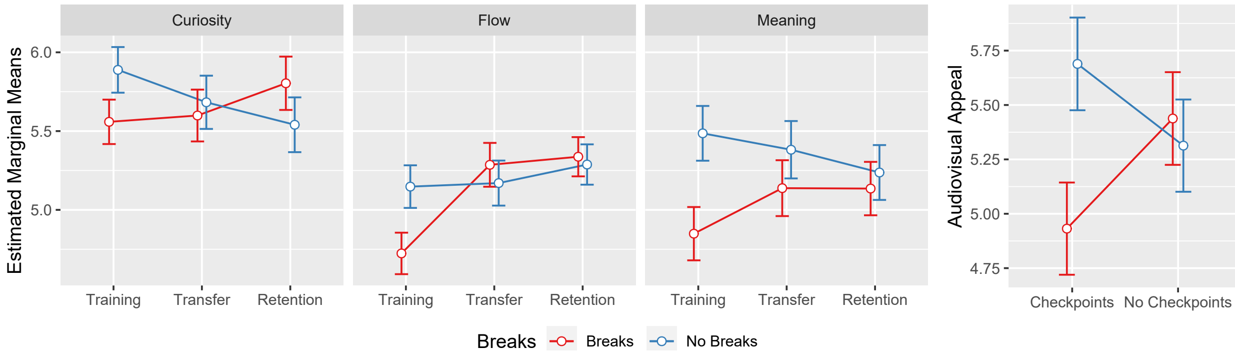


Figure 10.5: Estimated marginal means for curiosity, flow, meaning, and audiovisual appeal questions to help understand the significant interactions.

as between the Training session and Retention session ($p = .039$), but no significant difference between the Transfer and Retention sessions ($p = .125$).

For Curiosity, we found a significant interaction between Session and Breaks ($F_{2,228} = 5.5$, $p = .004$, $\eta_p^2 = .046$). Post-hoc tests do not show any significant differences between any pairs; however, when examining the means we see that when players were made to wait during training, their Curiosity about the game increased as they played the transfer and retention versions of the game that did not have players wait (see Figure 10.5). The opposite occurred for players who trained without Breaks — their Curiosity about the game decreased over the Sessions.

For Meaning, we found a significant interaction between Session and Breaks ($F_{2,228} = 4.9$, $p = .008$, $\eta_p^2 = .041$). Post-hoc tests do not show any significant differences between the groups, so we again look to the means to understand the interaction (see Figure 10.5). Participants who trained with Breaks found increased Meaning in the Transfer and Retention sessions compared to training; however, if participants trained without Breaks, then they perceived slightly less Meaning as they played the game for longer.

For the measures of Mastery, Immersion, Challenge, and Autonomy, we found no significant within-subject effects or interactions for Session, Breaks, or Checkpoints. We also found no significant between-subject effects or interactions for Breaks or Checkpoints.

10.7.2 Open-ended Responses

Game-Integrated Breaks

We asked participants who were made to wait to respawn after they died how they felt about it, what they did during their short breaks, and whether the need to wait changed how they approached the game. In terms of how they felt, the majority of participants (89%) gave negative comments, suggesting that the wait was pointless (i.e., there was no in-game context for it), frustrating, punishing (i.e., not only would they have to repeat part of the level, they had to wait to do so), or that it should be shorter. One noteworthy response stated that although the wait time was objectively short, it felt subjectively long. This could be because, as other participants stated, they were simply eager to get back to playing the game.

Most participants' (77%) focus remained on the game during the break. Either by simply watching the timer tick down (37%) or by considering what they were doing within the game (40%). Most participants (56%) who waited thought that the breaks did not change their approach to playing the game, although others stated that they did. In particular, many suggested that the breaks dissuaded them from taking risks in the game, especially as they got further through the level.

For participants who did not have to wait when they died (their respawn time was fixed to 1 second), we asked them to speculate about whether a break might have helped if they happened to take a break at any point, and if the ability to attempt the game again immediately affected how they approached the game. Only some participants thought that a break would be useful (21%), and over half (56%) thought a break would provide no benefit. Participants thought that a break could hinder their concentration or performance, but might help them deal with fatigue or frustration, particularly if the game had gone on longer. However, about a quarter of these players (24%) did in fact take a short break of some kind. Usually, it was to attend to a distraction, but some took a purposeful break (e.g., to stretch, refocus themselves, or hydrate).

When asked if they thought that being able to attempt the level again right away affected how they played, the majority (79%) said that it did, particularly by allowing them to approach the game more recklessly and by intentionally making many attempts to learn the game.

Participants who trained with game-integrated breaks had an opportunity to play the game without breaks in transfer. When asked which version of the game they preferred, about 43% said they preferred the training version (with breaks), compared to 52% who said that they preferred the transfer version (with no breaks). Of these, 15% stated explicitly that they liked not having to wait as long. When asked which version of the game was more difficult, 67% said that the second version was more difficult, compared to 25% who said that the first version was more difficult (the remainder did not answer the question).

Participants for whom the only change in the game was the new levels (i.e., they trained without checkpoints or breaks) preferred the first version of the game (60.5%), but they were split on which version of the game was more difficult (42.1% said version 1 as more difficult and 44.7% said version 2).

Checkpoints

For participants who trained with checkpoints enabled, we asked them how they felt about checkpoints, whether checkpoints made the game easier, and if they approached the game differently because of the checkpoints. The majority of participants seemed to like the checkpoints. No participant stated that they disliked the checkpoints, and the majority said that they were spaced about right (75%). Most participants (85%) said that checkpoints made the game easier, although two participants pointed out that the game wasn't necessarily easier, just that their progress was saved.

When asked if and how checkpoints altered their approach, most (63%) said that it did. The primary reason being that it allowed them to take risks that they otherwise would have avoided; another was that they felt less pressure knowing that they would not lose all of their progress.

For participants who trained with checkpoints disabled, we asked whether the walk back was easier than the section

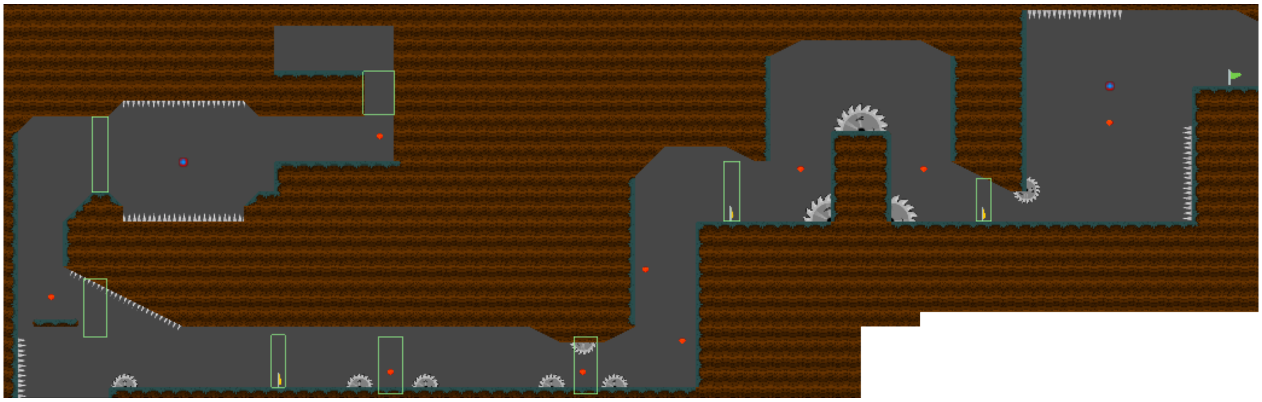
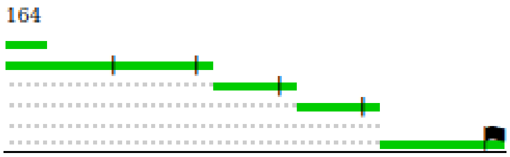


Figure 10.6: Example of one of the game's levels. The spots on the map where progress is logged are represented by the green rectangles. The checkpoints are the yellow flags, and the end of the level is the green flag.

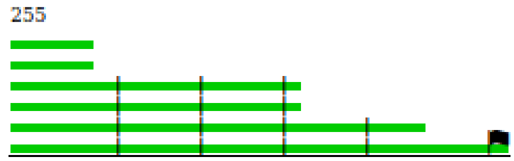
of the game where they died if they focused on anything while they walked back, if they approached the game differently due to needing to walk back, and if they had any general comments regarding this design choice. More than half (64%) said that this process was easier than the section of the game where they died. Most (73%) participants focused on some aspect of the game while they worked their way back to where they died. This included how to move through the level more efficiently, avoiding mistakes they made previously, thinking about how to overcome obstacles later in the level, and reflecting on the controls. On whether the walk back changed their approach to the game, many (75%) said that it did, by approaching the game more cautiously, putting more effort into playing the game, or by causing them to focus on learning aspects of the game or a specific level.

General comments on the need to walk back prompted more negative comments than positive comments (44% vs. 21%). Players said they were frustrated, annoyed, anxious about dying, or that they simply hated it. Some positive comments said that this process motivated them to put in more effort, or that it made the completion of a level feel more rewarding. About half of the comments (52%) discussed the game's design, for example, that the walk back was simply part of the game's challenge, or that this design was okay because it was done in other games. Other responses touch on the absence of checkpoints, with many stating that the game would benefit from checkpoints, although some stated that the levels were short enough to not need them. Related to this, several participants commented on the length of the level being a factor; they stated that the walk back was not a big deal as long as the level was relatively short. A few participants felt that repeating parts of the level allowed them to learn and improve at the game.

Participants who trained with checkpoints were able to play the game without checkpoints during transfer. These participants had a strong preference towards the training version of the game, with checkpoints enabled (73%). Of these participants, a third said they liked the checkpoints of the training version. In terms of difficulty, nearly all participants (87%) said that the second version of the game was harder, and 38% of these participants said that the lack of checkpoints was a reason why the transfer version of the game was more difficult.

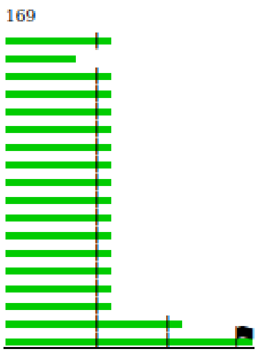


(a) Example with checkpoints enabled.

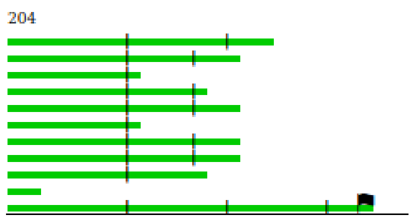


(b) Example with checkpoints disabled.

Figure 10.7: Examples of near-to-ideal scenarios in which players make consistent progress during training and do not require many attempts to overcome a challenge. The first attempt is the top green line and the last attempt is the bottom green line. The examples are from the same level; the number at the top is the participant's ID.



(a) Stalled progress. Lines of similar length indicate that the player is stuck and unable to get further in the level.



(b) Regressed progress. A line that is shorter than the line above it indicates that the player was unable to make as much progress as they had in previous attempts.

Figure 10.8: Examples visualizations of the two scenarios relating to lack of progress during training.

10.7.3 Visualizations

We generated visualizations from the game logs to represent player progress. These show each attempt made by each player as a green line (Figure 10.7); vertical bars indicate checkpoints. When a player dies, the green line restarts from the beginning (if no checkpoints), or from the last checkpoint reached; dashed lines represent the part of the level that was skipped due to the checkpoint. Completed levels are marked with a flag.

The power law of practice [214] suggests that players should continuously improve at the game and continue to make progress. Figure 10.7 shows two examples that approach the "ideal" learning situation: the player makes steady progress in the level without stalling. In contrast to this ideal, we observed two types of inhibited progress (see Figure 10.8):

- **Stalled progress**. The player is unable to progress beyond past attempts; they are "stuck" and unable to proceed, at least for a time.
- **Regressed progress**. The player fails to reach a part of the level that they had reached in a previous attempt.

These observations prompted us to quantify challenges to making progress within the game to better understand how checkpoints and breaks were affecting player progression.

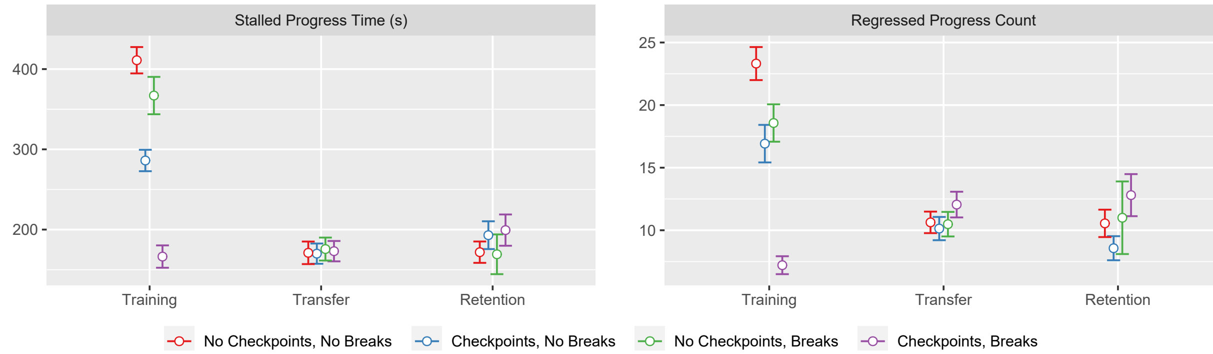


Figure 10.9: Results of our measures of players' progress being inhibited, in terms of stalled progress time and regressed progress count. Error bars are standard error.

Quantifying Stalled and Regressed Progress

Based on these observations, we used the game logs to work out two measures of ways that progress deviated from the ideal: the time for which players' progress was stalled, and the number of times that regressed progress occurred.

For *stalled progress time*, the logs of attempts were processed in the order in which they were collected. For each attempt, the total progress made was compared to past attempts. If the progress made was less than the best past attempt, then that attempt would be considered to be an attempt in which the player was stuck, and that attempt's time would be added to the total time spent stuck.

For *regressed progress count*, we calculated how their current attempt's progress compared to their best attempt. If the player failed to make as much progress, then the attempt was considered to be one where progress regressed. All regressed progress attempts are also stalled progress attempts as they are just a more severe form of being stalled.

The descriptive results of these analyses are presented in Figure 10.9. To check for significant differences between the groups, we used separate two-way ANOVAs for Stalled Progress Time and Regressed Progress Count, with Breaks and Checkpoints as between-subject factors. Any post-hoc tests used Bonferroni corrections.

In Training, we found significant main effects of Breaks ($F_{1,146} = 90.1$, $p < .001$) and Checkpoints ($F_{1,146} = 22.8$, $p < .001$) on Stalled Progress Time, with both decreasing Stalled Progress Time. There was also a significant interaction between Checkpoints and Breaks ($F_{1,146} = 4.9$, $p < .029$) — post-hoc tests show that the combination of Breaks and Checkpoints greatly reduce stuck time than either on their own. We also found significant main effects of Breaks ($F_{1,146} = 47.2$, $p < .001$) and Checkpoints ($F_{1,146} = 31.3$, $p < .001$) on Regressed Progress Count, with both reducing Regressed Progress Count. The interaction was not significant ($F_{1,146} = 31.3$, $p = .057$).

In the Transfer test, there were no significant main effects, nor any significant interaction effects for Stalled Progress Time or Regressed Progress Count (all $p \geq .280$). Similarly, in the retention test, there were no significant main effects, nor any significant interaction effects for Stalled Progress Time or Regressed Progress Count (all $p \geq .148$).

Therefore, Checkpoints and Breaks both helped players make progress while they were present, and this different training did not affect their progress in the later test sessions. There was also an interaction between Checkpoints and Breaks for Stalled Progress Time — Breaks were much more effective when Checkpoints were also present.

10.8 Discussion

The study provided several new findings about the effects of breaks and checkpoints on progress, skill development, and player experience:

- Players completed significantly more levels with both breaks and checkpoints (three additional levels in both cases compared to the baseline) during the training session (when the techniques were present), and the combination of breaks and checkpoints showed the largest increase in progress (eight levels better, although the interaction was not statistically significant);
- There were no significant differences in levels completed on either the transfer or retention tasks, suggesting that neither technique reduced (or improved) skill development beyond the training session;
- Both checkpoints and breaks showed improvements in our secondary measures of progress — breaks reduced player death rate (likely due to the forced pause after dying), and the combination of the two techniques significantly reduced the number of stalls and regressions;
- For measures of player experience, no differences were found between training and transfer sessions for checkpoints; for breaks, we found improvements in three measures (flow, curiosity, and meaning) when participants moved to the tasks where breaks were removed;
- No participants stated that they disliked checkpoints, and no participants stated that they liked the breaks.

10.8.1 Explanation of Results

Why Did Game-Integrated Breaks Help Progress?

Past work that has applied theories of spaced practice to games has done so on a fixed schedule, for example, a two-minute break after five minutes of play [151, 231], whereas this work presented breaks to players dynamically, after each death. We found benefits to presenting breaks in this way to players that were similar to the benefits found within other work that also applied spaced practice to games [151, 231]. Participants completed more levels, stalled fewer times, and spent less time not making progress when breaks were present. This strongly suggests that spaced practice can also be effective with a dynamic schedule and with dynamic break lengths. Past work not looking at games or even perceptual-motor skill have also found that shorter breaks can be effective [303, 132] and also that a variety of schedules for the breaks can be effective [257]. We find that the same is true for games.

More generally, spaced practice aids performance in a variety of ways. It aids skill development by helping learners develop memories that allow them to carry out a task [303]. This might be possible due to encoding variability, where events spaced by time can be encoded into memory in different ways [28, 303]. Spacing may also assist in the *consolidation* process, in which memories become more stable and resistant to decay [60, 282], as well as force a retrieval of the relevant memory traces when returning to the task, which reinforces them [321, 244]. Short break intervals in particular might be effective because of the concept of deficient processing, in which less attention is given to the second or subsequent attempts [132]. A short time delay between attempts could prompt a learner to direct more of their voluntary attention toward a task [132]. Similar to this is the thought that a short break could give one time to recover from physical [257] or cognitive [321, 3] fatigue.

Spaced practice helps learners transition from early stages of learning to later stages of learning [292], in which learners can execute a skill with more fluency and better attend to the relevant stimuli [302]. It also aids the process of *knowledge compilation* [10], in which declarative knowledge (verbal information about a skill) becomes encoded as procedural knowledge, which can be more directly applied to executing a skill.

Finally, we must acknowledge that game-integrated breaks did come at a cost — players overall spent less time actively playing the game when these types of breaks were used. When breaks were included, players spent an average of 274 seconds waiting (min=124, max=451, SD=68.4), compared to 61.2 seconds (min=29, max=124, SD=18.8). Therefore, unlike in past experiments, the total training time was not fixed; players who were given more breaks simply had less time to complete the levels. We corrected for this by including time spent training as a covariate in our analyses, and this variable was significantly related to the levels completed in training and death count in training, indicating that this reduction in training time did have a cost.

Why Did Checkpoints Help Progress?

We observed that checkpoints helped players make progress when they were present and did not get in the way of learning the game. Increased progress was made in part due to players making more consistent progress, stalling fewer times and for less time.

In terms of theory, in Section 10.3.2, we proposed that checkpoints could affect the variability of practice, as well as facilitate the process of part-task practice. Considering part-task practice, checkpoints implicitly allow this to occur by allowing a player to focus only on overcoming a single challenging obstacle or section of the game. Players can try to correct errors they are making by attempting different strategies. In the written feedback, we found that many participants took advantage of this learning opportunity. In cases where this worked, players overcame obstacles, which means that they progress in the game and can then be also exposed to a greater variety of practice.

Practice would have more variety if players were able to make progress in the game, and they did. In non-game contexts, variability in practice generally comes with increased errors in early learning [185], which we did observe somewhat, as players died more times (though only when game-integrated breaks were not present). However, due to the task being a game in which players make progress, the increased errors did not mean that performance was worse, as players were clearly able to make more progress when checkpoints were present.

We compared checkpoints to the absence of checkpoints, and so part of the reason why checkpoints were beneficial is that playing without checkpoints — making players work their way back through the level — was likely not helpful. We found no evidence that working through the potentially easier, early parts of the level was beneficial training.

We did not find any long-term benefits to progress due to checkpoints that would have indicated greater skill development. This does not mean there were no learning benefits at all, only that there were no additional benefits over the other variations tested. One important consideration, however, is that unlike traditional perceptual-motor skills, immediate performance can be more important than long-term performance in games, as a player's primary goal is often to make progress [167].

10.8.2 Do Players Hate Breaks?

None of our participants stated that they liked taking a break, and an overwhelming majority were critical of the wait. It was viewed as too long, punishing, and pointless. This makes sense considering that games are played for leisure, often as a change or a distraction from other aspects of a person's life. So why take a break? We show that there are good reasons to do so. In addition to any benefits of reducing fatigue or repetitive strain issues [218], we find clear and immediate performance benefits of taking a break, even one as short as ten seconds. And yet, the suggestion that players take a break to aid their progress might be met with even more opposition. When a player is engaged with a game, intently focusing on overcoming a challenge, all a player might want to do is keep trying. Despite our participants disliking the breaks, we did have participants use the breaks productively, as an opportunity to refocus or collect themselves. A player's dislike of breaks does not mean that they are ineffective, and without a break, they may have never been prompted to reflect on what they were doing. In fact, the players who want to get back to playing the game might be more inclined to use their break to consider what they are doing in the game.

Disliking the breaks within a game does not necessarily translate to disliking the game as a whole. In our game, the presence of game-integrated breaks did not significantly affect how they experienced the game in terms of ease of control, progress feedback, goals and rules, mastery, immersion, challenge, and autonomy. A similar lack of change in subjective experience was also found in past work [151, 231]. We only found that flow and meaning were slightly reduced when breaks were present. Flow might have been reduced due to the interruptions that our breaks created, and a reduction in meaning means that participants thought the task was less important. Even with this, when players had the opportunity to play the game again without breaks, there wasn't a strong preference for this version of the game, and players tended to think that the version of the game with breaks was easier, even though it really was not; they were just performing better.

If breaks do bother the player, this raises the question of whether it is more important that a player gets better at the game or enjoys it. This question is difficult to answer, as competence within the game is very much linked to a player's motivation to keep playing [247, 237]. However, not every game has an emphasis on performing procedural skills well. It may not make much sense, for example, to include breaks in games such as *The Sims* or *Animal Crossing*, which are simulation games that place fewer demands on perceptual-motor skill development. In a game like *Super Meat Boy*, in which the ability to execute a series of precisely timed inputs is needed to succeed, the benefits of breaks may greatly increase the overall enjoyment of the game.

Improving Acceptance of the Breaks

A criticism of breaks mentioned by some participants was that there was no context for the break. For example, one participant pointed out that if the game had a visible time limit, then the wait would make more sense (possibly because it would more explicitly be a form of punishing the player for failure). Other participants pointed out that a break would make more sense for different types of games such as multiplayer games, where players already accept breaks in the form of respawn timers or waiting for the matchmaking system to pair them with other players. If there seems to be a reason for the breaks, players may be more accepting of them.

Players may also be more accepting of the breaks in our game if their presentation to players was improved. When players died, we simply showed a timer. Death in other games involves animations, kill cams (replays where you can see how you die), spectating other players, or amusing cut-scenes such as how characters taunt Batman when he dies in the *Arkham* series of games [26]. Additionally, our breaks were predictable. Players learned that when they died they would need to take a break. It might be better to wait for indications that the player is struggling before trying to provide aid, such as after a certain length of time with stalled progress or after a certain number of progress regressions. Finally, it might be possible for players to continue to engage with the game in some way while they wait. For example, past work explored what players could do during breaks and found that simply switching to a separate but related task is beneficial [231].

10.8.3 Implications for Players and Game Designers

We learned that many players' intuition regarding breaks and checkpoints is wrong. Breaks, something that many players are opposed to, can be beneficial and enhance player performance. Checkpoints, unlike what many believe, do not necessarily make the game easier or result in players becoming reliant on them. In our game, checkpoints did not affect skill development and were beneficial because they helped players make progress. Game designers should consider whether there is value in including breaks or checkpoints in their games.

For checkpoints, we modelled our checkpoints on those already found within many commercial games. However, current biases about checkpoints and the notion that they may hinder learning could mean that there is room for improvement. Designers should carefully consider the frequency of checkpoints. We found no benefits to replaying earlier parts of the game compared to making use of checkpoints, and checkpoints can help players more rapidly apply a trial-and-error approach to overcome challenges. Checkpoints should not be considered to be a crutch, but instead a valuable tool.

In contrast to checkpoints, breaks are not typically included in games with the intention of helping players make progress and improve at the game. We extend past work by demonstrating that breaks can be presented with a dynamic schedule and the breaks can even be quite short — as short as ten seconds — while still improving performance. We gave players a break after each time they died, an implementation already found in some games (such as in many online first-person shooters) and one that is relatively easy to implement. While players disliked this, we note that it did not meaningfully affect their enjoyment of the game as a whole. Additionally, it should be possible to design breaks that are less apparent but still serve to improve performance. Finally, breaks prompt players to approach the game with more caution, taking fewer risks and dying fewer times. If players are repeatedly throwing themselves at a challenge without taking time to consider what they are doing or why they are failing, there is a good chance that a break will prompt them to do so.

Combined, checkpoints and breaks together provide game designers with a lot of control over how players make progress in a game. Being able to predict and dictate a player's progress can allow game designers to craft enjoyable game experiences by ensuring players are consistently and reliably progressing at the appropriate pace. For example, in the scenario presented in the introduction, what can be done if the player is repeatedly trying to overcome a challenge in a game but is unable to do so? If they are playing with checkpoints enabled, but no break, then it is likely they are attempting various strategies and observing the results. But they might also be simply throwing themselves at the problem without actively considering what they are doing. In this case, a short break may be just what the player needs to allow them to reconsider what they are doing and redirect their attention towards making the most of their attempts, rather than trying the same approach repeatedly.

10.8.4 Limitations and Future Work

The main limitation of this study is that we only tested one game, and one implementation of breaks and checkpoints. Furthermore, our game was specifically built for the study — it was not a commercial game. Further work is needed to determine if our results generalize to other side-scrolling platform games, and to other game genres. Future work could reproduce the results using other games, as well as testing different variations of our implementation of checkpoints and breaks. For breaks, given the variety of prior games in which spaced practice has been tested, we are confident that the benefits found with our specific implication of spaced practice will also be present in other games and game genres. In terms of checkpoints, this concept is also found in genres other than platformers and the way in which it is implemented is similar across genres. There is nothing specific about checkpoints found in side-scrolling platformers, compared to, for example, single-player first-person shooters; however, we must acknowledge that differences are a possibility. More generally, our platformer may not have been as polished as a commercial game. A professionally made and play-tested game designed to be played for many hours may result in a different experience.

Future work could investigate how breaks could be designed and implemented into commercial games. In particular, additional studies could explore how to improve the presentation of breaks and whether different frequencies or lengths would also be effective. The design of our breaks made them very obvious and uninteresting. This gives several opportunities for future work. Our breaks were highly predictable — players knew that any time they died they would need to wait. Future work could explore variations on dynamically presenting the breaks. Considering that spaced practice can enhance performance, our motivation was to provide aid to a player at a time when it might be effective, but it may be possible to time the breaks differently or present them only under specific circumstances. For example, by considering indicators of inhibited progress. Another opportunity to improve our breaks is to make them more interesting. Future work should investigate whether presenting the break differently can make the break less apparent or increase acceptance of the break. For example, instead of a simple timer, animations or short activities (like those in [231]) could be included to entertain the player and potentially improve the experience of breaks or waiting.

Testing different frequencies or lengths of breaks is also something future work should explore. Our breaks were only up to ten seconds long, but breaks as long as two minutes [231] or even one day [151] have also been tested. Based on past work, breaks of a variety of different lengths and frequencies could be useful. For short breaks, it is thought that a short timer interval between rehearsals prompts a learner to direct more of their voluntary attention towards the task [132] while also aiding with processing [303]. Longer breaks are thought to be beneficial due to strengthening the memory of how to execute the relevant skills [201, 238] or due to the second event being encoded differently in memory [28]. However, there may be upper or lower limits to explore. At some point, the breaks may be too long or too short, or presented too often or too infrequently to be of any benefit. It may also be that the most effective break length and frequency changes as players improve [49].

Additionally, future work could test whether breaks could be turned into short training scenarios. Prior work found that simply switching to a new task for an implicit break had benefits similar to a more explicit break [231] and so it may be possible to do this without losing the benefits of spaced practice. Further, it may be that some of the breaks found in commercial games are already doing this. Consider, for example, spectating your teammates in an online first-person shooter game while you wait to respawn. Past work has found that there are performance benefits to watching demonstrations of a game [232]. Simply spectating other players may also act as a demonstration that could benefit a player's performance.

There are opportunities for future work to test varying implementations of checkpoints. Checkpoints — or more generally, saving progress — can be presented in many different ways. Ours were presented as visible goal markers for players to reach, but they may also be invisible (i.e., auto-saves [309]). We placed checkpoints after what we felt were difficult obstacles in the level, but we could have spaced them more or less frequently than we did. It is possible that a system that prevents even the smallest loss of progress will affect performance and learning differently than our implementation.

We logged the number of deaths as a way of understanding how checkpoints and breaks were affecting player behaviour in the game. The death count was affected by breaks and checkpoints, but we do not know precisely why. With breaks, players may have avoided dying within the game to avoid the delay (some participants brought this up in their written responses). Those who played with checkpoints stated that the checkpoints allowed them to take additional risks (although we found no main effect of Checkpoints on death count). However, if breaks were given then the death count was low regardless of whether checkpoints were active. There is an interaction between the two that is worth further investigation — are players making an intentional decision to take more risks when checkpoints are present, but only if breaks are not also present? Furthermore, past work has found a relationship between death and a player's perception of challenge [157] — players view a game as more challenging if they die more frequently [64], yet we found no differences in challenge even though we found differences in death counts due to checkpoints and breaks. In our study, death was not necessarily an indicator of a lack of progress within the game in terms of levels completed — players might have died more often yet made more progress, and players may have even intentionally chosen to die in order to attempt different strategies. The relationship between deaths and perception of challenge is likely more complex than past work suggests, and future work should examine the relationship further.

A final limitation of our study is that our players were paid to play this game. This means that players might have been willing to put up with frustrations more than they typically would in a game. This also could have influenced them to rate their experience more positively than they may have otherwise, considering how interesting a game is when compared to other tasks on Amazon's Mechanical Turk.

10.9 Conclusion

Ensuring that players continue to make progress in a game is of great importance to both game designers and players. We carried out a study to test two techniques of supporting this, breaks and checkpoints. We found that both were effective at supporting players' progress without hindering skill development. Even with relatively short breaks (no longer than ten seconds) that were fully integrated into the game, players were able better able to overcome challenges and make progress, although participants tended to dislike the breaks. Checkpoints were also effective when integrated into the game before difficult obstacles, and were liked by our participants. Our work provides several contributions that can change the way players and game designers think about practice within games:

- We show that spaced practice can be integrated into games using a dynamic schedule rather than a fixed schedule and that the breaks can be as short as ten seconds while still effectively improving immediate performance and aiding progress.
- We show that checkpoints are an effective method of improving immediate performance and helping players make progress and that they do not come with any apparent drawbacks in terms of skill development.
- We show that benefits occur despite player beliefs about checkpoints and breaks, and the subjective experience of the game is largely unaffected.

Our results provide useful information for players who want to improve their skills, practical suggestions for designers who are interested in ways of helping their players make progress through their games and add to our overall understanding of how skill development occurs within games.

	Training						Transfer						Retention					
	Levels Completed			Death Count			Levels Completed			Death Count			Levels Completed			Death Count		
Covariate	F	p	η_p^2	F	p	η_p^2	F	p	η_p^2	F	p	η_p^2	F	p	η_p^2	F	p	η_p^2
Time spent training	**10.6**	**.001**	**.071**	**37.5**	**.001**	**.209**	4.5	.092	.031	**12.9**	**.001**	**.083**	2.3	.135	.020	**4.9**	**.029**	**.042**
Tutorial levels time	**69.7**	**.001**	**.334**	1.3	.254	.009		n/a		0.01	.972	.001	**196**	**.001**	**.231**	**17.9**	**.001**	**.138**
Platforming familiarity	**6.5**	**.012**	**.045**		n/a		**8.0**	**.005**	**.054**		n/a		**13.9**	**.001**	**.113**		n/a	
Attentional control	0.15	.697	.001	2.3	.133	.016	0.3	.590	.002	0.01	.939	.001	0.5	.495	.004	1.0	.329	.009
Probability of success	0.03	.952	.001		n/a			n/a			n/a		0.01	.926	.001		n/a	
Perceived task challenge	0.2	.865	.001		n/a			n/a			n/a			n/a			n/a	
Task-related anxiety	**4.9**	**.030**	**.033**	**5.9**	**.017**	**.040**	**4.2**	**.043**	**.029**		n/a		0.5	.485	.004		n/a	

Table 10.2: The effects of the covariates, from the separate ANCOVAs. A "n/a" indicates that the individual difference measure was not used as a covariate for that outcome measure.

Part III

Discussion

11 General Discussion

In the introduction of this dissertation, I stated that my overall goal was to address a current problem that exists in game design: there are times when players want to get better at the games they play yet game designers do not always know how to explicitly support skill development in their games. A single dissertation is insufficient for solving this problem entirely, but I did take several steps toward solutions for this. In particular, this dissertation had three sub-goals. First, I outlined the current approaches for supporting skill learning in games, which I did through the background chapter (Chapter 2).

My second sub-goal was to understand why existing approaches to skill learning used within games work. This was done by examining those games and the assistance within them through the lens provided by past research on perceptual-motor skill development. This was done in Manuscript A by testing out different types of navigation assistance that are currently applied within many games and investigating how they affect performance and learning. This was also done in Manuscript D when I looked at checkpoints and the ways in which checkpoints might affect what players are during practice (including how it might introduce part-task practice or affect the variety of practice), as well as the effects of checkpoints on performance and learning.

My third sub-goal was to explore ways of supporting skill learning within games that are novel in the games domain, but not novel in the perceptual-motor skill learning domain. This was done in Manuscripts B, C, and D. Manuscript B explored a novel method of navigation assistance based on the idea of concurrent mechanical guidance. Manuscript C applied different amounts of spaced practice to a game to evaluate the performance and learning effects, showing that spaced practice also applies to games. Manuscript D extended the work on spaced practice by integrating it into the experience of the game, presenting it as more frequent but shorter breaks, an approach to spaced practice that differs from what is done within experiments (but with some similarities to [201]), but could be more applicable to what is found in the wild.

11.1 Summaries of the Manuscripts

This section provides a brief summary of the manuscripts, with an emphasis on summarizing the hypotheses and results.

11.1.1 Manuscript A

One difficulty that many new players face is that they cannot find their way inside the game environment and become lost. From the literature on skill development, one common approach to assisting one to learn new skills is by providing guidance. The literature on this suggests that for simple skills [350] where trial-and-error is an effective approach, guidance improves performance but negatively affects learning when evaluated through retention and transfer tests [257, 16, 274, 135, 182, 343]. This is because the learner becomes reliant on the aid. Therefore, for Manuscript A, I explored the role of navigation guidance on performance and learning by testing navigation guidance systems commonly found in games.

I had two hypotheses:

1. Navigation guidance will improve performance when present.
2. Navigation guidance will negatively affect learning due to reliance on the guidance (i.e., Incidental learning will not occur).

To test these hypotheses, I had participants train in two different 3D game environments using either map guidance (called "no guidance" in the paper), position (or "moderate") guidance, or trail (or "strong") guidance. I then tested participants' spatial knowledge by having them navigate the environment a second time, where they navigated to specific landmarks without any help in the form of guidance (the transfer task). This was done over two studies, where the only difference between studies was the length of time spent training — 48 routes over 3 days for Study 2 compared to 16 routes over 1 day for Study 1. For both studies we found the same results:

- **Guidance helped considerably with immediate performance:** The trails guidance group performed better than any other group, and the position guidance group performed better than the map guidance group.
- **Wayfinding ability in testing was unaffected by guidance:** All groups performed the same in the transfer task.
- **Spatial knowledge ability was also unaffected by guidance:** All groups performed the same when answering the spatial knowledge questions.

These results were unexpected for two reasons. First, the map guidance group spent considerably more time training within the environments — approximately twice as much time as the trail guidance group. Second, the map guidance group needed to put in more effort to learning the environments; the retrieval effort hypothesis [238] as well as the findings on the effects of guidance on learning (e.g., [16, 234, 274]) suggests that this group should have learned their way around the environment better than the other groups.

11.1.2 Manuscript B

Manuscript A's results were unexpected in that they did not align with what past research had found when providing learners with augmented feedback and guidance, where the learners tended to become reliant on the aid (e.g., [16, 258]). Instead, all groups performed similarly to one another when they were tasked with navigating through the environment again without guidance. This raises the question of whether more severe forms of navigation guidance would result in similar performance and learning characteristics to the "strong" guidance used. The literature on applying guidance for skill development explores the use of "mechanical" guidance, where the learner is physically constrained in some way from making errors. Therefore, Manuscript B explored whether more "extreme" navigation guidance (where the player holds down a button to be moved in the correct direction) results in similar performance and learning to the "strong" visual guidance of Manuscript A (glowing trails).

Manuscript B makes new use of the the data from Manuscript A, but only from one of the one environments. This is because that same environment was used again in an new experiment, where the learning and performance effects of "mechanical" navigation guidance. In comparison to Manuscript A, we found similar results:

- **Guidance helped with immediate performance:** Position guidance or trail guidance resulted in significantly better performance when they were present, but trail guidance and position guidance performed similarly to one another.
- **Wayfinding ability was only slightly affected:** The unassisted group performed no better at the task than any other group, but the position guidance group performed better than the trails guidance group and the best overall.

The results differ from Manuscript A in that position guidance resulted in improved learning over the other groups, however, the trail and map guidance groups still showed similar learning. Therefore, in Study 2, we looked at a more extreme form of guidance where players simply held down a button to be taken along the correct path ("rails" guidance) in comparison to no guidance (not even a map) and the trail guidance used in Study 1. I had the following hypotheses:

1. Navigation guidance will improve performance when present.
2. Navigation guidance will negatively affect performance in the transfer test, where no guidance is given, and new routes must be formed.
3. Navigation guidance will result in no differences in performance on the retention test, where no guidance is given, but the routes were practiced in training.
4. Mechanical guidance will result in worse performance in testing than visual guidance.

The results from Study 2 showed that:

- **Guidance helped with immediate performance:** Both trail guidance and rail guidance groups performed better during training than the no guidance group. The trails and rails groups performed similarly to one another.
- **The ability to navigate new routes without guidance was affected:** We found that the unassisted group was better able to navigate the environment when asked to locate previously seen landmarks from new starting locations. The trail and rail groups performed similarly to one another.
- **All groups could navigate the trained routes once guidance was removed:** Every group was able to navigate the routes in a similar length of time once the guidance was removed. However, unassisted participants did make fewer errors.

The results from this study do line up more with prior results on the effects of guidance on learning perceptual-motor skills (e.g., [16, 234, 274]) in that there was a drawback to training with guidance. However, that drawback was not as pronounced as one might expect, especially considering that the unassisted group spent over twice as much time training within the environment compared to the guided groups (24.9 minutes compared to 10.4 minutes for trails and 9.3 minutes for rails).

11.1.3 Manuscript C

For Manuscripts C and D, the aim was to focus on ways to help players make continued progress at improving in the game. The literature on perceptual-motor skill development discusses several theories to help learners improve skills more quickly. One theory is *spaced practice*, the idea that breaking up training with breaks can improve performance and learning. Therefore, Manuscript C explored whether the benefits of spaced practice that are found in a variety of other perceptual-motor tasks are also found within digital games. This work had two hypotheses:

1. Spacing practice will result in improved performance.
2. Spacing practice will have lasting benefits to learning.

Secondly, Manuscript C explored whether the length of the rest would affect performance and learning. Therefore, a variety of rest intervals were tested (three seconds, two minutes, five minutes, ten minutes and one day) to explore whether one particular interval would result in better performance or learning over the others. Participants trained over four sessions and then returned for a retention test after one day. We had participants play a clone of the commercial game, *Super Hexagon* [45]. Performance was measured in terms of the average length of time alive. The results of this work are:

- **There were differences in performance due to the rest interval:** After training, the three-second group (i.e., continuous practice) performed worse than all other groups, except for the one-day group.
- **The gains to performance were temporary:** After a one-day break, the one-day and three-second rest interval groups effectively “caught up" to the others — all the groups reached the same level of performance.
- **There was no single “optimal” rest interval:** The ways in which the groups arrived at their final performance differed slightly, but no one rest period stood out as being clearly better than other others. However, our results do show that there was no benefit to taking breaks longer than two minutes.

These results are generally in line with what has been found for other perceptual-motor tasks that are not digital games (e.g., as summarized by [81]), indicating that spaced practice can be applied within the domain of digital games.

11.1.4 Manuscript D

The results of Manuscript C suggest that spaced practice does in fact have beneficial effects on performance in games. Manuscript C also found that spaced practice is generally better than continuous practice — one inter-session rest interval was not obviously better than any other. What Manuscript C did not do, however, was suggest or determine how to actually implement spaced practice within a game. Very rarely do games operate on a fixed schedule with play sessions of a fixed amount of time, so when can a break be introduced? How can breaks actually be applied to help players continue to make progress within the game?

Manuscript D extends the work on spaced practice by looking at a more complex game (a side-scrolling platform game) and dynamic breaks rather than static breaks. These breaks were more frequent, but shorter, and occurred as a consequence of events within the game rather. These "game-integrated breaks" occurred upon a player's death within the game, and could be up to ten seconds long, determined by how long a player was alive before their death.

In addition to spaced practice, Manuscript D also investigates how checkpoint systems affect performance and learning in games. Like spaced practice, checkpoint systems modify practice. In terms of theory, they modify practice in two ways. First, assuming that checkpoints do in fact help players make progress, then they change the *variability of practice* by allowing players to practice more variations of the task. This increased variability is thought to promote motor learning by helping a learner cope with novel situations. Second, if checkpoints are utilized before challenging sections of a level, then they may be modifying practice by providing *part-task practice*. That is, players are able to focus solely on a specific subset of skills being tested rather than the game or level as a whole. The conventional wisdom by players on discussion forums who oppose extreme use of checkpoint systems is that players are not engaging with the game properly and so are compromising their learning of the game. In the game, checkpoints were implemented as flags throughout the level. After players touch a flag, their progress is saved so that they are taken back to this checkpoint if they die rather than starting the level over from the beginning.

These two factors, spacing and checkpoints, were crossed in a 2x2 design to create four groups:

- Unmodified practice (no checkpoints, no spacing).
- Breaks (no checkpoints, but spacing is present).
- Checkpoints (checkpoints present, but spacing is not).
- Breaks & Checkpoints (checkpoints and spacing are both present).

We had players train with the game for twenty minutes under one of these four conditions and then played a transfer session, where the game contained brand-new levels and no spacing or checkpoints. Participants were invited back seven days later to complete a retention session, in which they played the game for an additional ten minutes, playing through the levels they trained on but this time without spacing or checkpoints. Performance was assessed by measuring the number of levels completed in each session. My hypotheses were:

1. Checkpoints will improve performance when they are present.
2. Checkpoints will negatively affect performance when they are taken away.
3. Spaced practice will improve performance when it is present.
4. Those performance gains will persist into testing tasks, where spaced practice is not present.

The results from the study show that:

- **Game-Integrated Breaks and Checkpoints helped performance while they were present:** We found that players were able to complete more levels during training when breaks or checkpoints were present. Furthermore, players experienced less stalled or regressed progress with either breaks or checkpoints, and even less if both breaks and checkpoints were present.

- **The learning of the game was unaffected:** There were no differences in performance between the groups in transfer and retention tests in terms of level completion. Additionally, there was no difference between the groups in terms of experiencing stalled or regressed progress.

- **Participants said they did not like breaks:** Our participants were eager to get back to playing the game and did not like the interruption, saying it was too long or pointless. However, this dislike of the breaks did not affect quantitative measures of the subjective experience of the game in a significant way.

- **Participants were accepting of checkpoints:** Our participants were generally okay with the checkpoints in the game, saying that there was an appropriate amount. However, even without checkpoints players were generally okay with needing to work their way back through the level. There were no differences due to checkpoints in terms of our quantitative measures of subjective experience.

11.2 How Support Affected the Learning of In-Game Skills

The manuscripts each present their own explanation of their results, but they do not explore in much detail how the support they provided interacts specifically how a learner is learning the trained skill. This section does this, and in particular it refers back to the concepts presented in the background chapter, Chapter 2.

11.2.1 How Guidance Affected Navigation Learning

In Manuscripts A and B, the studies involved novices who were in Fitts and Posner's [104] first stage of skill development. In this stage, the participants are trying to figure out how to carry out the task [302], in particular, the task of navigating through the environment. As described in Chapter 2, learning and improving at skills comes from a learner being able to leverage feedback to evaluate how well they did [302]. A learner will try out various strategies and use feedback to determine which strategy lead to the desired result [288]; in this case, to get to the desired location within an environment.

However, the navigation tasks used in Manuscripts A and B involved environments where participants would need to wander around for quite a while before they happened upon their desired destination. Feedback that provides the participant with information about the success of their attempt at navigation is therefore provided infrequently. Guidance is thought to generally have a strong and positive effect on performance, but not learning [253], possibly because it causes participants to pay attention to the guidance rather than to the feedback intrinsic to the skill that must be attended to if the guidance were ever removed [253, 249]. But in scenarios like the navigation task used in the manuscripts, where inherent feedback is less frequent, it appears to not negatively affect learning while enhancing performance when it is present. In other words, it provides the learner with knowledge of results when there normally would be little to no knowledge of results.

11.2.2 How Spaced Practice Affected Skill Learning in Manuscripts C and D

In Manuscript C, participants played a clone of *Super Hexagon* (Figure 8.2), a game that was intended to be easy to learn but difficult to play well. In Manuscript D, participants played a different game — a side-scrolling platform game (Figure 10.1).

Before describing how spaced practice affected the learning process of these games, I will first describe the learning process looks like in general for these games. After, I describe how spaced practice affected skill learning in Manuscripts C and D.

The Process of Learning *Super Hexagon*

Super Hexagon It involves only two possible inputs (move clockwise or counter-clockwise) and has a simple goal (avoid the obstacles). The intention was to limit how long participants would be in the first stage of skill development where they need to work out how to accomplish the goal [104]. With a game that was straightforward to learn, participants could transition more quickly to a stage of skill development where they knew what to do but had difficulty executing the correct response.

This is a scenario where some support methods would not be helpful. Additional instruction, for example, is not helpful if the learner already knows how to carry out the correct response. In this case, the issue is that the learner is not identifying and processing the stimulus fast enough to select the correct response (see the information processing model presented in Section 2.2.2). Furthermore, in this scenario, there is little to be improved about the task itself to improve how easy it is for a player to get better at it. This is because it already fulfills the conditions for deliberate practice [95] — it has well-defined, specific goals (avoid obstacles), clear feedback (touch an obstacle and you fail), and is at a level of difficulty just outside of a player's comfort zone (because the difficulty of the game increases slowly over time and resets with each new attempt so players will likely only fail at the game once it gets to just beyond their current skill level). This last point is crucial because a challenge well-matched to a player's ability makes an ideal learning environment [332].

Much of the difficulty of the game comes from its lack of stimulus-response compatibility [236]. For example, instead of the left arrow key simply moving your avatar left, it actually rotates your avatar counter-clockwise along a circular arc. So if your avatar is on the top of the arc, the mapping is natural and intuitive, but once your avatar is at the bottom of the arc, the mapping is reversed. Pressing the left arrow key moves the character right on the screen. This makes it less intuitive to select the appropriate key as a response. Combine this with a screen that is constantly rotating, which makes it more difficult to parse the on-screen stimuli. This is what makes up the challenge of the game.

This provides minimal options for aiding the player. A system might have been built that parses the stimuli for the player and tells them what response to select (i.e., a guidance system), but unlike navigation where there are usually other activities within a game that go alongside navigation, such a system would effectively play the game for the player. Therefore, because there is little that one could change about the game itself to support players, *Super Hexagon* is a good candidate for applying spaced practice as a performance aid.

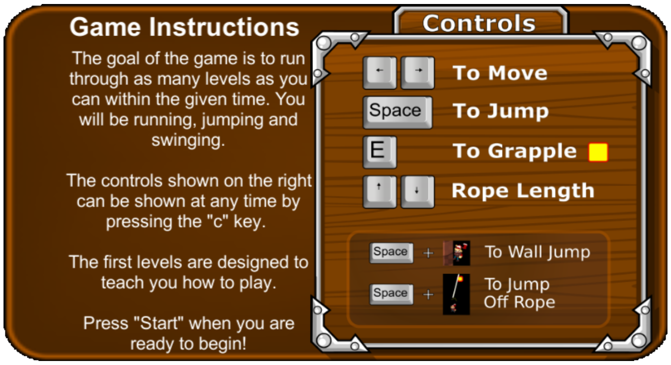


Figure 11.1: The screen at the start of the game used in Manuscript D, where the objective of the game and its controls are described.

Figure 11.2: The second tutorial level of the game used in Manuscript D.

The Process of Learning Our Side-Scrolling Platform Game

In Manuscript D, participants played a different game — a side-scrolling platform game. This game is differs from *Super Hexagon* because there were a collection of different skills (moving, jumping, avoiding obstacles, wall-jumping, and grappling) that needed to be applied to make progress. Therefore, we had to instruct players on how to play the game. At the start, players were given a description of the goal of the game as well as the controls (see Figure 11.1). After, players completed an in-game tutorial during the first three levels of the game. This tutorial (see Figure 11.2) first had players learn how to move and jump, then learn how to wall-jump, then learn about the grapple mechanic and how to swing from grapple points in the environment. In this game the challenge that players face comes from learning how the on-screen avatar responds to inputs and learning how to get past various obstacles within the environment.

It was expected that any participants who could make it through this tutorial would understand what they were doing as well have some sense of how to do it. They would therefore be in the second stage of learning, where their objective would be to make adjustments to how they are executing the skills they've learned [104]. We also recruited players who self-reported themselves as being familiar with side-scrolling platform games.

Spaced practice was provided as a possible support method. However, instead of providing the breaks on a schedule they were instead provided in response to player actions. When a player died, a break was given to them. Therefore, the length of each break was shorter, but they were provided more often.

How Did Spaced Practice Affect Skill Learning?

Manuscript C proposes that spaced practice might not benefit the players of games because many gamers are eager to make repeated attempts at overcoming a challenge within a game. However, there were benefits to spaced practice. Manuscript D further tests the assumption that players dislike breaks by asking them directly how they feel about them, and confirms that players are not in favour of taking breaks. Although it also showed that there were performance benefits to taking breaks. Despite the way that the learner feels about taking a break, it is beneficial. The literature on spaced practice strongly suggests that the benefits of spaced practice are ubiquitous, showing up in a diverse collection of different skills [303], and digital games are no different.

There are many different reasons why spaced practice could affect performance (recall that learning is inferred by measuring performance, see Section 2.3). It could simply be that providing players with breaks just long enough to give them a chance to recover from fatigue (whether that fatigue is physical [253] or cognitive [321, 3]) provides temporary performance improvements; however, research suggests that there are other factors also leading to performance improvements [253].

There are various theories that attempt to explain why spaced practice works and leads to improvements over continuous practice. Most theories propose that it works because it affects the formation of memories. For example, Benjamin and Tullis [28] argue that spaced practice works in part because of encoding variability — the idea that if two events are further apart in time, then the second event is likely to get encoded a second time in a different way [28, 303]. The chance of this happening is therefore thought to be higher if the time between events is increased [303]. Encoding variability is beneficial because if the information is encoded in multiple ways, it is thought to also give a learner more ways of retrieving the relevant information [303, 97]. Another possibility is the study-phase-retrieval mechanism, where the memory of a repeated item is strengthened if that item is presented again for study at a later time [303].

Some researchers have proposed that the effectiveness of a rest period is affected by how long it is, for example, Schmidt and Lee [253] state: "short rest periods degrade performance relative to performance with longer rest periods". Others, however, state that performance follows an inverted U function [48, 81]. Some research also suggests that spaced practice works best when the length of the rest increases over time [49]. It might be that the rest length is affected by how well-learned a task is [201]. For example, the retrieval effort hypothesis suggests that practice with the intention of strengthening memory is best done when retrieval of the relevant memory is difficult but can still succeed [238, 201] and retrieval effort can be manipulated by increasing the time between retrievals [201].

The relationship between break length and progress in skill learning could be why no specific break period performed better than any other in Manuscript C. To maximize the effectiveness of spaced practice, the breaks may have needed to be delivered dynamically, by considering how well the player was performing in the game. The rest intervals used in Manuscript D did this, although they were also more frequent (breaking up the play session more often) and shorter. They were also dynamic. It might be that as players got better at the game, they made fewer mistakes and therefore received fewer rest periods. However, if they ever came upon a challenge they couldn't handle, then they got a break. This is different than the suggestions that rests become longer as a learner becomes more skilled because the system is manipulating how often breaks are given as well as how long the breaks are (although the breaks are relatively short).

It is also thought that the underlying reasons why spaced practice is effective change when the rest period becomes relatively short (less than 15 seconds) [303, 132]. For example, Hintzman [132] proposed that less attention is given to the second attempt at rehearsing something, compared to the first attempt, and that a longer interval between attempts prompts a learner to give more of their voluntary attention toward the task. Similarly, the deficient-processing theory explains that the benefits of short time spacings come about due to a learner being unable to as effectively process a second rehearsal, so a time delay between rehearsals improves processing [303].

Apart from theories relating to memory and spaced practice in general, the participants in Manuscript D were doing things during their breaks that may have helped them improve at the game. In particular, some participants used the time to reconsider what they were doing within the game. The breaks also prompted players to approach the challenges in the game with more caution, possibly because needing to wait to try again was viewed as a form of punishment.

11.2.3 How Checkpoints Affected Skill Learning in Manuscript D

In Manuscript D, participants were given checkpoints as a possible support method to help them make progress within the game. Within the manuscript, I talk about how checkpoints could interact with different factors that could affect skill development. This included skill-challenge balance, variability of practice, and part-task practice.

Checkpoints are designed to help players continue to make progress within the game. Manuscript D confirmed that this is the case, playing with checkpoints enabled allowed participants to make more progress within the game than without. It however did not have a long terms effect on performance, so learning was ultimately unaffected. It is possible that checkpoints simultaneously provide benefits to learning as well as drawbacks to learning that effectively "cancel each other out" in the long term while providing a short-term benefit.

In terms of the benefits, checkpoints are likely to provide players with a form of part-task practice, which is thought to benefit perceptual-motor skill learning [253]. As players go up against a challenge and fail, they will be taken back to that challenge more promptly, without needing to overcome other obstacles before attempting the obstacle again. This should give players opportunities to try out different strategies and receive feedback about how effective those strategies are more quickly than without checkpoints.

Additionally, because checkpoints allow players to make progress more quickly they get exposed to more levels of the game, which exposes them to more variations of obstacles. This increases the variability of practice, which is thought to be beneficial to perceptual-motor skill learning [253]. However, the later levels in the game are not simply different, they are also more difficult. The skill-challenge balance of a game is known to affect learning — a challenge that is well-matched to a player's ability is a better learning environment [332] than when that challenge is mismatched. Having players further along within the game then they may be ready for may therefore be negatively affecting learning.

11.3 Contributions

At the start of this chapter, I stated that I had two goals with the manuscripts:

- To better understand why existent approaches to skill learning used within games work by examining them through the lens provided by past research on perceptual-motor skill development.
- To explore ways of supporting skill learning within games that are novel in the games domain, but not novel in the perceptual-motor skill learning domain.

Within the manuscripts, each state their contributions in their own way. In this section of the discussion, I provide a summary of these contributions contextualized by considering how they contribute to my goals. I consider my goals as they relate to game design and also how the manuscripts contribute to the underlying theories from the domain of perceptual-motor skill development.

11.3.1 Contributions to Game Design

I leveraged existing theories of skill development to better understand how skill development occurs in games as well as explore new ways of supporting players. The work presented in this dissertation provides several valuable lessons for others who wish to attempt the same within their own games. I found that guidance, spacing practice, and using checkpoints can help players improve their immediate performance, meaning that games can be made more accessible to players and they can be more effective within the game while they are learning how to play.

Guidance

In the background chapter, *guidance* was a support method introduced as a possible mechanism for helping players improve at a game. Guidance is something that can already be found in games, though it has not been evaluated experimentally in the games domain in much detail. The type of guidance explored in Manuscripts A and B (concurrent visual and mechanical guidance) has only been examined in one experiment that I am aware of, where the effects of aim assistance on learning were evaluated [122]. In that experiment, aim assistance helped when it was present and did not negatively affect learning how to aim without assistance.

This dissertation explored guidance in the context of the skill of *navigation*. As a game skill, navigation is an important aspect of many games, as knowing where to go is often the first step to accomplishing anything within a game. It is also interesting from a research perspective because it is a skill that people carry out in everyday life. Most people can navigate the environment around them without difficulty, yet if you put those same people into a virtual environment you would find that many of them struggle [68, 156]. Because of this, games often include various forms of navigation guidance. Manuscript A explored navigation guidance in the form of maps, GPS, and trails embedded into the environment. These are all systems currently present in games, though obvious glowing trails within the environment are only found in a small number of games [188] and typical implementations of trails are more subtle than what was used in my studies (e.g., collectable items might be placed linearly within an environment to indicate the path forward, or objects within the environment that players are expected to move toward will be brightly coloured [208]). In addition to existing forms of assistance, Manuscript B explored a system of navigation guidance not generally found in games, taking the player on a rail through the environment. Again, this isn't entirely unheard of within games (some shooter games are "on rails" and players don't navigate at all, e.g., *House of the Dead* [262]), but an analogue to this is showing a player where to go via a cutscene [208]. This "on rails" guidance functioned as a sort of "mechanical" guidance system, something that has been studied extensively in the perceptual-motor skill learning domain.

Over two studies, Manuscript A found that providing additional visual navigation guidance did not interfere with learning the environment, regardless of how much time was spent with that guidance. Further, the addition of guidance allowed players to more effectively navigate the environment while the guidance was present. Manuscript B confirmed this to still be the case, but less so. With no assistance at all, players did end up performing better on tests; that is, they had learned more. That extra learning came at a very steep cost, however, as not being given assistance meant that training took about over twice as long as the trails group. Additionally, the gains in performance in testing were less pronounced than the gains that having assistance during training provided — it seems that players who trained with trails did learn the environment even if they were less successful at learning the environment than not having assistance. For the more extreme and slightly novel "on rails" guidance in Study two Manuscript B, the results are identical to the trail results just presented.

In terms of actionable information for game designers, the results of Manuscript A suggest that there is no downside to providing players with navigation assistance and so game designers should do this without delay. However, the results of Manuscript B provide a more nuanced view of the situation. Study two of Manuscript B did find that having *no* guidance resulted in improved learning of the environment when forming *new routes* within the environment, but no significant benefit in terms of traversing *previously learned routes* without guidance. Therefore, if a game is going to task a player with forming new routes to in-game locations, then there is a *slight* downside to providing guidance. However, even with this downside, the performance gains while guidance is present and the evidence suggesting that players do still learn the environment suggests that game designers should not hesitate to add this type of guidance into their games. Having said that, if being able to navigate through an environment in novel ways is a key part of a game's experience (as found in games like *Dark Souls* [108]), then a game designer may want to make the decision to omit navigation assistance, even though this will make a game more difficult.

Considering how navigation assistance affects a game's difficulty is also an important thing to consider if a game has several other difficult skills to master in addition to navigation skill. For such a game, a game designer may want to provide navigation assistance to the player so that the player can focus on learning other skills within the game. By decreasing the difficulty of the task overall through navigation assistance the learning environment might be enhanced by bringing the task closer to an optimal level of difficulty for learning [332]. This may allow players to direct their limited attention toward skills other than navigation. As a bonus, our results also suggest that some incidental spatial learning will occur, so if at a later point the navigation assistance is withdrawn, then players will have learned the environment at least somewhat.

Finally, if navigation guidance is to be withdrawn at some point, it need not be done all at once or abruptly. In the case of trail guidance, for example, the trails can be gradually faded out, presented less often, or made less obvious in some other way.

Modifying Practice

In the introduction, I describe the scenario in which a player knows what needs to be done, and has acquired the necessary skills required to succeed, but yet they still struggle. In this situation, practice can be modified to help a player out so that they can make progress within the game or train in a way that leads to improved learning.

Manuscript C explored how *spaced practice* could be used to modify practice within the game *Super Hexagon*.

Spaced practice is an approach that has been explored for many tasks other than games [81], but to the best of my knowledge is not currently leveraged by commercial games. We found that spacing practice resulted in immediate and obvious gains in performance. Taking a break as short as two minutes meant that players performed better in the next five minutes of play. Additionally, the length of the break appeared to not matter so much, so long as there was a break. Breaks of two minutes, five minutes, and ten minutes all resulted in immediate performance gains compared to playing continuously. This initial work exploring spaced practice in games demonstrated that there is a lot of potential for applying spaced practice as a beneficial aid for when players are struggling, yet provided little in terms of how spacing ought to be incorporated into a game.

This was addressed in Manuscript D, wherein I tested what the manuscript calls "game-integrated breaks". These are breaks that are given to the player after each failure, and the time is determined by the length of time the player was alive, up to a maximum of ten seconds. These breaks are shorter but more frequent and incidentally encouraged players to play the game in a more risk-averse way. Like the breaks in Manuscript C, these breaks improved player performance (when controlling for total training time) when they were present, and did not affect learning. This suggests that spaced practice can be implemented in games in a variety of ways, not just via a fixed schedule.

However, one downside of breaks is that players tend to dislike them. In Manuscript D, players were asked to share how they felt about breaks and the response was generally negative, or at best, neutral. A game designer must consider then if a player's objective performance improvements are worth the potential subjective experience cost of implementing breaks. It should also be noted, however, that the breaks implemented in the game used in the study were simple timers counting down to zero — very obvious and simplistic. More interesting breaks might be received more warmly by players. In fact, other work suggests that the breaks do not need to be as simple as a timer counting down. As a side project not included as part of this dissertation, we (Piller et al. [231]) investigated what players should be doing with their time during a break. Three different breaks were used: an interactive narrative dialogue, a grapple mini-game, and a maze mini-game. This work found that all break types resulted in a similar performance, and when compared to a timer counting down, there was also little difference.

Overall, spaced practice *can* help players who know what to do but are simply struggling. While the best way to present it is not entirely clear, it is clear that the effect is robust enough that it can provide benefits when presented in a variety of different ways. For example, they can be short, and still effective. We also know that players can continue to engage with the game during the break rather than needing to switch away from the game entirely.

Manuscript D also looked at modifying practice in a different way, through saving progress with *checkpoints*. Checkpoints can modify practice by affecting the variety of practice, as well as facilitating the process of part-task practice. Checkpoints are very common in commercial games and so the checkpoints were modelled on those found in many commercial games. Despite how common checkpoints are, to my knowledge, this is the first study that explores whether checkpoints affect performance and learning. Checkpoints do in fact help players make progress and improve immediate performance, and if they do affect learning, the net result is no different than playing without checkpoints. Overall, checkpoints should not be considered a crutch for weak players, and they can be a valuable tool for helping players make progress within the game.

11.3.2 Contributions to Theory

The work presented in these manuscripts was designed primarily to investigate applying theories of learning and skill development to learning within games. However, there are still takeaways from this work that adds to our understanding of the theories.

Guidance

Quite a bit of research that looks into applying guidance to help learners found that it is harmful to learning [257, 16] because learners do not attend to the task's inherent response-produced feedback [288, 257] and instead become reliant on the guidance [249, 248]. It seems that this phenomenon is not present in the context of navigation. There are two possible reasons for this. First, navigation is a task that is carried out without much in terms of feedback — it isn't immediately obvious to the navigator when they have made a wrong turn. Second, navigation learning appears to be something that can occur incidentally — a navigator can simply follow a glowing line or be taken automatically in the correct direction and they are able to retrace those steps later on.

Spaced Practice

My results indicate that spaced practice is a robust and beneficial under a variety of settings. It worked under two different games with a variety of different rest periods. First, I showed that it was beneficial to players of a clone of *Super Hexagon* as well as a bespoke side-scrolling platform game. I also showed that it can be implemented without a fixed schedule but instead dynamically in response to in-game events and continue to be beneficial. Those dynamic breaks could also be short (less than ten seconds) and spaced practice continues to be beneficial over playing continuously.

12 Thoughts on Future Work

The work presented in this dissertation took several steps toward solving the problem outlined in the introduction, that players want to get better at the games they play yet game designers do not always know how to explicitly support skill development. If work were to continue on this problem, there are several follow-up steps to take to conduct future research into this topic.

12.1 Further Exploration of the Support Methods

The work in this dissertation looked at guidance in the context of navigation (Manuscripts A and B) and spaced practice in the context of *Super Hexagon* (Manuscript C) and a side-scrolling platform game (Manuscript D). It also incidentally looked at part-task practice and variety of practice (Manuscript D). However, this is only scratching the surface.

12.1.1 Demonstrations

The work in this dissertation explored concurrent visual guidance in the context of game navigation, but there are many other contexts where guidance could be beneficial for gamers as well as contexts where guidance is *already* found within games. As outlined in the background chapter, visual guidance that that precedes an action takes the form of videos, charts, visual aids, or demonstrations [136, 216]. Demonstrations can be found during an arcade cabinet's attract mode (short pre-recorded gameplay segments that play when nobody interacts with the cabinet [306]) and serve to demonstrate how one might play the game. If the arcade cabinet was occupied, one might watch another player play through the game. This concept of watching another person play the game is very common these days, with many players watching one another play on social media platforms like Twitch or YouTube.[1] This also occurs within games themselves. For example, in multiplayer games, it is quite common for one player to spectate another. If the results about the efficacy of demonstrations for other perceptual-motor skills also apply to digital games, then watching others in this way would serve as a valuable learning opportunity for new players.

12.1.2 Spaced Practice

Spaced practice shows much potential as a mechanism for strategically providing players with a performance boost, and the experiment in Manuscript D demonstrated one method for integrating spaced practice into a commercial game. However, there are still some potential avenues for future research.

[1] Hundreds of thousands of people watch live streams of games on `http://twitch.tv` daily.

Figure 12.1: The different break types used within [231]. Left: Cutscene break. Center: Grapple mini-game. Right: Maze mini-game.

First, future work might explore how the presentation of the break. Written comments from our participants in Manuscript D did show that the break was somewhat unwelcome. It might be possible to present the break in a way that makes it less obvious that it is a break. The presentation in Manuscript D was simply to display numbers counting down, which would not be difficult to improve upon. It also may be possible to have players engage with the game in a completely different way during the break. For example, work that explored having players do a different task during a two-minute break and found benefits that were similar to spaced practice (see Figure 12.1) [231].

Second, if it is possible to have players do something else during the break while receiving benefits to performance, might it be possible to have players do something productive? For example, what might happen if, instead of presenting players with a number counting down, they were presented with a recording of what they were doing the previous ten seconds before their death? A "break" of this type might also serve as a demonstration, so would it then be more beneficial than a break alone?

12.1.3 Task Decomposition

A series of papers from 1989 used the purpose-built research game *Space Fortress* (Figure 2.10) to study the acquisition of complex skills [79, 187]. A feature of many approaches was to make use of *task decomposition*, where the game was decomposed into individual components that players could focus on [99, 118, 107]. One approach looked at part-task practice where players were placed in an environment where each component of the game or a subset of the game was practised in isolation [107], while another instructed players to direct their attention towards specific parts of the game [118]. Both approaches were effective, though the part-task practice approach has the drawback that players eventually need to transfer to the full version of the game with all components of the game present, and when they do there is a slight initial drop in performance [99].

To my knowledge, no more recent work has properly revisited the ideas presented in these papers. Games are significantly more complex now than they were in 1989, but this older work can still provide inspiration for designing training systems for our modern, more complex games. The idea of task decomposition is central to making this work. The closest we have gotten to recently has been the notion of creating "skill chains" that are made up of "skill atoms" (e.g., [75, 138, 22, 203], see also Section 2.1), but there has not yet been work that demonstrates the efficacy of instructing players based on the specific skill chains identified.

12.2 Incorporating Other Players Into the Learning Process

The introduction pointed out that in multiplayer games, a game's difficulty is determined primarily by the other players. When a novice wants to play a game with their friend, but that friend is an expert, a mismatch in expertise is the result. This can lead to unbalanced teams or a lack of outcome uncertainty, which makes the game less enjoyable for both players [2]. However, this problem could be turned into an asset if the stronger player is willing to instruct the weaker player. This has been observed, for example, among siblings, where the older sibling acts as instructional scaffolding for the younger sibling when they try to play a game together [117]. If the players are willing, then the same can be done in online play. Weaker players could *intentionally* pair up with stronger players with the goal of seeking instruction.

In a different scenario, a weaker player may want to play a game but not have anyone in particular to play it with. This player may find matches through the game's matchmaking system. This system has difficulty matching new players who do not have a rating and so these players' first matches are likely to be unbalanced and therefore frustrating. Alternatively, the game could provide players with an online training option, where the player could learn the game alongside other novices, essentially a collaborative learning environment [169]. If a player matches with other players who are at similar skill levels, the player should be in a better learning environment [332] in which they are motivated to improve the execution of their skills. This would be particularly true if all players involved are opting to play for the same reason: to get better at the game.

Either of these approaches — receiving help from an experienced player or learning alongside other players — results in increased relatedness compared to playing alone [247].

12.3 Complex Games

Games that are more complex present unique challenges in terms of finding opportunities to support a player's learning and how to know if any support being provided has actually been helpful. Solving the problem outlined in the dissertation for more complex games than those used in the manuscripts could require additional solutions.

12.3.1 Finding Opportunities to Support Learning

Chapter 2 discussed how games that are linear in presentation present more opportunities for scaffolding skill development. Leveraging this linearity to scaffold players and support their learning is very common, however, not all times are designed in this way. For games with different designs, support needs to be designed into the game at different times.

Multiplayer games, particularly competitive ones, tend not to include this linearity, so game designers need to explore other opportunities for in-game timing to support skill development. One aspect of these types of games is that they often include repetitive activities that are completed in roughly the same sequence every time. This repetition could serve as a sort of short time-span linearity and be leveraged to provide the player with a simple real-time tutorial.

For example, in *Counter-Strike* [130], each game round starts with the same sequence of actions (buying weapons, moving to the objective, and then engaging in a firefight with the enemy team to defend a bomb site or plant a bomb). As players get better at the game, they eventually learn to buy specific weapons and move to specific locations on the map. It would be possible to support players in making these decisions using systems already presented in this dissertation. For example, a just-in-time tool tip could tell the player to purchase a specific weapon based on their available funds, then a glowing trail could tell the player how to reach the objective on the map. As the game progresses, the guidance would then need to dynamically adjust depending on the state of the game. For example, if the bomb were dropped, the player would need to be directed toward the bomb to either defend it or recover it. Essentially, the goal would be to develop a real-time system to "coach" the player into making better decisions that is made up of components that are proven to work in isolation.

Open-world games present similar challenges for supporting learning as multi-player games. It is difficult to design a specific tutorial to teach a player a specific skill when you are not sure when a player might require that instruction. Future work could explore this topic in more detail and classify different ways in which a player's progress in a game becomes stalled.

12.3.2 What Does it Mean to "Get Better"?

The problem my dissertation aimed to solve (that players want to get better at the games they play yet game designers do not always know how to explicitly support skill development), cannot be effectively evaluated unless one can quantify specifically what it means to get better at a specific game. If we want to study a support method and determine its efficacy, we need to know *specifically* what it means to "get better" at a game.

In Chapter 2, "skill" was defined for the context of this dissertation as "a task carried out to accomplish a specific goal with efficiency and a high success rate". With this definition, an entire game is unlikely to be a skill unless that game is quite simple. For example, the *Super Hexagon* clone used in Manuscript C is a game made up of one skill, so getting better at *Super Hexagon* happens to be the same as getting better at the skill of dodging obstacles. But the game used in Manuscript D involves running, jumping, wall-jumping, and grappling. Therefore, getting better at that game involves improving one's performance in all of those skills. Even though the game involves multiple skills, performance was still boiled down into a single variable (number of levels completed) because there was a single goal within the game. Therefore, knowing whether players got better was easy enough to determine. But what about more complex games with multiple goals?

There is at least some interest in studying skill development in more complex games [283, 33] (in particular, Boot et al. [33] also advocate for moving beyond aggregate measures of performance), and so this is a problem that researchers will need to address at some point. There are several approaches that could be taken. For example, some games, such as older arcade games, attempt to boil down performance in multiple goals into a single measure (which would then be found on a high-score table [35]). This is also the approach used for the research game, *Space Fortress* [79]. However, this approach has the inherent problem that it introduces some bias in what aspects of a game a training system will instruct players on. Past research looking into comparing speed to accuracy as performance metrics, for example, has found that instructing learners to focus on accuracy will result in improved accuracy at the cost of speed and vice versa [279]. For example, take a simple arcade game such as *Space Invaders*. The player's main goal is to avoid incoming projectiles while shooting back at enemies. However, every so often a "mothership" passes by that provides bonus points. Destroying the mothership does not help the player work toward the goal of defeating enemies to progress to the next level, but does raise a player's score. Is a player who prioritizes destroying motherships actually better at the game, even if that goal does not contribute to their progress within the game?

Complex tasks involve multiple skills that can be learned or trained in isolation. When studying complex games, it might be advantageous to consider studying the component skills, or at least study performance in terms of specific in-game goals. One could then demonstrate that the considered training system is effective in isolation before trying to devise a system that trains a player in all of a game's skills. As an example, consider how Frederiksen and White trained *Space Fortress* players in individual in-game skills in isolated environments before having them play the game as a whole [107]. In particular, they trained players on the skills of hitting fortresses without being hit themselves, detecting and destroying mines, and allocating resources to maximize score. The training scenarios for those individual skills all included high-quality feedback that could tell the player how well they were doing as well as be used to measure how well a player was executing that skill. With all of these training environments developed, it then becomes possible to determine not only if a player is good at the game, but the ways in which they are good at the game. Therefore, I argue that if future work were to study more complex games, then it should avoid boiling down all aspects of a player's performance into one dependent variable and instead focus on a few key variables that are indicators of success at specific goals within the game.

12.3.3 Game Selection

In terms of *which* games to study, I would suggest researchers explore the option of making use of or modifying existing open-source games[2]. Of particular interest might be games that were originally released as a commercial game, but later had their source-code released[3].

[2]See https://en.wikipedia.org/wiki/List_of_open-source_video_games

[3]See https://en.wikipedia.org/wiki/List_of_commercial_video_games_with_later_released_source_code

13 Conclusion

The introduction introduced the idea that players want to get better at the games they play, and that game designers don't always know how to facilitate skill development in their games. This dissertation takes several steps in the direction of showing how game designers can support in-game skill development and understanding why some of our current approaches (such as checkpoints) work.

The first step was to outline how skill development occurs in general, and the different ways in which it can be supported. This included guidance, feedback, spaced practice, directing attention, part-task practice, and variety in practice. By categorizing and discussing these methods, this dissertation contributes to the understanding of how game designers can facilitate skill development within their games. Further, I also described how skill development is currently handled within commercial games. I outlined the strengths of the current approaches and where those current approaches are lacking. This allowed me to contribute new ways of supporting skill development that avoid potential pitfalls of past approaches.

The second step was an examination of using guidance to support players as they learn a new skill. The skill was one that is important, but one that new players generally struggle with — navigation. Manuscripts A and B showed that guidance substantially improved performance — navigating an environment while following a glowing trail was 2.4 times as fast as navigating with just a map, and navigating with either glowing trails or being taken in the correct direction on rails was 2.6 times as fast as navigating with no assistance. Furthermore, this assistance came at little cost in terms of learning the environment. The gains of playing with assistance strongly outweigh any drop in performance if the assistance were ever taken away. In Manuscript A, Study 2, the groups were all within 16 seconds of each other when navigating new routes, on a navigation task that took about 7 minutes on average. In Manuscript B, Study 2, when navigating new routes, the fastest group (no assistance) was 1.4 times as fast as the slowest (trail assistance). This difference was about 2 minutes in actual navigation time (6.17 minutes versus 8.25 minutes). When navigating the previously seen routes, the differences were smaller — the fastest group was only about 1.2 times faster than the slowest, a difference of about a minute (5 minutes versus 6.17 minutes). These differences in training are much smaller than those present when the assistance was present. When Manuscript A was published (2017), navigation in virtual environments and digital games had been studied extensively, but no work had looked at what happens if the assistance was taken away. Based on the guidance hypothesis, it was thought that a player would become reliant on any assistance (e.g., [249, 255]), but we showed that this was not necessarily the case. This work contributes evidence that navigation assistance can be a valuable tool to support new players in games and one that comes with little downside.

The third step was an exploration of spaced practice to help improve players' performance. This step involved a study that replicated past findings on spaced practice, applying them to a new context that had not yet been properly explored. Manuscript C showed that spaced practice improves immediate performance over continuous practice, with little change to long-term performance (i.e., learning). This is an important contribution because during gameplay, immediate performance improvements could allow a player to overcome a tough challenge and progress in the game beyond what would have been possible without the immediate performance gain. This is additionally important because it goes against many players' desires or intuition; the last thing many players want to do is to stop playing, and yet doing so for a short time could help them. Furthermore, past work on spaced practice has shown that it applies very generally, that is, it seems to be beneficial when considering many different skills [81]. If it is helpful in one digital game, it is likely helpful in others.

The fourth step was to test the idea that spaced practice would be beneficial for different games and under different circumstances, and directly help a player make progress within a game. Manuscript D explored the concept of "game-integrated breaks" — dynamic breaks presented after each failure rather than scheduled to occur at specific intervals like in prior implementations of spaced practice. The robustness of spaced practice was further put to the test by making the breaks shorter (no more than ten seconds long) and testing them with a game of an entirely different genre. Manuscript D demonstrates that spaced practice is a robust effect that can be applied in a variety of different ways, at different times, and in different games, and yet it is still beneficial. It also demonstrates what we proposed in Manuscript C, but did not demonstrate explicitly — that spaced practice can help players make progress within a game, including by helping them make more consistent progress, get stuck fewer times, and complete more levels.

Finally, Manuscript D also compared this new approach for helping players make progress in a game (i.e., game-integrated breaks) to a common approach in current commercial games — the use of checkpoints. Manuscript D also uses the theories presented in this dissertation to explain how checkpoints could affect a player's learning and performance. Checkpoints could be considered to modify practice by introducing part-task practice (allowing players to focus on specific challenges they are struggling with) and increasing the variety of practice (players that progress further in the game will be exposed to a greater variety of obstacles than if they were not able to make progress in the game). In Manuscript D, we found that checkpoints led to similar progress improvements in the game as did game-integrated breaks, without introducing any detrimental effects on learning (just like game-integrated breaks). This comparison shows that new techniques for supporting in-game progress can be developed that are just as effective as existing techniques.

This dissertation takes several steps toward increasing our understanding of how to support in-game skill development. The greater our understanding of how to accomplish this, the more interesting and complex games can become.

Part IV

References and Appendices

References

[1] Vero Vanden Abeele, Katta Spiel, Lennart Nacke, Daniel Johnson, and Kathrin Gerling. 2020. Development and validation of the player experience inventory: A scale to measure player experiences at the level of functional and psychosocial consequences. *International Journal of Human-Computer Studies* 135 (3 2020), 102370. https://doi.org/10.1016/j.ijhcs.2019.102370

[2] Sami Abuhamdeh, Mihaly Csikszentmihalyi, and Baland Jalal. 2014. Enjoying the possibility of defeat: Outcome uncertainty, suspense, and intrinsic motivation. *Motivation and Emotion* 39, 1 (sep 2014), 1–10. https://doi.org/10.1007/s11031-014-9425-2

[3] Phillip L Ackerman, Ruth Kanfer, Stacey W Shapiro, Sunni Newton, and Margaret E Beier. 2010. Cognitive Fatigue During Testing: An Examination of Trait, Time-on-Task, and Strategy Influences. *Human Performance* 23, 5 (10 2010), 381–402. https://doi.org/10.1080/08959285.2010.517720

[4] E. Adams. 2013. *Fundamentals of Game Design*. Pearson Education.

[5] Jack A. Adams. 1987. Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. *Psychological Bulletin* 101, 1 (1987), 41–74. https://doi.org/10.1037/0033-2909.101.1.41

[6] Susan A Ambrose, Michael W Bridges, Michele DiPietro, Marsha C Lovett, and Marie K Norman. 2010. *How learning works: Seven research-based principles for smart teaching*. John Wiley & Sons.

[7] Robert B. Ammons. 1950. Acquisition of motor skill: III. Effects of initially distributed practice on rotary pursuit performance. *Journal of Experimental Psychology* 40, 6 (1950), 777–787. https://doi.org/10.1037/h0061049

[8] Erik Andersen, Eleanor O'Rourke, Yun-En Liu, Richard Snider, Jeff Lowdermilk, David Truong, Seth Cooper, and Zoran Popovic. 2012. The impact of tutorials on games of varying complexity. *SIGCHI Conference on Human Factors in Computing Systems* Pajitnov 1984 (2012), 59–68. https://doi.org/10.1145/2207676.2207687

[9] Fraser Anderson and Walter F. Bischof. 2013. Learning and performance with gesture guides. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. https://doi.org/10.1145/2470654.2466143

[10] John R. Anderson. 1982. Acquisition of cognitive skill. *Psychological Review* 89, 4 (1982), 369–406. https://doi.org/10.1037/0033-295x.89.4.369

[11] Jackie Andrade and Peter Meudell. 1993. Short Report: Is Spatial Information Encoded Automatically in Memory? *The Quarterly Journal of Experimental Psychology Section A* 46, 2 (may 1993), 365–375. https://doi.org/10.1080/14640749308401051

[12] Mathieu Andrieux and Luc Proteau. 2016. Observational Learning: Tell Beginners What They Are about to Watch and They Will Learn Better. *Frontiers in Psychology* 7 (jan 2016), 51. https://doi.org/10.3389/fpsyg.2016.00051

[13] Angel Studies. 2000. Midtown Madness 2. Game [PC].

[14] ArcadeControls.com. 2022. Donkey Kong Instruction Sticker Vector Rework. Retrieved September 21, 2022 from http://forum.arcadecontrols.com/index.php?topic=161126.0

[15] Anthony J. Aretz and Christopher D. Wickens. 1992. The Mental Rotation of Map Displays. *Human Performance* 5, 4 (dec 1992), 303–328. https://doi.org/10.1207/s15327043hup0504_3

[16] Terry R Armstrong. 1970. Training for the Production of Memorized Movement Patterns. In *Proceedings*, Vol. 6. Scientific and Technical Information Division, National Aeronautics and . . . , 449.

[17] Graham Aston. 2017. What is the Optimum Training Time for Esports Players? *The Esports Observer* (December 2017). Retrieved April 4, 2019 from https://esportsobserver.com/optimum-player-training-time/

[18] Atari. 1972. Pong.

[19] Atari Games. 1985. Gauntlet. Game [Arcade].

[20] Auction Game Sales. 2018. The Golden Age of Arcade Games – A Brief History. https://auctiongamesales.com/2018/12/17/the-golden-age-of-arcade-games-a-brief-history/

[21] Automated Simulations. 1979. Temple of Apshai. Game [PC].

[22] Batu Aytemiz, Isaac Karth, Jesse Harder, Adam M Smith, and Jim Whitehead. 2018. Talin: a framework for dynamic tutorials based on the skill atoms theory. In *Fourteenth Artificial Intelligence and Interactive Digital Entertainment Conference*.

[23] A. D. Baddeley and D. J. A. Longman. 1978. The Influence of Length and Frequency of Training Session on the Rate of Learning to Type. *Ergonomics* 21, 8 (aug 1978), 627–635. https://doi.org/10.1080/00140137808931764

[24] Jonathan Z. Bakdash, Sally A. Linkenauger, and Dennis Proffitt. 2008. Comparing Decision-Making and Control for Learning a Virtual Environment: Backseat Drivers Learn Where They are Going. https://doi.org/10.1037/e578342012-007

[25] Scott Bateman, Regan L. Mandryk, Tadeusz Stach, and Carl Gutwin. 2011. Target assistance for subtly balancing competitive play. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*. ACM Press. https://doi.org/10.1145/1978942.1979287

[26] Batman: Arkham Wiki. 2021. Game over scenes. Retrieved July 22, 2021 from https://batmanarkham.fandom.com/wiki/Game_over_scenes

[27] Roy F. Baumeister. 1984. Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology* 46, 3 (1984), 610–620. https://doi.org/10.1037/0022-3514.46.3.610

[28] Aaron S. Benjamin and Jonathan Tullis. 2010. What makes distributed practice effective? *Cognitive Psychology* 61, 3 (nov 2010), 228–247. https://doi.org/10.1016/j.cogpsych.2010.05.004

[29] Bethesda Game Studios. 2011. The Elder Scrolls V: Skyrim. Game [PC].

[30] Blizzard Entertainment. 1998. StarCraft. Game [PC].

[31] Blizzard Entertainment. 2004. World of Warcraft. Game [PC].

[32] Vicarious Visions Blizzard Entertainment. 2021. Diablo II: Resurrected. Game [PC].

[33] Walter R. Boot, Anna Sumner, Tyler J. Towne, Paola Rodriguez, and K. Anders Ericsson. 2016. Applying Aspects of the Expert Performance Approach to Better Understand the Structure of Skill and Mechanisms of Skill Acquisition in Video Games. *Topics in Cognitive Science* 9, 2 (nov 2016), 413–436. https://doi.org/10.1111/tops.12230

[34] Lyle E. Bourne and E. James Archer. 1956. Time continuously on target as a function of distribution of practice. *Journal of Experimental Psychology* 51, 1 (1956), 25–33. https://doi.org/10.1037/h0039887

[35] Jason T. Bowey, Max V. Birk, and Regan L. Mandryk. 2015. Manipulating Leaderboards to Induce Player Experience. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play* (London, United Kingdom) *(CHI PLAY '15)*. ACM, New York, NY, USA, 115–120. https://doi.org/10.1145/2793107.2793138

[36] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (jan 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[37] Gerard J.P. Van Breukelen. 2006. ANCOVA versus change from baseline had more power in randomized studies and more bias in nonrandomized studies. *Journal of Clinical Epidemiology* 59, 9 (sep 2006), 920–925. https://doi.org/10.1016/j.jclinepi.2006.02.007

[38] Gerard J.P. Van Breukelen and Koene R.A. Van Dijk. 2007. Use of covariates in randomized controlled trials. *Journal of the International Neuropsychological Society* 13, 05 (aug 2007). https://doi.org/10.1017/s1355617707071147

[39] Ryan Brydges, Heather Carnahan, David Backstein, and Adam Dubrowski. 2007. Application of Motor Learning Principles to Complex Surgical Tasks: Searching for the Optimal Practice Schedule. *Journal of Motor Behavior* 39, 1 (jan 2007), 40–48. https://doi.org/10.3200/jmbr.39.1.40-48

[40] budda loves you. 2015. Titan Souls General Discussions - Is there any reason why it doesn't just spawn you outside of the boss door when you die. *Steam Community* (2015). Retrieved September 13, 2018 from https://steamcommunity.com/app/297130/discussions/0/618458030655379691/

[41] Stefano Burigat and Luca Chittaro. 2007. Navigation in 3D virtual environments: Effects of user experience and location-pointing navigation aids. *International Journal of Human-Computer Studies* 65, 11 (nov 2007), 945–958. https://doi.org/10.1016/j.ijhcs.2007.07.003

[42] Gary E. Burnett and Kate Lee. 2005. The Effect of Vehicle Navigation Systems on the Formation of Cognitive Maps. In *Traffic and Transport Psychology*. Elsevier, 407–418. https://doi.org/10.1016/b978-008044379-9/50188-6

[43] Capcom. 1993. Mega Man X. Game [SNES].

[44] Stuart K Card, Allen Newell, and Thomas P Moran. 1983. *The Psychology of Human-Computer Interaction*. L. Erlbaum Associates Inc.

[45] Terry Cavanagh. 2012. Super Hexagon. Game [PC].

[46] Terry Cavanagh. 2012. *Super Hexagon*. Retrieved August 30, 2018 from https://www.superhexagon.com/

[47] Jared E. Cechanowicz, Carl Gutwin, Scott Bateman, Regan Mandryk, and Ian Stavness. 2014. Improving player balancing in racing games. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play*. ACM. https://doi.org/10.1145/2658537.2658701

[48] Nicholas J. Cepeda, Noriko Coburn, Doug Rohrer, John T. Wixted, Michael C. Mozer, and Harold Pashler. 2009. Optimizing Distributed Practice: Theoretical Analysis and Practical Implications. *Experimental Psychology* 56, 4 (jan 2009), 236–246. https://doi.org/10.1027/1618-3169.56.4.236

[49] Nicholas J. Cepeda, Harold Pashler, Edward Vul, John T. Wixted, and Doug Rohrer. 2006. Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin* 132, 3 (2006), 354–380. https://doi.org/10.1037/0033-2909.132.3.354

[50] William G. Chase. 1983. Spatial Representations of Taxi Drivers. In *The Acquisition of Symbolic Skills*. Springer US, 391–405. https://doi.org/10.1007/978-1-4613-3724-9_43

[51] Chien-Hsiung Chen, Wen-Chih Chang, and Wen-Te Chang. 2009. Gender differences in relation to wayfinding strategies, navigational support design, and wayfinding task difficulty. *Journal of Environmental Psychology* 29, 2 (jun 2009), 220–226. https://doi.org/10.1016/j.jenvp.2008.07.003

[52] Jenova Chen. 2007. Flow in games (and everything else). *Commun. ACM* 50, 4 (apr 2007), 31. https://doi.org/10.1145/1232743.1232769

[53] Jui Lin Chen and Kay M. Stanney. 1999. A Theoretical Model of Wayfinding in Virtual Environments: Proposed Strategies for Navigational Aiding. *Presence: Teleoperators and Virtual Environments* 8, 6 (dec 1999), 671–685. https://doi.org/10.1162/105474699566558

[54] Gifford K. Cheung, Thomas Zimmermann, and Nachiappan Nagappan. 2014. The First Hour Experience: How the Initial Play Can Engage (or Lose) New Players. *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play - CHI PLAY '14* (2014), 57–66. http://dl.acm.org/citation.cfm?id=2658537.2658540

[55] Elizabeth R. Chrastil and William H. Warren. 2011. Active and passive contributions to spatial learning. *Psychonomic Bulletin & Review* 19, 1 (nov 2011), 1–23. https://doi.org/10.3758/s13423-011-0182-x

[56] Andy Cockburn, Carl Gutwin, Joey Scarr, and Sylvain Malacria. 2014. Supporting Novice to Expert Transitions in User Interfaces. *Comput. Surveys* 47, 2 (nov 2014), 1–36. https://doi.org/10.1145/2659796

[57] CogWorks Experiments. 2022. Space Fortress - John's First Game – YouTube. Retrieved September 21, 2022 from https://www.youtube.com/watch?v=tmRITTg70VE

[58] Daniel Cook. 2007. The Chemistry Of Game Design. https://www.gamasutra.com/view/feature/1524/the_chemistry_of_game_design.php

[59] Fergus I.M. Craik and Robert S. Lockhart. 1972. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior* 11, 6 (dec 1972), 671–684. https://doi.org/10.1016/s0022-5371(72)80001-x

[60] S. E. Criscimagna-Hemminger and R. Shadmehr. 2008. Consolidation Patterns of Human Motor Memory. *Journal of Neuroscience* 28, 39 (sep 2008), 9610–9618. https://doi.org/10.1523/jneurosci.3071-08.2008

[61] E. R. F. W. Crossman. 1959. A Theory of the Acquisition of Speed-Skill. *Ergonomics* 2, 2 (1959), 153–166. https://doi.org/10.1080/00140135908930419

[62] Cryptic Studios. 2013. Neverwinter. Game [PC].

[63] Mihaly Csikszentmihalyi. 2000. *Beyond boredom and anxiety*. Jossey-Bass.

[64] Marjorie Ann M. Cuerdo, Anika Mahajan, and Edward F. Melcer. 2021. Die-r Consequences: Player Experience and the Design of Failure through Respawning Mechanics. In *2021 IEEE Conference on Games (CoG)*. IEEE. https://doi.org/10.1109/cog52621.2021.9618894

[65] Rudolph P Darken and Barry Peterson. 2014. Spatial Orientation, Wayfinding, and Representation. In *Handbook of Virtual Environments*. CRC Press, 497–522. https://doi.org/10.1201/b17360-28

[66] Rudy P. Darken and John L. Sibert. 1993. A toolset for navigation in virtual environments. In *Proceedings of the 6th annual ACM symposium on User interface software and technology - UIST '93*. ACM Press. https://doi.org/10.1145/168642.168658

[67] Rudolph P. Darken and John L. Sibert. 1996. Navigating large virtual spaces. *International Journal of Human-Computer Interaction* 8, 1 (jan 1996), 49–71. https://doi.org/10.1080/10447319609526140

[68] Rudolph P. Darken and John L. Sibert. 1996. Wayfinding strategies and behaviors in large virtual worlds. In *Proceedings of the SIGCHI conference on Human factors in computing systems common ground - CHI '96*. ACM Press. https://doi.org/10.1145/238386.238459

[69] K Davids, C Button, and S Bennett. 2008. *Dynamics of Skill Acquisition: A Constraints-led Approach*. Human Kinetics. https://books.google.ca/books?id=0N-ffZNdEUMC

[70] Joey Davidson. 2011. The Agonizing Death of Video Game Instruction Booklets. *TechoBuffalo* (2011). Retrieved November 27, 2018 from https://www.technobuffalo.com/2011/11/16/the-agonizing-death-of-video-game-instruction-booklets/

[71] Lee C Deighton. 1971. *The encyclopedia of education*. Vol. 4. Macmillan New York.

[72] Peter F. Delaney, Peter P.J.L. Verkoeijen, and Arie Spirgel. 2010. Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. In *Psychology of Learning and Motivation*. Elsevier, 63–147. https://doi.org/10.1016/s0079-7421(10)53003-2

[73] Frank N. Dempster. 1988. The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist* 43, 8 (1988), 627–634. https://doi.org/10.1037/0003-066x.43.8.627

[74] Douglas Derryberry and Marjorie A. Reed. 2002. Anxiety-related attentional biases and their regulation by attentional control. *Journal of Abnormal Psychology* 111, 2 (2002), 225–236. https://doi.org/10.1037/0021-843x.111.2.225

[75] Sebastian Deterding. 2015. The lens of intrinsic skill atoms: A method for gameful design. *Human-Computer Interaction* 30, 3-4 (2015), 294–335. https://doi.org/10.1080/07370024.2014.993471

[76] DICE. 2008. Mirror's Edge. Game [Xbox 360, PS3].

[77] Betsy Van Dijk, Rieks Op Den Akker, Anton Nijholt, and Job Zwiers. 2003. Navigation Assistance in Virtual Worlds. *Informing Science: The International Journal of an Emerging Transdiscipline* 6 (2003), 115–125. https://doi.org/10.28945/519

[78] Kody R. Dillman, Terrance Tin Hoi Mok, Anthony Tang, Lora Oehlberg, and Alex Mitchell. 2018. A Visual Interaction Cue Framework from Video Game Environments for Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, Article 140, 12 pages. https://doi.org/10.1145/3173574.3173714

[79] Emanuel Donchin. 1989. The learning strategies project. *Acta Psychologica* 71, 1-3 (aug 1989), 1–15. https://doi.org/10.1016/0001-6918(89)90002-4

[80] Emanuel Donchin. 1995. Video games as research tools: The Space Fortress game. *Behavior Research Methods, Instruments, & Computers* 27, 2 (jun 1995), 217–223. https://doi.org/10.3758/bf03204735

[81] John J. Donovan and David J. Radosevich. 1999. A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology* 84, 5 (1999), 795–805. https://doi.org/10.1037/0021-9010.84.5.795 arXiv:00028282

[82] DoubleDutch Games. 2016. SpeedRunners. Game [PC].

[83] Julie S. Downs, Mandy B. Holbrook, Steve Sheng, and Lorrie Faith Cranor. 2010. Are Your Participants Gaming the System?: Screening Mechanical Turk Workers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. ACM, New York, NY, USA, 2399–2402. https://doi.org/10.1145/1753326.1753688

[84] Stephen Draper. 2017. Esports Terminology: Skill Ceilings and Skill Floors. *Esports Edition* (September 2017). Retrieved April 4, 2019 from https://esportsedition.com/general/skill-ceiling-skill-floor-esports-terminology/

[85] Hippolyte Dubois, Patrick Le Callet, and Antoine Coutrot. 2021. Visualizing navigation difficulties in video game experiences. In *2021 13th International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE. https://doi.org/10.1109/qomex51781.2021.9465423

[86] H. Ebbinghaus, H.A. Ruger, and C.E. Bussenius. 1913. *Memory: A Contribution to Experimental Psychology*. Teachers College, Columbia University.

[87] Darren Edge, Elly Searle, Kevin Chiu, Jing Zhao, and James A. Landay. 2011. MicroMandarin. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*. ACM Press. https://doi.org/10.1145/1978942.1979413

[88] Brian D. Ehret. 2002. Learning where to look. In *Proceedings of the SIGCHI conference on Human factors in computing systems Changing our world, changing ourselves - CHI '02*. ACM Press. https://doi.org/10.1145/503376.503414

[89] H C Ellis. 1978. *Fundamentals of Human Learning, Memory, and Cognition*. W. C. Brown Company.

[90] Stefan Engeser and Falko Rheinberg. 2008. Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion* 32, 3 (sep 2008), 158–172. https://doi.org/10.1007/s11031-008-9102-4

[91] Entertainment Software Association. 2020. Report: Video Games Contribute $90 Billion+ to U.S. Economy. https://www.theesa.com/news/report-video-games-contribute-90-billion-to-u-s-economy/

[92] Entertainment Software Association of Canada. 2021. The Canadian Video Game Industry 2021. https://theesa.ca/wp-content/uploads/2021/11/esac-2021-final-report.pdf

[93] Epic Games. 2017. Fortnite Battle Royale. Game [PC].

[94] K. Anders Ericsson, Ralf T. Krampe, and Clemens Tesch-Römer. 1993. The role of deliberate practice in the acquisition of expert performance. *Psychological Review* 100, 3 (1993), 363–406. https://doi.org/10.1037/0033-295X.100.3.363 arXiv:/doi.apa.org/psycinfo/1993-40718-001 [http:]

[95] K. Anders Ericsson and Robert Pool. 2016. *Peak: Secrets from the new science of expertise*. Houghton Mifflin Harcourt.

[96] K. Anders Ericsson, Michael J. Prietula, and Edward T. Cokely. 2007. The Making of an Expert. *Harvard Business Review* 85, 11 (2007), 146–147. https://doi.org/Article arXiv:arXiv:1011.1669v3

[97] W. K. Estes. 1955. Statistical theory of spontaneous recovery and regression. *Psychological Review* 62, 3 (1955), 145–154. https://doi.org/10.1037/h0048509

[98] M W Eysenck, M T Keane, and G Mather. 2010. *Cognitive Psychology: A Student's Handbook, 6th Edition*. Routledge Chapman & Hall.

[99] Monica Fabiani, Jean Buckley, Gabriele Gratton, Michael G H Coles, Emanuel Donchin, and Robert Logie. 1989. The training of complex task performance. *Acta Psychologica* 71, 1 (1989), 259–299. https://doi.org/10.1016/0001-6918(89)90012-7

[100] Franz Faul, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang. 2009. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods* 41, 4 (nov 2009), 1149–1160. https://doi.org/10.3758/brm.41.4.1149

[101] A. Field. 2018. *Discovering Statistics Using IBM SPSS Statistics*. SAGE Publications. https://books.google.ca/books?id=0QTyswEACAAJ

[102] Andy Field and Graham Hole. 2002. *How to design and report experiments*. Sage.

[103] Paul M Fitts. 1964. Perceptual-motor skill learning. In *Categories of human learning*. Elsevier, 243–285.

[104] Paul Morris Fitts and Michael I. Posner. 1967. *Human Performance*. Brooks/Cole.

[105] Fabio E. Fontana, Ovande Furtado, Oldemar Mazzardo, and Jere D. Gallagher. 2009. Whole and Part Practice: A Meta-Analysis. *Perceptual and Motor Skills* 109, 2 (oct 2009), 517–530. https://doi.org/10.2466/pms.109.2.517-530

[106] Fortnite. 2020. We defeated him! A record 15.3 million concurrent players joined forces in our biggest event ever to fight back galactus in today's in-game event, while more than 3.4 million cheered and watched on @youtubegaming and @Twitch! pic.twitter.com/iacnpcpkew. Retrieved January 28, 2022 from https://twitter.com/FortniteGame/status/1333954074371383296

[107] John R. Frederiksen and Barbara Y. White. 1989. An approach to training based upon principled task decomposition. *Acta Psychologica* 71, 1-3 (1989), 89–146. https://doi.org/10.1016/0001-6918(89)90006-1

[108] FromSoftware. 2011. Dark Souls. Game [PC].

[109] Aaron L. Gardony, Tad T. Brunyé, Caroline R. Mahoney, and Holly A. Taylor. 2013. How Navigational Aids Impair Spatial Memory: Evidence for Divided Attention. *Spatial Cognition & Computation* 13, 4 (sep 2013), 319–350. https://doi.org/10.1080/13875868.2013.792821

[110] Rosemary Garris, Robert Ahlers, and James E. Driskell. 2002. Games, Motivation, and Learning: A Research and Practice Model. *Simulation & Gaming* 33, 4 (dec 2002), 441–467. https://doi.org/10.1177/1046878102238607

[111] John Garvin, Christopher Reese, Darren Yager, and Ron Allen. 2013. Game aim assist. US Patent 8,342,926.

[112] G.M. Gazda and R.J. Corsini. 1980. *Theories of Learning: A Comparative Approach*. F. E. Peacock Publishers.

[113] G.M. Gazda and R.J. Corsini. 1980. *Theories of Learning: A Comparative Approach*. F. E. Peacock Publishers.

[114] James Paul Gee. 2005. Learning by Design: Good Video Games as Learning Machines. *E-Learning* 2, 1 (2005), 5–16. https://doi.org/10.2304/elea.2005.2.1.5

[115] Giant Bomb. 2019. Checkpoints (Concept). https://www.giantbomb.com/checkpoints/3015-292/. Accessed February 5, 2021.

[116] Diane L. Gill and Thomas E. Deeter. 1988. Development of the Sport Orientation Questionnaire. *Research Quarterly for Exercise and Sport* 59, 3 (sep 1988), 191–202. https://doi.org/10.1080/02701367.1988.10605504

[117] Janet Go, Rafael Ballagas, and Mirjana Spasojevic. 2012. Brothers and sisters at play. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work - CSCW '12*. ACM Press. https://doi.org/10.1145/2145204.2145316

[118] Daniel Gopher, Maya Weil, and Daniel Siegel. 1989. Practice under changing priorities: An approach to the training of complex skills. *Acta Psychologica* 71, 1 (1989), 147–177. https://doi.org/10.1016/0001-6918(89)90007-3

[119] Graham Grindlay. 2008. Haptic guidance benefits musical motor learning. *Symposium on Haptics Interfaces for Virtual Environment and Teleoperator Systems 2008 - Proceedings, Haptics* (2008), 397–404. https://doi.org/10.1109/HAPTICS.2008.4479984

[120] Thomas Grip. 2010. Why Trial and Error will Doom Games. https://www.gamasutra.com/blogs/ThomasGrip/20100412/87102/Why_Trial_and_Error_will_Doom_Games.php

[121] Frédéric Guay, Robert J Vallerand, and Céline Blanchard. 2000. On the assessment of situational intrinsic and extrinsic motivation: The Situational Motivation Scale (SIMS). *Motivation and emotion* 24, 3 (2000), 175–213. https://doi.org/10.1037/t12678-000

[122] Carl Gutwin, Rodrigo Vicencio-Moreira, and Regan L Mandryk. 2016. Does Helping Hurt?: Aiming Assistance and Skill Development in a First-Person Shooter Game. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*. ACM, 338–349. https://doi.org/10.1145/2967934.2968101

[123] Arin Hanson. 2011. Sequelitis - Mega Man Classic vs. Mega Man X. https://www.youtube.com/watch?v=8FpigqfcvlM

[124] Douglas A. Hardwick, Curtis W. McIntyre, and Herbert L. Pick. 1976. The Content and Manipulation of Cognitive Maps in Children and Adults. *Monographs of the Society for Research in Child Development* 41, 3 (1976), 1. https://doi.org/10.2307/1165952

[125] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*. Elsevier, 139–183. https://doi.org/10.1016/s0166-4115(08)62386-9

[126] Lynn Hasher and Rose T. Zacks. 1979. Automatic and effortful processes in memory. *Journal of Experimental Psychology: General* 108, 3 (1979), 356–388. https://doi.org/10.1037/0096-3445.108.3.356

[127] Thomas Heinen, Alexandra Pizzera, and Jorge Cottyn. 2009. When is manual guidance effective for the acquisition of complex skills in Gymnastics? *International Journal of Sport Psychology* 40 (2009), 1–22.

[128] Franklin M Henry and Donald E Rogers. 1960. Increased response latency for complicated movements and a "memory drum" theory of neuromotor reaction. *Research Quarterly. American Association for Health, Physical Education and Recreation* 31, 3 (1960), 448–458.

[129] William E Hick. 1952. On the rate of gain of information. *Quarterly Journal of Experimental Psychology* 4, 1 (1952), 11–26.

[130] Hidden Path Entertainment. 2012. Counter-Strike: Global Offensive. Game [PC].

[131] Daniel Himmelstein, Yitong Liu, and Jamie L. Shapiro. 2017. An Exploration of Mental Skills Among Competitive League of Legend Players. *International Journal of Gaming and Computer-Mediated Simulations* 9, 2 (apr 2017), 1–21. https://doi.org/10.4018/ijgcms.2017040101

[132] Douglas L Hintzman. 1974. *Theoretical Implications of the Spacing Effect*. Lawrence Erlbaum Associates, 77–99.

[133] Nicola J. Hodges, Raoul Huys, and Janet L. Starkes. 2012. Methodological Review and Evaluation of Research in Expert Performance in Sport. In *Handbook of Sport Psychology*. John Wiley & Sons, Inc., 159–183. https://doi.org/10.1002/9781118270011.ch7

[134] Dennis H. Holding. 1989. Skills Research. In *Human Skills*. John Wiley & Sons, 1–16.

[135] D. H. Holding and A. W. Macrae. 1964. Guidance, Restriction and Knowledge of Results. *Ergonomics* 7, 3 (1964), 289–295. https://doi.org/10.1080/00140136408930748

[136] J Honeybourne, M Hill, and H Moors. 2000. *AS Level Textbook*. Nelson Thornes Limited.

[137] Britton Horn, Seth Cooper, and Sebastian Deterding. 2017. Adapting Cognitive Task Analysis to Elicit the Skill Chain of a Game. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3116595.3116640

[138] Britton Horn, Seth Cooper, and Sebastian Deterding. 2017. Adapting Cognitive Task Analysis to Elicit the Skill Chain of a Game. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3116595.3116640

[139] James Thomas Howard. 2003. *Physical Guidance in Motor Learning*. Ph. D. Dissertation. Queensland University of Technology.

[140] HowLongToBeat. 2018. How long is Super Hexagon? (2018). Retrieved September 18, 2018 from https://howlongtobeat.com/game.php?id=9353

[141] Jeff Huang, Eddie Yan, Gifford Cheung, Nachiappan Nagappan, and Thomas Zimmermann. 2017. Master Maker: Understanding Gaming Skill Through Practice and Habit From Gameplay Behavior. *Topics in Cognitive Science* 9, 2 (feb 2017), 437–466. https://doi.org/10.1111/tops.12251

[142] Alycia M. Hund and Jennifer L. Minarik. 2006. Getting From Here to There: Spatial Anxiety, Wayfinding Strategies, Direction Type, and Wayfinding Efficiency. *Spatial Cognition &amp Computation* 6, 3 (sep 2006), 179–201. https://doi.org/10.1207/s15427633scc0603_1

[143] Ioanna Iacovides, Anna L. Cox, Ara Avakian, and Thomas Knoll. 2014. Player Strategies: Achieving Breakthroughs and Progressing in Single-Player and Cooperative Games. *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play - CHI PLAY '14* (2014), 131–140. https://doi.org/10.1145/2658537.2658697

[144] id Software. 2010. Quake Live. Game [PC].

[145] id Software. 2018. *GtkRadiant*. http://icculus.org/gtkradiant/

[146] Toru Ishikawa, Hiromichi Fujiwara, Osamu Imai, and Atsuyuki Okabe. 2008. Wayfinding with a GPS-based mobile navigation system: A comparison with maps and direct experience. *Journal of Environmental Psychology* 28, 1 (mar 2008), 74–82. https://doi.org/10.1016/j.jenvp.2007.09.002

[147] JC Fletcher. 2012. Super Hexagon has sold approximately 45,000 copies. *Engadget* (2012). Retrieved September 18, 2018 from https://www.engadget.com/2012/09/21/super-hexagon-has-sold-approximately-45-000-copies/

[148] Charlene Jennett, Anna L. Cox, Paul Cairns, Samira Dhoparee, Andrew Epps, Tim Tijs, and Alison Walton. 2008. Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies* 66, 9 (sep 2008), 641–661. https://doi.org/10.1016/j.ijhcs.2008.04.004

[149] Colby Johanson. 2016. *Scaffolding Novices to Leverage Auditory Awareness Cues in First-Person Shooters*. Master's thesis. University of Saskatchewan, Saskatoon, Canada.

[150] Colby Johanson. 2020. *colbyj/bride-of-frankensystem 1.1*. https://doi.org/10.5281/zenodo.3738761

[151] Colby Johanson, Carl Gutwin, Jason T. Bowey, and Regan L. Mandryk. 2019. Press Pause when you Play. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3311350.3347195

[152] Colby Johanson, Carl Gutwin, and Regan Mandryk. 2023. Trails, rails, and over-reliance: How navigation assistance affects route-finding and spatial learning in virtual environments. *International Journal of Human-Computer Studies* 178 (2023), 103097. https://doi.org/10.1016/j.ijhcs.2023.103097

[153] Colby Johanson, Carl Gutwin, and Regan L. Mandryk. 2017. The Effects of Navigation Assistance on Spatial Learning and Performance in a 3D Game. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116602

[154] Colby Johanson and Regan L Mandryk. 2016. Scaffolding Player Location Awareness through Audio Cues in First-Person Shooters. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 3450–3461. https://doi.org/10.1145/2858036.2858172

[155] A. Johnson. 2004. We Learn from our Mistakes – Don't We? *Ergonomics in Design: The Quarterly of Human Factors Applications* 12, 2 (2004), 24–27. https://doi.org/10.1177/106480460401200206

[156] Susanne Jul and George W. Furnas. 1997. Navigation in electronic worlds. *ACM SIGCHI Bulletin* 29, 4 (oct 1997), 44–49. https://doi.org/10.1145/270950.270979

[157] Jesper Juul. 2009. Fear of failing? the many meanings of difficulty in video games. *The video game theory reader* 2, 237-252 (2009).

[158] Jesper Juul. 2013. *The art of failure: An essay on the pain of playing video games*. Mit Press.

[159] Ruth Kanfer and Phillip Ackerman. 2000. Individual differences in work motivation: Further explorations of a trait framework. *Applied Psychology* 49, 3 (2000), 470–482.

[160] Tadhg Kelly. 2018. Conveyance. *What Games Are* (2018). Retrieved November 28, 2018 from https://www.whatgamesare.com/conveyance.html

[161] Numair Khan and Anis Ur Rahman. 2017. Rethinking the Mini-Map: A Navigational Aid to Support Spatial Learning in Urban Game Environments. *International Journal of Human–Computer Interaction* 34, 12 (dec 2017), 1135–1147. https://doi.org/10.1080/10447318.2017.1418804

[162] Kristian Kiili. 2005. Digital game-based learning: Towards an experiential gaming model. *The Internet and higher education* 8, 1 (2005), 13–24.

[163] Jong Wook Kim and Frank E. Ritter. 2015. Learning, Forgetting, and Relearning for Keystroke- and Mouse-Driven Tasks: Relearning Is Important. *Human–Computer Interaction* 30, 1 (2015), 1–33. https://doi.org/10.1080/07370024.2013.828564

[164] Jong W. Kim, Frank E. Ritter, and Richard J. Koubek. 2013. An integrated theory for improved skill acquisition and retention in the three stages of learning. *Theoretical Issues in Ergonomics Science* 14, 1 (2013), 22–37. https://doi.org/10.1080/1464536X.2011.573008

[165] Aniket Kittur, Ed H. Chi, and Bongwon Suh. 2008. Crowdsourcing user studies with Mechanical Turk. In *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*. ACM Press. https://doi.org/10.1145/1357054.1357127

[166] D.A. A Kolb. 1984. *Experiential Learning: Experience as the Source of Learning and Development*. Prentice Hall, Englewood Cliffs, NJ. https://books.google.se/books?id=jpbeBQAAQBAJ&pg=PR4&dq=10:0-13-389240-9&hl=en&sa=X&ved=0ahUKEwjH2sTY9vPRAhXBHJoKHYuUDIQQuwUIHTAA#v=onepage&q=10%3A0-13-389240-9&f=false

[167] Raph Koster. 2013. *Theory of fun for game design*. O'Reilly Media, Inc.

[168] Ben Kuchera. 2014. The passion of the checkpoint: Why gaming's most frustrating failure is so hard to fix. *Polygon* (2014). Retrieved February 5, 2021 from https://www.polygon.com/2014/2/25/5422328/the-passion-of-the-checkpoint-why-gamings-most-frustrating-failure-is

[169] Marjan Laal and Mozhgan Laal. 2012. Collaborative learning: what is it? *Procedia - Social and Behavioral Sciences* 31 (2012), 491–495. https://doi.org/10.1016/j.sbspro.2011.12.092

[170] Daniël Lakens. 2022. *Improving Your Statistical Inferences*. https://lakens.github.io/statistical_inferences/index.html

[171] LaunchBox Games Database. 2022. Pong Details. Retrieved September 21, 2022 from https://gamesdb.launchbox-app.com/games/images/7563

[172] Carol A Lawton and Janos Kallai. 2002. Gender differences in wayfinding strategies and anxiety about wayfinding: A cross-cultural comparison. *Sex roles* 47, 9 (2002), 389–401.

[173] HyunKyu Lee, Walter R. Boot, Chandramallika Basak, Michelle W. Voss, Ruchika Shaurya Prakash, Mark Neider, Kirk I. Erickson, Daniel J. Simons, Monica Fabiani, Gabriele Gratton, Kathy A. Low, and Arthur F. Kramer. 2012. Performance gains from directed training do not transfer to untrained tasks. *Acta Psychologica* 139, 1 (jan 2012), 146–158. https://doi.org/10.1016/j.actpsy.2011.11.003

[174] Timothy D. Lee and Elizabeth D. Genovese. 1988. Distribution of Practice in Motor Skill Acquisition: Learning and Performance Effects Reconsidered. *Research Quarterly for Exercise and Sport* 59, 4 (dec 1988), 277–287. https://doi.org/10.1080/02701367.1988.10609373

[175] Timothy D. Lee and Elizabeth D. Genovese. 1989. Distribution of Practice in Motor Skill Acquisition: Different Effects for Discrete and Continuous Tasks. *Research Quarterly for Exercise and Sport* 60, 1 (mar 1989), 59–65. https://doi.org/10.1080/02701367.1989.10607414

[176] Gilly Leshed, Theresa Velden, Oya Rieger, Blazej Kot, and Phoebe Sengers. 2008. In-Car GPS Navigation: Engagement with and Disengagement from the Environment. In *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*. ACM Press. https://doi.org/10.1145/1357054.1357316

[177] Jing Lin, Lijun Cao, and Nan Li. 2019. Assessing the influence of repeated exposures and mental stress on human wayfinding performance in indoor environments using virtual reality technology. *Advanced Engineering Informatics* 39 (2019), 53–61. https://doi.org/10.1016/j.aei.2018.11.007

[178] Jonas Linderoth. 2012. Why gamers don't learn more: An ecological approach to games as learning environments. *Journal of Gaming & Virtual Worlds* 4, 1 (mar 2012), 45–62. https://doi.org/10.1386/jgvw.4.1.45_1

[179] Lionhead Studios. 2008. Fable II. Game [Xbox 360].

[180] Ricardo Lopes and Rafael Bidarra. 2011. Adaptivity challenges in games and simulations: a survey. *IEEE Transactions on Computational Intelligence and AI in Games* 3, 2 (2011), 85–99. https://doi.org/10.1109/tciaig.2011.2152841

[181] I Scott MacKenzie. 2012. *Human-computer interaction: An empirical research perspective*. Newnes.

[182] A. W. Macrae and D. H. Holding. 1966. Transfer of training after guidance or practice. *Quarterly Journal of Experimental Psychology* 18, 4 (1966), 327–333. https://doi.org/10.1080/14640746608400050

[183] R.A. Magill. 1989. *Motor Learning: Concepts and Applications*. W.C. Brown.

[184] R.A. Magill. 1998. *Motor Learning: Concepts and Applications*. McGraw-Hill.

[185] R.A. Magill. 2007. *Motor Learning and Control: Concepts and Applications*. McGraw-Hill. https://books.google.ca/books?id=rCPdAAAACAAJ

[186] R. Magill. 2010. *Motor Learning and Control: Concepts and Applications*. McGraw-Hill Education.

[187] Amir Mané and Emanuel Donchin. 1989. The space fortress game. *Acta Psychologica* 71, 1-3 (aug 1989), 17–22. https://doi.org/10.1016/0001-6918(89)90003-6

[188] Daryl Marples. 2017. *The Influence of Intrinsic Perceptual Cues on Navigation and Route Selection in Virtual Environments*. Ph. D. Dissertation. University of Huddersfield.

[189] Theresa M. Marteau and Hilary Bekker. 1992. The development of a six-item short-form of the state scale of the Spielberger State—Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology* 31, 3 (sep 1992), 301–306. https://doi.org/10.1111/j.2044-8260.1992.tb00997.x

[190] Winter Mason and Siddharth Suri. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods* 44, 1 (2012), 1–23.

[191] RSW Masters, Jon Law, and Jon Maxwell. 2002. Implicit and explicit learning in interceptive actions. *Interceptive actions in sport: Information and movement* (2002), 126–143.

[192] J.P. Maxwell, R.S.W. Masters, E. Kerr, and E. Weedon. 2001. The implicit benefit of learning without errors. *The Quarterly Journal of Experimental Psychology A* 54, 4 (2001), 1049–1068. https://doi.org/10.1080/02724980143000073

[193] Edward McAuley, Terry Duncan, and Vance V. Tammen. 1989. Psychometric Properties of the Intrinsic Motivation Inventory in a Competitive Sport Setting: A Confirmatory Factor Analysis. *Research Quarterly for Exercise and Sport* 60, 1 (mar 1989), 48–58. https://doi.org/10.1080/02701367.1989.10607413

[194] Vivienne Robison McClatchy. 1925. The optimal position of a rest period in learning. *Journal of Experimental Psychology* 8, 4 (1925), 251–277. https://doi.org/10.1037/h0068331

[195] Hugh D. McCracken and George E. Stelmach. 1977. A Test of the Schema Theory of Discrete Motor Learning. *Journal of Motor Behavior* 9, 3 (sep 1977), 193–201. https://doi.org/10.1080/00222895.1977.10735109

[196] Penny McCullagh and Korinne N Meyer. 1997. Learning versus correct models: Influence of model type on the learning of a free-weight squat lift. *Research Quarterly for Exercise and Sport* 68, 1 (1997), 56–61.

[197] Chris McEntee. 2012. Rational Design The Core of Rayman Origins. http://www.gamasutra.com/view/feature/167214/rational{_}design{_}the{_}core{_}of{_}.php

[198] Adam W Meade and S Bartholomew Craig. 2012. Identifying careless responses in survey data. *Psychological methods* 17, 3 (2012), 437.

[199] Metacritic. 2018. Super Hexagon for PC Reivews. (2018). Retrieved September 18, 2018 from https://www.metacritic.com/game/pc/super-hexagon

[200] Spero A. Metalis. 1985. Effects of Massed versus Distributed Practice on Acquisition of Video Game Skill. *Perceptual and Motor Skills* 61, 2 (oct 1985), 457–458. https://doi.org/10.2466/pms.1985.61.2.457

[201] Everett Mettler, Christine M. Massey, and Philip J. Kellman. 2016. A comparison of adaptive and fixed schedules of practice. *Journal of Experimental Psychology: General* 145, 7 (jul 2016), 897–917. https://doi.org/10.1037/xge0000170

[202] Microsoft. 2001. Internet Backgammon. Game [PC].

[203] Josh Aaron Miller, Britton Horn, Matthew Guthrie, Jonathan Romano, Guy Geva, Celia David, Amy Robinson Sterling, and Seth Cooper. 2021. How do Players and Developers of Citizen Science Games Conceptualize Skill Chains? *Proceedings of the ACM on Human-Computer Interaction* 5, CHI PLAY (oct 2021), 1–29. https://doi.org/10.1145/3474671

[204] Matthew K. Miller and Regan L. Mandryk. 2016. Differentiating in-Game Frustration from at-Game Frustration using Touch Pressure. In *Proceedings of the 2016 ACM on Interactive Surfaces and Spaces - ISS '16*. ACM Press. https://doi.org/10.1145/2992154.2992185

[205] R. Moirn, P.-M. Léger, S. Senecal, M.-C.B. Roberge, M. Lefebvre, and M. Fredette. 2016. The effect of game tutorial: A comparison between casual and hardcore gamers. *CHI PLAY 2016 - Proceedings of the Annual Symposium on Computer-Human Interaction in Play Companion* (2016), 229–237. https://doi.org/10.1145/2968120.2987730

[206] Daniel R. Montello. 2005. Navigation. In *The Cambridge Handbook of Visuospatial Thinking*. Cambridge University Press, 257–294. https://doi.org/10.1017/cbo9780511610448.008

[207] Vanessa D Moss. 1995. *The efficacy of massed versus distributed practice as a function of desired learning outcomes and grade level of the student*. Ph. D. Dissertation. Utah State University. https://doi.org/10.26076/92f1-7957

[208] Dinara Moura and Magy Seif El-Nasr. 2014. Design Techniques for Planning Navigational Systems in 3-D Video Games. *Computers in Entertainment* 12, 2 (jun 2014), 1–25. https://doi.org/10.1145/2701657.2633421

[209] Stefan Münzer, Hubert D. Zimmer, and Jörg Baus. 2012. Navigation assistance: A trade-off between wayfinding support and configural learning support. *Journal of Experimental Psychology: Applied* 18, 1 (2012), 18–37. https://doi.org/10.1037/a0026553

[210] Daniel Natapov, Steven J Castellucci, and I Scott MacKenzie. 2009. ISO 9241-9 evaluation of video game controllers. In *Proceedings of Graphics Interface 2009*. Canadian Information Processing Society, 223–230.

[211] James C. Naylor and George E. Briggs. 1963. Effects of task complexity and task organization on the relative efficiency of part and whole training methods. *Journal of Experimental Psychology* 65, 3 (1963), 217–224. https://doi.org/10.1037/h0041060

[212] Nazca Corporation. 1996. Metal Slug. Game [Arcade].

[213] David M Neves and John R Anderson. 1981. Knowledge compilation: Mechanisms for the automatization of cognitive skills. *Cognitive skills and their acquisition* (1981), 57–84.

[214] Allen Newell and Paul S Rosenbloom. 1981. Mechanisms of skill acquisition and the law of practice. *Cognitive skills and their Acquisition* 6 (1981), 1–55.

[215] K M Newell. 1981. Skill Learning. In *Human Skills*. John Wiley & Sons, 203–226.

[216] K M Newell. 1981. Skill Learning. In *Human Skills*. 203–226.

[217] Newzoo. 2019. Global Esports Market Report 2019. (2019). Retrieved July 30, 2019 from https://newzoo.com/insights/trend-reports/newzoo-global-esports-market-report-2019-light-version/

[218] Nintendo. 2021. Wii - Health and Safety Precautions. *Nintendo* (2021). Retrieved July 12, 2021 from https://www.nintendo.com/consumer/systems/wii/en_na/health_safety.jsp

[219] Nintendo Creative Department. 1985. Super Mario Bros. Game [NES].

[220] Nintendo EAD. 1990. Super Mario World. Game [SNES].

[221] Nintendo EAD. 2006. Wii Sports. Game [Wii].

[222] Nintendo EPD. 2017. The Legend of Zelda: Breath of the Wild. Game [Switch].

[223] Nintendo Research & Development 1. 1981. Donkey Kong. Game [Arcade].

[224] Kent L. Norman. 2011. *Assessing the Components of Skill Necessary for Playing Video Games*. Technical Report. University of Maryland. 1–12 pages.

[225] Kent L. Norman, Yu-Chi Wang, Joseph Barnet, and Reze Mahmud. 2016. *Assessing the Necessary Skill Profiles for Playing Video Games*. Technical Report. University of Maryland. 1–16 pages.

[226] Thomas Pagulayan, Randy J and Keeker, Kevin and Wixon, Dennis and Romero, Ramon L and Fuller. 2002. User-centered Design in Games. In *Human-Computer Interaction: Designing for Diverse Users and Domains*. CRC Press Boca Raton, FL, Chapter 13, 217–236.

[227] Gabriele Paolacci and Jesse Chandler. 2014. Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science* 23, 3 (jun 2014), 184–188. https://doi.org/10.1177/0963721414531598

[228] Gabriele Paolacci, Jesse Chandler, Panagiotis G Ipeirotis, et al. 2010. Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making* 5, 5 (2010), 411–419.

[229] Felipe Pepe. 2016. The history of the Quest Compass & its dreadful convenience. *Gamasutra* (2016). Retrieved August 8, 2021 from http://www.gamasutra.com/blogs/FelipePepe/20160412/270100/The_history_of_the_Quest_Compass__its_dreadful_convenience.php

[230] Serge Petralito, Florian Brühlmann, Glena Iten, Elisa D Mekler, and Klaus Opwis. 2017. A Good Reason to Die: How Avatar Death and High Challenges Enable Positive Experiences. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2017)* (2017). https://doi.org/10.1145/3025453.3026047

[231] Brandon Piller, Colby Johanson, Cody Phillips, Carl Gutwin, and Regan L. Mandryk. 2020. Is a Change as Good as a Rest? Comparing Break Types for Spaced Practice in a Platformer Game. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3410404.3414225

[232] Barbara J Pollock and Timothy D Lee. 1992. Effects of the model's skill level on observational motor learning. *Research quarterly for exercise and sport* 63, 1 (1992), 25–29.

[233] Muhammad Farrel Pramono, Kevin Renalda, Dedy Prasetya Kristiadi, Harco Leslie Hendric Spits Warnars, and Worapan Kusakunniran. 2018. Matchmaking Problems in MOBA Games. *Indonesian Journal of Electrical Engineering and Computer Science* 11, 3 (sep 2018), 908. https://doi.org/10.11591/ijeecs.v11.i3.pp908-917

[234] Dirk C Prather. 1971. Trial-and-error versus errorless learning: Training, transfer, and stress. *The American journal of psychology* (1971), 377–386.

[235] R.W. Proctor, A. Dutta, and A. Ehrenstein. 1995. *Skill Acquisition and Human Performance*. SAGE Publications.

[236] Robert W. Proctor, T. Gilmour Reeve, and Daniel J. Weeks. 1990. A Triphasic Approach to the Acquisition of Response-Selection Skill. *Psychology of Learning and Motivation - Advances in Research and Theory* 26, C (1990), 207–240. https://doi.org/10.1016/S0079-7421(08)60055-9

[237] Andrew K Przybylski, C Scott Rigby, and Richard M Ryan. 2010. A motivational model of video game engagement. *Review of General Psychology* 14, 2 (2010), 154–166. https://doi.org/10.1037/a0019440

[238] Mary A. Pyc and Katherine A. Rawson. 2009. Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language* 60, 4 (may 2009), 437–447. https://doi.org/10.1016/j.jml.2009.01.004

[239] Relic Entertainment. 2021. Age of Empires IV. Game [PC].

[240] Falko Rheinberg, Regina Vollmeyer, and Bruce D Burns. 2001. QCM: A questionnaire to assess current motivation in learning situations. *Diagnostica* 47, 2 (2001), 57–66.

[241] Riot Games. 2009. League of Legends. Game [PC].

[242] Scott Rogers. 2014. *Level Up! The guide to great video game design*. John Wiley & Sons.

[243] Hassan Rohbanfard and Luc Proteau. 2011. Learning through observation: a combination of expert and novice models favors learning. *Experimental Brain Research* 215, 3-4 (2011), 183–197.

[244] Kelly Braun David C Rubin. 1998. The Spacing Effect Depends on an Encoding Deficit, Retrieval, and Time in Working Memory: Evidence. *Memory* 6, 1 (1 1998), 37–66. https://doi.org/10.1080/741941599

[245] Roy A Ruddle. 2001. Navigation: Am i really lost or virtually there? *Engineering psychology and cognitive ergonomics* 6 (2001), 135–142.

[246] Richard M. Ryan and Edward L. Deci. 2000. Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology* 25, 1 (jan 2000), 54–67. https://doi.org/10.1006/ceps.1999.1020

[247] Richard M. Ryan, C. Scott Rigby, and Andrew Przybylski. 2006. The Motivational Pull of Video Games: A Self-Determination Theory Approach. *Motivation and Emotion* 30, 4 (nov 2006), 344–360. https://doi.org/10.1007/s11031-006-9051-8

[248] Alan W. Salmoni. 1989. Motor Skill Learning. In *Human Skills*. John Wiley & Sons, 197–228.

[249] Alan W. Salmoni, Richard A. Schmidt, and Charles B. Walter. 1984. Knowledge of results and motor learning: A review and critical reappraisal. *Psychological Bulletin* 95, 3 (1984), 355–386. https://doi.org/10.1037/0033-2909.95.3.355

[250] Caroline Q Santos, Porto Alegre, Porto Alegre, Angelina C A Ziesemer, Porto Alegre, Milene S Silveira, Porto Alegre, Luciana S Espindola, Pedro A Pires, Porto Alegre, and Porto Alegre. 2015. "How can I help you?" Preliminary studies about user strategies and preferences during a game. August 2016 (2015).

[251] Florian Schimanke, Robert Mertens, and Oliver Vornberger. 2014. Spaced repetition learning games on mobile devices. *Interactive Technology and Smart Education* 11, 3 (sep 2014), 201–222. https://doi.org/10.1108/itse-07-2014-0017

[252] Florian Schimanke, Robert Mertens, and Oliver Vornberger. 2017. Designing for Motivation: Design-Considerations for Spaced-Repetition-Based Learning Games on Mobile Devices. *International Journal on E-Learning* 16, 3 (2017), 287–311.

[253] R.A. Schmidt and T.D. Lee. 2011. *Motor Control and Learning: A Behavioral Emphasis*. Human Kinetics.

[254] Richard A. Schmidt. 1975. A schema theory of discrete motor skill learning. *Psychological Review* 82, 4 (jul 1975), 225–260. https://doi.org/10.1037/h0076770

[255] Richard A. Schmidt. 1991. Frequent Augmented Feedback Can Degrade Learning: Evidence and Interpretations. In *Tutorials in Motor Neuroscience*. Springer Netherlands, 59–75. https://doi.org/10.1007/978-94-011-3626-6_6

[256] Richard A. Schmidt and Timothy D. Lee. 2005. *Motor Control and Learning: A Behavioral Emphasis*. Human Kinetics.

[257] Richard A Schmidt, Timothy D Lee, C J Winstein, G Wulf, and H N Zelaznik. 2018. *Motor Control and Learning, 6E*. Langara College.

[258] Richard A Schmidt, Douglas E Young, Stephan Swinnen, and Diane C Shapiro. 1989. Summary knowledge of results for skill acquisition: support for the guidance hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15, 2 (1989), 352.

[259] Bruce A. Schneider, Meital Avivi-Reich, and Mindaugas Mozuraitis. 2015. A cautionary note on the use of the Analysis of Covariance (ANCOVA) in classification designs with and without within-subject factors. *Frontiers in Psychology* 6 (apr 2015). https://doi.org/10.3389/fpsyg.2015.00474

[260] Walter Schneider. 1985. Training High-Performance Skills: Fallacies and Guidelines. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 27, 3 (jun 1985), 285–300. https://doi.org/10.1177/001872088502700305

[261] Walter Schneider and Richard M. Shiffrin. 1977. Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review* 84, 1 (1977), 1–66. https://doi.org/10.1037/0033-295x.84.1.1

[262] Wow Entertainment Sega. 1996. The House of the Dead. Game [Arcade].

[263] Yuri Seo. 2016. Professionalized consumption and identity transformations in the field of eSports. *Journal of Business Research* 69, 1 (jan 2016), 264–272. https://doi.org/10.1016/j.jbusres.2015.07.039

[264] C.H. Shea, W. Shebilske, and S. Worchel. 1993. *Motor Learning and Control*. Prentice Hall.

[265] Charles H. Shea and Robert M. Kohl. 1990. Specificity and Variability of Practice. *Research Quarterly for Exercise and Sport* 61, 2 (jun 1990), 169–177. https://doi.org/10.1080/02701367.1990.10608671

[266] Charles H. Shea and Robert M. Kohl. 1991. Composition of Practice: Influence on the Retention of Motor Skills. *Research Quarterly for Exercise and Sport* 62, 2 (jun 1991), 187–195. https://doi.org/10.1080/02701367.1991.10608709

[267] Charles H. Shea, Gabriele Wulf, Chad A. Whitacre, and Jin-Hoon Park. 2001. Surfing the implicit wave. *The Quarterly Journal of Experimental Psychology A* 54, 3 (aug 2001), 841–862. https://doi.org/10.1080/02724980042000381

[268] Wayne L Shebilske, Barry P Goettl, Kip Corrington, and Eric Anthony Day. 1999. Interlesson spacing and task-related processing during complex skill acquisition. *Journal of Experimental Psychology: Applied* 5, 4 (1999), 413–437. https://doi.org/10.1037/1076-898x.5.4.413

[269] Steve Sheng, Mandy Holbrook, Ponnurangam Kumaraguru, Lorrie Faith Cranor, and Julie Downs. 2010. Who Falls for Phish?: A Demographic Analysis of Phishing Susceptibility and Effectiveness of Interventions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. ACM, New York, NY, USA, 373–382. https://doi.org/10.1145/1753326.1753383

[270] Elizabeth L. Shoenfelt, Leslie A. Snyder, Allison E. Maue, C. Patrick McDowell, and Christopher D. Woolard. 2002. Comparison of Constant and Variable Practice Conditions on Free-Throw Shooting. *Perceptual and Motor Skills* 94, 3_suppl (jun 2002), 1113–1123. https://doi.org/10.2466/pms.2002.94.3c.1113

[271] H. A. Simon. 1956. Rational choice and the structure of the environment. *Psychological Review* 63, 2 (1956), 129–138. https://doi.org/10.1037/h0042769

[272] R N Singer. 1980. *Motor Learning and Human Performance: An Application to Motor Skills and Movement Behaviors*. MacMillan. https://books.google.ca/books?id=zD1kQgAACAAJ

[273] Robert N Singer and Leonide Gaines. 1975. Brief Note: Effects of Prompted and Problem-solving Approaches on Learning and Transfer of Motor Skills. *American Educational Research Journal* 12, 3 (1975), 395–403.

[274] Robert N Singer and Dale Pease. 1976. A comparison of discovery learning and guided instructional strategies on motor skill learning, retention, and transfer. *Research Quarterly. American Alliance for Health, Physical Education and Recreation* 47, 4 (1976), 788–796.

[275] Robert N Singer and Dale Pease. 1978. Effect of guided vs. discovery learning strategies on initial motor task learning, transfer, and retention. *Research Quarterly. American Alliance for Health, Physical Education and Recreation* 49, 2 (1978), 206–217.

[276] Timothy W. Smith. 1982. Olympic Decathlon. Game [PC].

[277] G.S. Snoddy. 1926. Learning and stability: a psychophysiological analysis of a case of motor learning with clinical applications. *Journal of Applied Psychology* 10, 1 (1926), 1–36. https://doi.org/10.1037/h0075814

[278] Nicholas C. Soderstrom and Robert A. Bjork. 2015. Learning Versus Performance. *Perspectives on Psychological Science* 10, 2 (2015), 176–199. https://doi.org/10.1177/1745691615569000

[279] William H Solley. 1952. The effects of verbal instruction of speed and accuracy upon the learning of a motor skill. *Research Quarterly. American Association for Health, Physical Education and Recreation* 23, 2 (1952), 231–240.

[280] Splash Damage. 2003. Wolfenstein: Enemy Territory. Game [PC].

[281] Tom Stafford, Sam Delvin, Rafet Sifa, and Anders Drachen. 2017. Exploration and Skill Acquisition in a Major Online Game. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*.

[282] Tom Stafford and Erwin Haasnoot. 2016. Testing Sleep Consolidation in Skill Learning: A Field Study Using an Online Game. *Topics in Cognitive Science* 9, 2 (nov 2016), 485–496. https://doi.org/10.1111/tops.12232

[283] Tom Stafford and Nemanja Vaci. 2022. Maximizing the Potential of Digital Games for Understanding Skill Acquisition. *Current Directions in Psychological Science* 31, 1 (jan 2022), 49–55. https://doi.org/10.1177/09637214211057841

[284] Steam Database. 2018. Super Hexagon - AppID: 221640. (2018). Retrieved September 18, 2018 from https://steamdb.info/app/221640/graphs/

[285] Dan Stemkoski. 2017. StarCraft Remastered - Artosis Talks about Deliberate Practice and Answering Questions. https://www.youtube.com/watch?v=y2nyWIlA_CY

[286] J Ridley Stroop. 1935. Studies of interference in serial verbal reactions. *Journal of experimental psychology* 18, 6 (1935), 643.

[287] StudioMDHR. 2017. Cuphead. Game [PC].

[288] Jeffery J Summers. 1989. Motor Programs. In *Human Skills*. John Wiley & Sons, 49–70.

[289] John Sweller. 1988. Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science* 12, 2 (apr 1988), 257–285. https://doi.org/10.1207/s15516709cog1202_4

[290] Holly A. Taylor, Susan J. Naylor, and Nicholas A. Chechile. 1999. Goal-specific influences on the representation of spatial perspective. *Memory & Cognition* 27, 2 (mar 1999), 309–319. https://doi.org/10.3758/bf03211414

[291] Team Meat. 2010. Super Meat Boy. Game [PC].

[292] Caitlin Tenison and John R. Anderson. 2017. The impact of practice frequency on learning and retention. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*. 1175–1180.

[293] The Fable Wiki. 2012. Glowing Trail. Retrieved April 12, 2017 from http://fable.wikia.com/wiki/Glowing_Trail

[294] The Jamovi project. 2022. jamovi (Version 3.2) [Computer Software]. Retrieved from https://www.jamovi.org.

[295] C Therrien. 2011. "To get help, please press X" the rise of the assistance paradigm in video game design. In *Proceedings of DiGRA 2011 Conference: Think Design Play*.

[296] L Thomsen, F Petersen, A Drachen, and P Babaei. 2016. Identifying Onboarding Heuristics for Free-to- Play Mobile Games : A Mixed Methods Approach Identifying Onboarding Heuristics for Free-to-Play. October (2016), 1–8.

[297] Line E. Thomsen, Falko Weigert Petersen, Anders Drachen, and Pejman Mirza-Babaei. 2016. Identifying Onboarding Heuristics for Free-to-Play Mobile Games: A Mixed Methods Approach. In *Entertainment Computing - ICEC 2016*. Springer International Publishing, 241–246. https://doi.org/10.1007/978-3-319-46100-7_24

[298] Edward Lee Thorndike. 1913. *The Original Nature of Man*. Teachers College, Columbia University.

[299] Perry W. Thorndyke and Sarah E. Goldin. 1983. Spatial Learning and Reasoning Skill. In *Spatial Orientation*. Springer US, 195–217. https://doi.org/10.1007/978-1-4615-9325-6_9

[300] Perry W Thorndyke and Barbara Hayes-Roth. 1982. Differences in spatial knowledge acquired from maps and navigation. *Cognitive Psychology* 14, 4 (oct 1982), 560–589. https://doi.org/10.1016/0010-0285(82)90019-6

[301] Perry W. Thorndyke and Cathleen Stasz. 1980. Individual differences in procedures for knowledge acquisition from maps. *Cognitive Psychology* 12, 1 (jan 1980), 137–175. https://doi.org/10.1016/0010-0285(80)90006-7

[302] P D Tomporowski. 2003. *The Psychology of Skill: A Life-span Approach*. Praeger Publishers. https://books.google.ca/books?id=QaN9AAAAMAAJ

[303] Thomas C. Toppino and Emilie Gerbier. 2014. About Practice. In *Psychology of Learning and Motivation*. Elsevier, 113–189. https://doi.org/10.1016/b978-0-12-800090-8.00004-4

[304] Melih Turgut. 2014. Development of the spatial ability self-report scale (SASRS): reliability and validity studies. *Quality & Quantity* 49, 5 (aug 2014), 1997–2014. https://doi.org/10.1007/s11135-014-0086-8

[305] Turn 10 Studios. 2011. Forza Motorsport 4. Game [Xbox 360].

[306] TV Tropes. 2018. Attract Mode. Retrieved November 27, 2018 from https://tvtropes.org/pmwiki/pmwiki.php/Main/AttractMode

[307] TV Tropes. 2018. Hint System. Retrieved November 27, 2018 from https://tvtropes.org/pmwiki/pmwiki.php/Main/HintSystem

[308] TV Tropes. 2018. Speedrun. Retrieved February 25, 2022 from https://tvtropes.org/pmwiki/pmwiki.php/Main/Speedrun

[309] TV Tropes. 2021. Auto-Save - TV Tropes. Retrieved July 26, 2021 from https://tvtropes.org/pmwiki/pmwiki.php/Main/Autosave

[310] u/0li0li. 2021. r/ArenaFPS - What is wrong with us? *reddit* (2021). Retrieved January 28, 2022 from https://www.reddit.com/r/ArenaFPS/comments/l93yse/what_is_wrong_with_us/

[311] u/CherrySlurpee. 2016. r/gaming - CMV: Quest markers are one of the worst things to ever happen to video games, and they're ruining the concept of immersion. *reddit* (2016). Retrieved October 8, 2021 from https://www.reddit.com/r/changemyview/comments/4nnt79/cmv_quest_markers_are_one_of_the_worst_things_to/

[312] u/Cosmic_Colin. 2014. r/truegaming - Is using save states really cheating? *reddit* (2014). Retrieved February 5, 2021 from https://www.reddit.com/r/truegaming/comments/24miq7/is_using_save_states_really_cheating/

[313] u/imaunitard. 2012. r/gaming - Why not take a break? *reddit* (2012). Retrieved July 31, 2019 from https://www.reddit.com/r/gaming/comments/x6exy/why_not_take_a_break/

[314] u/Kizuxtheo. 2018. r/Megaman - Thoughts on save states? *reddit* (2018). Retrieved August 31, 2020 from https://www.reddit.com/r/Megaman/comments/7ij4so/thoughts_on_save_states/

[315] u/Lopsidation. 2017. r/darksouls - Why are there long walks from bonfires to bosses? *reddit* (2017). Retrieved September 13, 2018 from https://www.reddit.com/r/darksouls/comments/6f8dwm/why_are_there_long_walks_from_bonfires_to_bosses/

[316] u/MetroidAndZeldaFan. 2015. r/truegaming - Do you find the lack of checkpoints in Dark Souls to be fair. *reddit* (2015). Retrieved September 15, 2018 from https://www.reddit.com/r/truegaming/comments/3k1bhp/do_you_find_the_lack_of_checkpoints_in_dark_souls/

[317] Unity Technologies. 2018. *Unity*. http://unity.com

[318] University of Washington. 2008. Foldit. Game [PC].

[319] Valve Corporation. 2007. Portal. Game [PC].

[320] Marieke van Asselen, Eva Fritschy, and Albert Postma. 2005. The influence of intentional and incidental learning on acquiring spatial knowledge during navigation. *Psychological Research Psychologische Forschung* 70, 2 (apr 2005), 151–156. https://doi.org/10.1007/s00426-004-0199-0

[321] F Martijn Verhoeven and Karl M Newell. 2018. Unifying practice schedules in the timescales of motor learning and performance. *Human Movement Science* 59 (2018), 153–169. https://doi.org/10.1016/j.humov.2018.04.004

[322] Rodrigo Vicencio-Moreira, Regan L Mandryk, and Carl Gutwin. 2015. Now You Can Compete With Anyone: Balancing Players of Different Skill Levels in a First-Person Shooter Game. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2255–2264. https://doi.org/10.1145/2702123.2702242

[323] Rodrigo Vicencio-Moreira, Regan L Mandryk, Carl Gutwin, and Scott Bateman. 2014. The Effectiveness (or Lack Thereof) of Aim-Assist Techniques in First-Person Shooter Games. In *CHI'14: Proceedings of the 2014 CHI international conference on Human factors in computing systems*. Toronto, Canada, 937–946.

[324] Regina Vollmeyer and Falko Rheinberg. 2006. Motivational Effects on Self-Regulated Learning with Different Tasks. *Educational Psychology Review* 18, 3 (oct 2006), 239–253. https://doi.org/10.1007/s10648-006-9017-0

[325] L.S. Vygotsky. 1978. Interaction between learning and development. , 79–91 pages. https://doi.org/10.1016/S0006-3495(96)79572-3

[326] Michael G Wagner. 2006. On the Scientific Relevance of eSports.. In *International conference on internet computing*. 437–442.

[327] Szymon Walkowiak, Tom Foulsham, and Alison F. Eardley. 2015. Individual differences and personality correlates of navigational performance in the virtual route learning task. *Computers in Human Behavior* 45 (apr 2015), 402–410. https://doi.org/10.1016/j.chb.2014.12.041

[328] David Waller, Earl Hunt, and David Knapp. 1998. The Transfer of Spatial Knowledge in Virtual Environment Training. *Presence: Teleoperators and Virtual Environments* 7, 2 (apr 1998), 129–143. https://doi.org/10.1162/105474698565631

[329] Waluigilicious. 2016. Save states on old games. Let's end this argument for good. *NeoGAF* (2016). Retrieved August 31, 2020 from https://www.neogaf.com/threads/save-states-on-old-games-lets-end-this-argument-for-good.1209168/

[330] R. H. Waters. 1930. The Influence of Large Amounts of Manual Guidance upon Human Maze Learning. *The Journal of General Psychology* 4, 1-4 (1930), 213–228. https://doi.org/10.1080/00221309.1930.9918311

[331] Helen Wauck and Wai-Tat Fu. 2017. A Data-Driven, Multidimensional Approach to Hint Design in Video Games. *Iui '17* (2017), 137–147. https://doi.org/10.1145/3025171.3025224

[332] Jane Webster, Linda Klebe Trevino, and Lisa Ryan. 1993. The dimensionality and correlates of flow in human-computer interactions. *Computers in Human Behavior* 9, 4 (dec 1993), 411–426. https://doi.org/10.1016/0747-5632(93)90032-n

[333] Alan T Welford. 1976. *Skilled performance: Perceptual and motor skills.* Scott & Foresman.

[334] Wikipedia contributors. 2018. Bushnell's Law — Wikipedia, The Free Encyclopedia. Retrieved October 2, 2018 from https://en.wikipedia.org/w/index.php?title=Bushnell%27s_Law&oldid=825057314

[335] Wikipedia contributors. 2018. Nonlinear gameplay — Wikipedia, The Free Encyclopedia. Retrieved November 26, 2018 from https://en.wikipedia.org/w/index.php?title=Nonlinear_gameplay&oldid=865540573

[336] Wikipedia contributors. 2019. Video Game Masters Tournament — Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Video_Game_Masters_Tournament&oldid=895291688 [Online; accessed 21-May-2019].

[337] Wikipedia contributors. 2022. File:Foldit screenshot 2.jpg — Wikimedia Commons. Retrieved September 21, 2022 from https://commons.wikimedia.org/wiki/File:Foldit_screenshot_2.jpg

[338] Wikipedia contributors. 2022. File:Long tail.svg — Wikimedia Commons. Retrieved September 5, 2022 from https://commons.wikimedia.org/wiki/File:Long_tail.svg

[339] Daniel B. Willingham and Kelly Goedert-Eschmann. 1999. The Relation Between Implicit and Explicit Learning: Evidence for Parallel Development. *Psychological Science* 10, 6 (1999), 531–534. https://doi.org/10.1111/1467-9280.00201

[340] A. Wingfield. 1979. *Human learning and memory: an introduction*. Harper & Row.

[341] A Wingfield. 1979. *Human learning and memory: an introduction*. Harper & Row. https://books.google.ca/books?id=dEMQAQAAIAAJ

[342] Brian M. Winn. 2009. The Design, Play, and Experience Framework. In *Handbook of Research on Effective Electronic Gaming in Education*. IGI Global, 1010–1024. https://doi.org/10.4018/9781599048086.ch058

[343] Carolee J. Winstein, Patricia S. Pohl, and Rebecca Lewthwaite. 1994. Effects of Physical Guidance and Knowledge of Results on Motor Learning: Support for the Guidance Hypothesis. *Research Quarterly for Exercise and Sport* 65, 4 (1994), 316–323. https://doi.org/10.1080/02701367.1994.10607635

[344] Carolee J. Winstein and Richard A. Schmidt. 1989. Sensorimotor Feedback. In *Human Skills*. John Wiley & Sons, 17–47.

[345] Melody Wiseheart, Annalise A. D'Souza, and Jacey Chae. 2017. Lack of spacing effects during piano learning. *PLoS ONE* 12, 8 (2017), 1–13. https://doi.org/10.1371/journal.pone.0182986

[346] Bob G. Witmer and Michael J. Singer. 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments* 7, 3 (jun 1998), 225–240. https://doi.org/10.1162/105474698565686

[347] Robert Sessions Woodworth and Harold Schlosberg. 1954. *Experimental Psychology*. Oxford and IBH Publishing.

[348] WoWWiki. 2021. Quest. Retrieved August 8, 2021 from http://wowwiki.wikia.com/wiki/Quest

[349] Gabriele Wulf, Nancy H. McNevin, Thomas Fuchs, Florian Ritter, and Tonya Toole. 2000. Attentional Focus in Complex Skill Learning. *Research Quarterly for Exercise and Sport* 71, 3 (sep 2000), 229–239. https://doi.org/10.1080/02701367.2000.10608903

[350] Gabriele Wulf and Charles H. Shea. 2002. Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychonomic Bulletin & Review* 9, 2 (jun 2002), 185–211. https://doi.org/10.3758/bf03196276

[351] Gabriele Wulf, Charles H. Shea, and Chad A. Whitacre. 1998. Physical-Guidance Benefits in Learning a Complex Motor Skill. *Journal of Motor Behavior* 30, 4 (1998), 367–380. https://doi.org/10.1080/00222899809601351

[352] Gabriele Wulf and Cornelia Weigelt. 1997. Instructions about Physical Principles in Learning a Complex Motor Skill: To Tell or Not to Tell.... *Research Quarterly for Exercise and Sport* 68, 4 (dec 1997), 362–367. https://doi.org/10.1080/02701367.1997.10608018

[353] Su Xue, Meng Wu, John Kolen, Navid Aghdaie, and Kazi A Zaman. 2017. Dynamic Difficulty Adjustment for Maximized Engagement in Digital Games. In *Proceedings of the 26th International Conference on World Wide Web Companion - WWW '17 Companion*. {ACM} Press. https://doi.org/10.1145/3041021.3054170

[354] Tiffany Zachry, Gabriele Wulf, John Mercer, and Neil Bezodis. 2005. Increased movement accuracy and reduced EMG activity as the result of adopting an external focus of attention. *Brain Research Bulletin* 67, 4 (oct 2005), 304–309. https://doi.org/10.1016/j.brainresbull.2005.06.035

[355] Krzysztof Zagata and Beata Medyńska-Gulij. 2023. Mini-Map Design Features as a Navigation Aid in the Virtual Geographical Space Based on Video Games. *ISPRS International Journal of Geo-Information* 12, 2 (feb 2023), 58. https://doi.org/10.3390/ijgi12020058

[356] Zelda Dungeon. 2018. Breath of the Wild Walkthrough – Great Plateau. Retrieved November 26, 2018 from https://www.zeldadungeon.net/breath-of-the-wild-walkthrough/great-plateau/

[357] Mohammad Zohaib. 2018. Dynamic Difficulty Adjustment (DDA) in Computer Games: A Review. *Advances in Human-Computer Interaction* 2018 (nov 2018), 1–12. https://doi.org/10.1155/2018/5681652

Appendix A

Consent Forms and Ethical Approval

A.1 Consent Forms

A.1.1 Manuscript 1

Title: Personalizing, Adapting, and Balancing Computer Games

Researcher(s):
Colby Johanson, Ph.D. Student, Department of Computer Science, University of Saskatchewan, 306-966-2327, colby.johanson@usask.ca
Dr. Regan Mandryk, Associate Professor, Department of Computer Science, University of Saskatchewan, 306-966-4888, regan@usask.ca

Purpose(s) and Objective(s) of the Research: The purpose of this study is to understand how players navigate 3D virtual environments in the context of games.

Procedures:

- In this study, you will be asked to complete questionnaires asking questions about yourself and your experience level. Next, you will navigate two levels from first-person shooter games. Following this, you will be asked to complete additional questionnaires relating to your experience and complete tasks relating to navigating a level.
- This study will take approximately 40 minutes to complete.

Funded by: The Natural Sciences and Engineering Research Council of Canada (NSERC).

Potential Risks and Benefits: There are no known or anticipated risks to you by participating in this research. Your participation will help us better understand how player navigation compares to navigation in real life.

Confidentiality:

- Confidentiality will be maintained throughout the study. The entire process and data will be anonymized. Data will only be presented in the aggregate and any individual user comments will be anonymized prior to presentation in academic venues.
- Only the principal researcher and his research assistants will have access to the data to ensure that your confidentiality is protected.
- Storage of Data
  - Data (including survey and interview responses, logs of computer use, and videos of interaction) will be stored on a secure password-protected server for 7 years after data collection.
  - After 7 years, the data will be destroyed. Paper data will be shredded and digital data will be wiped from hard disks beyond any possibility for data recovery.

Right to Withdraw:

- Your participation is voluntary. You may withdraw from the research project for any reason, at any time without explanation.
- Should you wish to withdraw, you may do so at any point, and we will not use your data; we will destroy all records of your data.

- Your right to withdraw data from the study will apply until the data have been aggregated (one week after study completion). After this date, it is possible that some form of research dissemination will have already occurred and it may not be possible to withdraw your data

Follow up: To obtain results from the study, please contact Colby Johanson (colby.johanson@usask.ca).

Questions or Concerns:

- Contact the researcher(s) using the information at the top.
- This research project has been approved on ethical grounds by the University of Saskatchewan Research Ethics Board. Any questions regarding your rights as a participant may be addressed to that committee through the Research Ethics Office ethics.office@usask.ca (306) 966-2975. Out of town participants may call toll free (888) 966-2975.

A.1.2 Manuscript 2

Title: Personalizing, Adapting, and Balancing Computer Games

Researcher(s):
Colby Johanson, Ph.D. Student, Department of Computer Science, University of Saskatchewan, 306-966-2327, colby.johanson@usask.ca
Dr. Regan Mandryk, Associate Professor, Department of Computer Science, University of Saskatchewan, 306-966-4888, regan@usask.ca
Dr. Carl Gutwin, Professor, Department of Computer Science, University of Saskatchewan, 306-966-8646, gutwin@usask.ca

Purpose(s) and Objective(s) of the Research: The purpose of this study is to understand how players navigate 3D virtual environments in the context of games with the presence of varying assistance techniques.

Procedures:

- You will train yourself to navigate an environment.
- You will answer questionnaires regarding the task and your experience in virtual environments .

Funded by: The Natural Sciences and Engineering Research Council of Canada (NSERC).

Potential Risks and Benefits: There are no known or anticipated risks to you by participating in this research. Your participation will help us better understand how guidance impacts learning in virtual environments.

Confidentiality:

- Confidentiality will be maintained throughout the study. The entire process and data will be anonymized. Data will only be presented in the aggregate and any individual user comments will be anonymized prior to presentation in academic venues.
- Only the principal researcher and his research assistants will have access to the data to ensure that your confidentiality is protected.
- Storage of Data
  - Data (including survey and interview responses, logs of computer use, and videos of interaction) will be stored on a secure password-protected server for 7 years after data collection.
  - After 7 years, the data will be destroyed. Paper data will be shredded and digital data will be wiped from hard disks beyond any possibility for data recovery.

Right to Withdraw:

- Your participation is voluntary. You may withdraw from the research project for any reason, at any time without explanation.
- Should you wish to withdraw, you may do so at any point, and we will not use your data; we will destroy all records of your data.
- Your right to withdraw data from the study will apply until the data have been aggregated (one week after study completion). After this date, it is possible that some form of research dissemination will have already occurred and it may not be possible to withdraw your data

Follow up: To obtain results from the study, please contact Colby Johanson (colby.johanson@usask.ca).

Questions or Concerns:

- Contact the researcher(s) using the information at the top.
- This research project has been approved on ethical grounds by the University of Saskatchewan Research Ethics Board. Any questions regarding your rights as a participant may be addressed to that committee through the Research Ethics Office ethics.office@usask.ca (306) 966-2975. Out of town participants may call toll free (888) 966-2975.

A.1.3 Manuscript 3

Title: The Effects of Spaced Practice on Experience and Performance in a Digital Game

Researcher(s):

- Colby Johanson, Ph.D. Student, Department of Computer Science, University of Saskatchewan, 306-966-2327, colby.johanson@usask.ca
- Dr. Regan Mandryk, Professor, Department of Computer Science, University of Saskatchewan, 306-966-4888, regan@usask.ca
- Dr. Carl Gutwin, Professor, Department of Computer Science, University of Saskatchewan, 306-966-8646, gutwin@usask.ca

Purpose(s) and Objective(s) of the Research: The purpose of this study is to understand how spaced practice sessions impact player experience performance in a digital game.

Procedures:

- You will complete questionnaires asking questions about yourself, and your experience with games (about 10 minutes).
- You will play a digital game for 20 minutes.
- You will then complete questionnaires relating to your experience playing the game.

Funded by: The Natural Sciences and Engineering Research Council of Canada (NSERC).

Potential Risks and Benefits: There are no known or anticipated risks to you by participating in this research. Your participation will help us better understand how spacing out practice sessions impacts play experience and learning.

Confidentiality:

- Confidentiality will be maintained throughout the study. The entire process and data will be anonymized. Data will only be presented in the aggregate and any individual user comments will be anonymized prior to presentation in academic venues.
- Only the principal researcher and their research assistants will have access to the data to ensure that your confidentiality is protected.
- Storage of Data
  - Data (including questionnaire responses and logs of computer use) will be stored on a secure password-protected server for 5 years after data collection.
  - After 5 years, the data will be destroyed. Digital data will be wiped from hard disks beyond any possibility for data recovery.

Right to Withdraw:

- Your participation is voluntary. You may withdraw from the research project for any reason, at any time without explanation.
- Should you wish to withdraw, you may do so at any point, and we will not use your data; we will destroy all records of your data.
- Withdrawal requests can be made by contacting us through the Mechanical Turk website.
- Your right to withdraw data from the study applies until September 1, 2018. After this date, it is possible that some form of research dissemination will have already occurred and it may not be possible to withdraw your data.

Follow up: To obtain results from the study, please contact Colby Johanson (colby.johanson@usask.ca).

Questions or Concerns:

- Any questions you many have regarding consent can be sent to us by contacting us through the Mechanical Turk website or by sending an email to any of the contact emails listed in this consent form.
- This research project has been approved on ethical grounds by the University of Saskatchewan Research Ethics Board. Any questions regarding your rights as a participant may be addressed to that committee through the Research Ethics Office `ethics.office@usask.ca` (306) 966-2975. Out of town participants may call toll free (888) 966-2975.

Copies:

- If you would like to keep a copy of this consent form for your records, simply right-click this web page, click "Save Page As..." and follow the prompts provided by your web browser.

By clicking the consent button below, you are indicating that you...

- Have read and understand the description provided.
- Have had an opportunity to ask questions and your questions have been answered.
- Consent to participate in the research project.
- Understand that copy of this Consent Form is available to you for your records.

A.1.4 Manuscript 4

Title: The Effects of Spaced Practice on Experience and Performance in a Digital Game

Researcher(s):

- Colby Johanson, Ph.D. Student, Department of Computer Science, University of Saskatchewan, 306-966-2327, colby.johanson@usask.ca
- Brandon Piller, Masters Student, Department of Computer Science, University of Saskatchewan, 306-966-2327 brandon.piller@usask.ca
- Dr. Regan Mandryk, Professor, Department of Computer Science, University of Saskatchewan, 306-966-4888, regan@usask.ca
- Dr. Carl Gutwin, Professor, Department of Computer Science, University of Saskatchewan, 306-966-8646, gutwin@usask.ca

Purpose(s) and Objective(s) of the Research: The purpose of this study is to understand how spaced practice sessions impact player experience performance in a digital game.

Procedures:

- You will complete questionnaires asking questions about yourself, and your experience with games (about 10 minutes).
- You will play a digital game for 20 minutes.
- You will then complete questionnaires relating to your experience playing the game.

Funded by: The Natural Sciences and Engineering Research Council of Canada (NSERC).

Potential Risks and Benefits: There are no known or anticipated risks to you by participating in this research. Your participation will help us better understand how spacing out practice sessions impacts play experience and learning.

Confidentiality:

- Confidentiality will be maintained throughout the study. The entire process and data will be anonymized. Data will only be presented in the aggregate and any individual user comments will be anonymized prior to presentation in academic venues.
- Only the principal researcher and their research assistants will have access to the data to ensure that your confidentiality is protected.
- Storage of Data
  - Data (including questionnaire responses and logs of computer use) will be stored on a secure password-protected server for 5 years after data collection.
  - After 5 years, the data will be destroyed. Digital data will be wiped from hard disks beyond any possibility for data recovery.

Right to Withdraw:

- Your participation is voluntary. You may withdraw from the research project for any reason, at any time without explanation.
- Should you wish to withdraw, you may do so at any point, and we will not use your data; we will destroy all records of your data.
- Withdrawal requests can be made by contacting us through the Mechanical Turk website.
- Your right to withdraw data from the study applies until August 15, 2020. After this date, it is possible that some form of research dissemination will have already occurred and it may not be possible to withdraw your data.

Follow up: To obtain results from the study, please contact Colby Johanson (colby.johanson@usask.ca).

Questions or Concerns:

- Any questions you many have regarding consent can be sent to us by contacting us through the Mechanical Turk website or by sending an email to any of the contact emails listed in this consent form.
- This research project has been approved on ethical grounds by the University of Saskatchewan Research Ethics Board. Any questions regarding your rights as a participant may be addressed to that committee through the Research Ethics Office ethics.office@usask.ca (306) 966-2975. Out of town participants may call toll free (888) 966-2975.

Copies:

- If you would like to keep a copy of this consent form for your records, simply right-click this web page, click "Save Page As..." and follow the prompts provided by your web browser.

By clicking the consent button below, you are indicating that you...

- Have read and understand the description provided.
- Have had an opportunity to ask questions and your questions have been answered.
- Consent to participate in the research project.
- Understand that copy of this Consent Form is available to you for your records.

A.2 Certificates of Approval

UNIVERSITY OF SASKATCHEWAN

Behavioural Research Ethics Board (Beh-REB) 14-Oct-2021

Certificate of Re-Approval

Ethics Number: 14-225

Principal Investigator: Regan Mandryk

Department: Department of Computer Science

Locations Where Research Activities are Conducted: University of Saskatchewan, Canada

Student(s): Ansgar Depping
Colby Johanson
Jason Bowey
Matthew Miller
Max Birk
Ricardo Rheeder
Sarah Vedress

Funder(s): Canada Foundation for Innovation
College of Arts and Science
College of Graduate and Postdoctoral Studies
Mitacs
Natural Sciences and Engineering Research Council of Canada
Networks of Centres of Excellence

Sponsor:

Title: Personalizing, Adapting, and Balancing Computer Games

Approval Effective Date: 13-Oct-2021

Expiry Date: 13-Oct-2022

Acknowledgment Of: N/A

Review Type: Delegated Review

* This study, inclusive of all previously approved documents, has been re-approved until the expiry date noted above

CERTIFICATION

The University of Saskatchewan Behavioural Research Ethics Board (Beh-REB) is constituted and operates in accordance with the current version of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2 2014). The University of Saskatchewan Behavioural Research Ethics Board has reviewed the above-named project. The proposal was found to be acceptable on ethical grounds. The principal investigator has the responsibility for any other administrative or regulatory approvals that may pertain to this project, and for ensuring that the authorized project is carried out according to the conditions outlined in the original protocol submitted for ethics review. This Certificate of Approval is valid for the above time period provided there is no change in experimental protocol or consent process or documents.

ONGOING REVIEW REQUIREMENTS

In order to receive annual renewal, a status report must be submitted to the REB Chair for Board consideration within one month prior to the current expiry date each year the project remains open, and upon project completion. Please refer to the following website for further instructions: https://vpresearch.usask.ca/researchers/forms.php.

Digitally Approved by Diane Martz
Chair, Behavioural Research Ethics Board
University of Saskatchewan

1 / 1

UNIVERSITY OF SASKATCHEWAN

Behavioural Research Ethics Board (Beh-REB) 09-Sep-2021

Certificate of Re-Approval

Application ID:	181
Principal Investigator:	Carl Gutwin
Department:	Department of Computer Science
Locations Where Research Activities are Conducted:	United States, United States of America
Student(s):	Brandon Piller Colby Johanson
Funder(s):	Natural Sciences and Engineering Research Council of Canada
Sponsor:	
Title:	The Effects of Spaced Practice on Experience and Performance in a Digital Game
Approval Effective Date:	20-Aug-2021
Expiry Date:	20-Aug-2022
Acknowledgment Of:	N/A
Review Type:	Delegated Review

* This study, inclusive of all previously approved documents, has been re-approved until the expiry date noted above

CERTIFICATION

The University of Saskatchewan Behavioural Research Ethics Board (Beh-REB) is constituted and operates in accordance with the current version of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2 2014). The University of Saskatchewan Behavioural Research Ethics Board has reviewed the above-named project. The proposal was found to be acceptable on ethical grounds. The principal investigator has the responsibility for any other administrative or regulatory approvals that may pertain to this project, and for ensuring that the authorized project is carried out according to the conditions outlined in the original protocol submitted for ethics review. This Certificate of Approval is valid for the above time period provided there is no change in experimental protocol or consent process or documents.

ONGOING REVIEW REQUIREMENTS

In order to receive annual renewal, a status report must be submitted to the REB Chair for Board consideration within one month prior to the current expiry date each year the project remains open, and upon project completion. Please refer to the following website for further instructions: https://vpresearch.usask.ca/researchers/forms.php.

Digitally Approved by Melanie Bayly
Vice-Chair, Behavioural Research Ethics Board
University of Saskatchewan

1 / 1

Appendix B

Preregistrations

For Manuscripts C and D, I pre-registered the design of the experiments on the OSF Registries. For convenience, I have reproduced these pre-registrations in this appendix.

Manuscript C:

- Colby Johanson, Carl Gutwin, and Regan Mandryk. 2018. Spaced Practice in Video Games. `https://doi.org/10.17605/OSF.IO/SK2W9`

Manuscript D:

- Colby Johanson and Brandon Piller. 2020. PlayPause Checkpoints. `https://doi.org/10.17605/OSF.IO/S6ZDT`

B.1 OSF Preregistration for Manuscript C

B.1.1 Title

The Effects of Spaced Practice in a Digital Game

B.1.2 Research Questions

This project consists of two studies.

First Study

The benefits of distributed practice (practice with rest intervals) over massed practice (continuous practice without rest) have been found for dozens of different tasks, from verbal recall tasks, to tasks which make use of procedural skills. These benefits appear to be essentially universal. However, video games may be different because they are played for intrinsic enjoyment. Therefore, we seek to determine whether or not this effect is present when considering performance in video games. (1)

Additionally, because short-term performance is not necessarily reflective of long term learning, we will have participants complete a delayed retention test, where participants will play another session of the game after a one day rest period. The second research question is therefore: Do the short-term differences in performance persist in the long-term? (2)

Since we expect that this effect will be found, our followup question is, what practice schedule is optimal for video games? (3) We will test a variety of rest periods to see which one results in the best performance improvements over a set time period, including: No rest, 1-minute rest, 2-minute rest, 5-minute rest, 10-minute rest, and 1-day rest.

After answering these questions, we will more closely examine the role that individual differences have in improving performance, and if there are any interactions between these individual differences and performance and subjective experience. (4)

Second Study

Past studies examining the effect of massed versus distributed practice have held the total training time (or trials) consistent between the groups. What might happen if we look at the problem from the perspective of someone having limited access to a computer or game console; is it better to take breaks, or play continuously? (5) After determining the optimal rest period, we will recruit more participants to practice the game continuously for the total practice + rest time and compare between those two groups.

B.1.3 Hypotheses

1. Based on the apparent ubiquity of the distribution of practice effect, we expect that we will find that it also is present when considering video game practice. This will be evident by examining performance metrics, and we expect to observe that distributed practice will outperform massed practice.
2. Past studies have found that the distribution of practice effect affects not only short-term performance, but also long-term learning. We expect to find the same.
3. Based on a meta-analysis by Donovan and Radosevich (1999), we expect that the optimal inter-session rest period will be somewhere around 1 to 10 minutes.
4. This is exploratory, and we are not sure what to expect.
5. We expect that the distribution of practice effect will be strong, but if the inter-session rest period is long enough, then continuous practice including practicing throughout what would be a rest period would result in better performance than the group taking rests. Without knowing the optimal inter-sesson rest period at this time we cannot make a prediction.

B.1.4 Sampling Plan

Existing Data

Registration prior to creation of data

Explanation of existing data

N/A

Data collection procedures

We will recruit participants through Amazon's Mechanical Turk platform. Participants will be paid $10 USD per hour of their time. We will recruit from the general Mechanical Turk population in the United States, with the default Mechanical Turk filters (minimum 500 accepted HITs and 90% approval rate).

We only want to recruit participants who have never played Super Hexagon before, therefore we will run a pre-screen task on Mechanical Turk that asks whether or not a prospective participant has played Super Hexagogn previously. Based on the results of this pre-screen task, they will receive a qualification that allows them to complete the actual task.

No files selected

Sample size

For Study 1, we will collect 216 participants, or 36 participants per group (6 groups).

For Study 2, we will collect an additional 36 participants and compare that to one of the groups from Study 1.

Sample size rationale

We used the program, G*Power, to conduct a power analysis for a one-way ANOVA. We specified a medium effect size of 0.25, with 6 groups, and a power of 0.80. The standard alpha error probability of 0.05 was used. This gave a suggested sample size of 216 total.

Stopping rule

N/A

B.1.5 Variables

Manipulated variables

Study 1: The length of the inter-session rest period. Study 2: Total combined length of the play sessions.

No files selected

Measured variables

In-Game Performance

High Score. The game will be played in multiple sessions, with each session consisting of a variable number of rounds, depending on the player's performance. Because each round ends when the player hits an obstacle, we will measure the participant's overall high score (longest time alive) in each session.

Average Life Time. For each session, the participant will play multiple rounds. Looking at the sessions individually, we will take the average of the scores (time) among each session's rounds.

Subjective Measures

Intrinsic Motivation. We will evaluate the participants intrinsic motivation towards the game by using the Intrinsic Motivation Inventory (IMI) (McAuley, E., Duncan, T., & Tammen, V. 1989). The IMI measures the participant's interest-enjoyment, effort-importance, competence, and tension-pressure.

Flow. We will use Jackson et al.'s (2008) Flow Short Scale (FSS) to measure the participant's experience of flow while playing the game. We will use the fluency of performance and absorption by activity subscales.

Immersion. We will use Jannett et al.'s (2008) questionnaire on in-game immersion. The questionnaire measures the participant's attention, temporal dissociation, transportation, challenge, emotional involvement, and enjoyment

Individual Trait Differences and Demographic Measures

Gaming Experience. We expect that individuals who enjoy playing video games regularly might enjoy our game more than those who don't, and so may perform better. We therefore will ask participants a number of questions to evaluate their level of experience with video games, including:

- "Are you a gamer?"
- "Are you experienced at playing video games?"
- "Do you ever become so involved in a video game that it is as if you are inside the game rather than moving a joystick or pressing buttons and watching the screen?"
- "How many years have you been playing video games for?"
- "How often (on average) do you play video games?"
- "If you have played games more often in the past, how often were you playing at peak times?"

Attentional Control. We anticipate that participants who are better at maintaining attention and focus on the task might perform better than those who don't. We therefore will use Derryberry and Reed's (2002) Attentional Control Scale (ACS) to measure each participant's attentional control.

Immersive Tendencies. We used Whitmer and Singer's (1998) Immersive Tendencies Questionnaire (ITQ) to measure participants' tendency to experience presence in virtual environments. The questionnaire consists of three subscales: involvement (propensity to get involved with an activity), focus (ability to concentrate on enjoyable activities), and games (how much they play games and whether they become involved enough to feel like they are inside the game).

Current Motivation. Because each participant may have different levels of interest in completing our task, we used Guay et al.'s (2000) Situational Motivation Scale (SIMS) to measure the participant's intrinsic motivation, identified regulation, external regulation, and amotivation towards our experiment. We also used Rheinberg et al.'s (2001) Questionnaire on Current Motivation (QCM) to measure the participant's task-related anxiety, probability of success, interest, and challenge.

Achievement Orientation. Because participants with a competitive nature may invest more effort into the task, we measured their competitiveness, win orientation, and goal orientation using Gill and Deeter's (1988) Sport Orientation Questionnaire.

References

1. Falko Rheinberg, Regina Vollmeyer, and Bruce D Burns. 2001. QCM: A questionnaire to assess current motivation in learning situations. Diagnostica 47, 2: 57–66.
2. Charlene Jennett, Anna L. Cox, Paul Cairns, Samira Dhoparee, Andrew Epps, Tim Tijs, and Alison Walton. 2008. Measuring and defining the experience of immersion in games. International Journal of Human Computer Studies 66, 9: 641–661. https://doi.org/10.1016/j.ijhcs.2008.04.004
3. D Derryberry and M Reed. 2002. Anxiety Related Attentional Biases and Their Regulation by Attentional Control. Journal of Abnormal Psychology 111, 2: 225–236. https://doi.org/10.1037//0021-843X.111.2.225
4. Bob G. Witmer and Michael J. Singer. 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. Presence: Teleoperators and Virtual Environments 7, 3: 225–240. https://doi.org/10.1162/105474698565686
5. Susan A. Jackson, Andrew J. Martin, and Robert C. Eklund. 2008. Long and Short Measures of Flow: The Construct Validity of the FSS-2, DFS-2, and New Brief Counterparts. Journal of Sport and Exercise Psychology 30, 5: 561–587. https://doi.org/10.1123/jsep.30.5.561
6. Diane L. Gill and Thomas E. Deeter. 1988. Development of the sport orientation questionnaire. Research Quarterly for Exercise and Sport 59, 3: 191–202. https://doi.org/10.1080/02701367.1988.10605504
7. Frederic Guay, Robert J. Vallerand, and Celine Blanchard. 2000. On the Assessment of Situational Intrinsic and Extrinsic Motivation: The Situational Motivation Scale (SIMS). Motivation and Emotion 24, 3: 175–213. https://doi.org/10.1023/A:1005614228250

- questionnaires.pdf

Indices

In the previous question, we indicated 7 existing (published) questionnaires. These questionnaires have various subscales, and some of the items may be reverse coded. After re-coding the reversed questions to match the coding of the others, the average value of the questions in each subscale will be computed, and those values will be used.

In particular, the following questionnaires and the subscales we are using are as follows:

Outcomes

Intrinsic Motivation Inventory

- Interest and enjoyment
- Competence
- Effort
- Pressure

Flow Short Scale

- Absorption
- Fluency
- Importance

Immersion

- Attention
- Temporal dissociation
- Transportation
- Challenge
- Emotional involvement
- Enjoyment

Trait Differences

Gaming Experience

Attentional Control Scale

- No subscales – take overall measure

Immersive Tendencies

- Focus
- Involvement

Situational Motivation Scale

- Intrinsic motivation towards the task
- Identified regulation
- External regulation
- Amotivation

Questionnaire on Current Motivation

- Probability of success
- Task-related anxiety
- Challenge

Sport Orientation Questionnaire

- Competitiveness
- Win orientation
- Goal orientation

No files selected

B.1.6 Design Plan

Study type

Experiment - A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.

Blinding

For studies that involve human subjects, they will not know the treatment group to which they have been assigned.

Study design

We will use a between-subjects design.

For Study 1, the between-subject factor will be the length of the inter-session rest period. Instead of using our performance measures as a repeated measure, we will consider the final performance measures and use the initial performance as a covariate. When considering the performance retention, we will consider the retention session and again use the initial performance as a covariate. The measures of subjective experience (Intrinsic Motivation, Flow, and Immersion) will not be repeated. They will be measured once after the last play session.

For Study 2, the between-subject factor will be the total practice time. Everything else will be the same.

The game we are using for both studies is a clone of the commercial game, Super Hexagon.

No files selected

Randomization

N/A

B.1.7 Analysis Plan

Statistical models

Hypothesis 1 will be tested with a confirmatory analysis. We will use two one-way ANCOVAs — one each for the dependent measures of final session high score and final session average life time. The treatment group (length of inter-session rest period) will be the between-subject variable and the initial, first-session, performance will be used as a covariate (as in Van Breukelen 2006).

Hypothesis 2 will be tested similarly to Hypothesis 1. We will use two one-way ANCOVAs — one each for the dependent measures of retention session high score and retention session average life time. The treatment group (length of inter-session rest period) will be the between-subject variable and the initial, first-session, performance will be used as a covariate.

Hypothesis 3 will also be tested with a confirmatory analysis. We will perform post-hoc pairwise analyses of the ANOVAs of Hypothesis 1 and 2. The pairwise comparisons will reveal which of the inter-session rest periods we used is optimal for performance and learning.

Hypothesis 4 is exploratory. The analysis to be used is reported below, under "Exploratory analysis."

Hypothesis 5 will be tested with a confirmatory analysis. We will consider both performance measures (high score and average life time) in terms of the difference between the first session and last session to represent the performance improvement. We will use two one way ANOVAs, one for each performance measure, with the treatment group (total play time) as the between-subject variable.

References:

- Van Breukelen, G. J. (2006). ANCOVA versus change from baseline had more power in randomized studies and more bias in nonrandomized studies. Journal of clinical epidemiology, 59(9), 920-925. DOI: `https://doi.org/10.1016/j.jclinepi.2006.02.007`

No files selected

Transformations

Some of the individual differences questionnaires have reverse-coded questions.

Follow-up analyses

These were mentioned in the answer to "Statistical models." Hypothesis 3 will be tested by performing post-hoc pairwise analyses.

Inference criteria

p less than .05.

Data exclusion

We will only exclude data points from participants who failed to attempt to play the game, those who stopped playing the game part way through the experiment, or those who had significant performance issues (framerate less than 300 frames a second).

Missing data

Missing data will result in the participant being removed from the analysis.

Exploratory analysis

Hypothesis 4 is exploratory. We will examine the correlations between our measures of individual differences and our outcome variables (both performance and experiential measures). We will choose measures of individual differences to use as covariates in separate ANCOVAs for each of our outcome variables. Each ANCOVA will use the treatment group as the between-subject measure. This will allow us to examine whether or not individual differences impact performance, learning, and the subjective experience. Additionally, we can determine whether or not there are any interaction effects between our covariates and the treatment group to determine how individual differences impact the distribution of practice effect.

B.1.8 Scripts

Upload an analysis script with clear comments

- PreregSyntax.sps

B.1.9 Other

Other

No response

B.2 OSF Preregistration for Manuscript D

B.2.1 Study Information

Hypotheses

If checkpoints act as part-task practice, then they will improve performance during “training" but hinder performance on retention and transfer tasks.

If breaks after death act as spaced practice, then they will improve performance during training, as well as on retention and transfer tasks.

Participants will dislike taking forced breaks upon death and starting the level over from the beginning.

B.2.2 Design Plan

Study type

Experiment - A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.

Blinding

For studies that involve human subjects, they will not know the treatment group to which they have been assigned.

Personnel who interact directly with the study subjects (either human or non-human subjects) will not be aware of the assigned treatments. (Commonly known as “double blind”)

Is there any additional blinding in this study?

The study will be administered over the internet, so participants will be assigned to treatment groups based on which treatment group has the fewest participants at the time the participant start. Personnel cannot interfere with this process and under normal circumstances, will not interact with participants directly (only in the case of technical errors).

Study design

The study will use a between subjects design with 2 factors with 2 levels each (explicit break or no explicit break, and part or whole practice).

No files selected

Randomization

No response

B.2.3 Sampling Plan

Existing Data

Registration prior to creation of data

Explanation of existing data

No response

Data collection procedures

Participants will be recruited through Amazon's Mechanical Turk. Participants will be pre-screened to ensure a basic level of competence in playing games similar to our game, so that they are capable of completing the levels in our game.

Participants will be paid $10 USD per hour for agreeing to complete our task. Participants that fully complete our task will be invited back at a later time (via an email sent with Mechanical Turk's API) to complete retention and transfer tasks, and they will again be compensated $10 USD per hour for agreeing to complete these follow-up tasks.

All participants will be at least 18 years of age with active worker accounts on Amazon Mechanical Turk with at least a 97% approval rating and more than 500 HITs completed.

The game will be developed in Unity and deployed online via a WebGL build. The game and all questionnaires will be presented to participants via a website built using the Bride of Frankensystem framework (`https://github.com/colbyj/bride-of-frankensystem/`).

No files selected

Sample size

Based on an a priori power analysis, we may recruit up to 179 participants. Given that participants may not complete the task or may not return when recruited for the follow-up tasks, we may end up recruiting more participants to compensate for this.

Sample size rationale

An a prior power analysis in G*Power used the following parameters: .25 effect size, alpha of .05, power of .80, numerator df of 3, 4 groups, and 6 covariates.

Stopping rule

No response

B.2.4 Variables

Manipulated variables

The study uses a 2x2 design with two factors: Checkpoints and Explicit Breaks.

Checkpoints:
If participants play the version of the game with checkpoints, then throughout the level will be flags that participants will touch that save their progress. If they die while traversing the level, they will be re-started at the checkpoint. If participants are playing the version of the game without checkpoints, then when they die they will re-start from the beginning of the current level.

Presence of explicit breaks after death:
If participants play the version of the game that has an explicit break, then any time they die while traversing the level they will have to wait some number of seconds before they can attempt the level again. This length of time should be

similar to the length of time players need to traverse the level to reach the point where they died. This time will be determined by pilot testing; we will have participants play the game without any breaks or checkpoints and calculate the average length of time between each death. This will inform our selection of how long the break should be.

No files selected

Measured variables

Measured Predictors:

- Gaming and Platforming Expertise, consisting of:
  - How much do you self-identify as a gamer on the following scale? (0 to 100)
  - How familiar are you with side-scrolling platform games? (0 to 100)
  - How familiar are you with Super Mario games? (0 to 100)
  - How familiar are you with the game “Super Meat Boy"? (0 to 100)
  - How familiar are you with the game “Speedrunners"? (0 to 100)
  - How familiar are you with the game “Celeste"? (0 to 100)
- Derryberry and Reed’s (2002) Attentional Control Scale (ACS)
- Rheinberg et al.’s (2001) Questionnaire on Current Motivation (QCM)
- Gill and Deeter’s (1988) Sport Orientation Questionnaire (SOQ)
- Completion time of the first four levels.

Objective measures:

- Total levels completed
- Total deaths

Objective measures will be collected three times: the initial training session, the retention session, and the transfer session.

Subjective measures:

- Vollmeyer and Rheinberg’s (2006) Flow Scale Short (FSS)
- Abeele et al.’s (2020) Player Experience Inventory (PXI)
- A specific question depending on their group:
  - "How did you feel about waiting to play the game after each death?"
  - "How did you feel about the checkpoints in the game?"
  - "How did you feel about starting each level from the beginning after each death?"
  - "Do you think you would have benefited from taking a break while you played?"
  - "Do you have any further comments about the game?"

See uploaded pdf.

PXI and FSS will be collected twice. Once after the initial training session and again following retention/transfer session.

- questionnaires.pdf

Indices

- Gaming and Platforming Expertise will be the mean of the questions.
- Completion time of the first four levels will be the sum of the number of seconds
- Total levels completed (a count)
- Total deaths (a count)

The published questionnaires have their own subscales.

No files selected

B.2.5 Analysis Plan

Statistical models

To ensure there are no trait differences between the groups, we will use a one-way ANOVA for each measure (Gaming and Platforming Experience, Attentional Control, Current Motivation, Sport Orientation, Early level completion time). If there are significant differences between groups for one of the measures, then that measure will be used as a covariate in all other analyses.

Separate ANCOVAs will be used for each outcome measure (total levels completed, total deaths, and each of the subscales for FKS and PXI). Presence of Explicit Break and Checkpoints will be used as between subjects factors. The covariate selection will be made based on a correlation matrix and whether each covariate correlates with the measure.

No files selected

Transformations

All covariates will be mean centred.

Inference criteria

$p < .05$

Data exclusion

We will exclude people:

- by age (<18, or >90).
- those who did not complete any levels beyond the tutorial
- if they are bots (determine by inspecting whether written responses are appropriate)
- if their framerate within the game is low (<30)

Missing data

If the participant does not complete the experiment, they will not be included in the analysis.

Exploratory analysis

Most subjective measures will be used in an exploratory way, as we do not have any specific predictions.

B.2.6 Other

Other

No response

Appendix C

Measures of Individual Differences

C.1 Attentional Control Scale [74]

Read each item carefully. Using the provided scale, please indicate your agreement with each item.

	Almost never	Sometimes	Often	Always
When I am working hard on something, I still get distracted by events around me.	○	○	○	○
It is difficult for me to coordinate my attention between the listening and writing required when taking notes during lectures.	○	○	○	○
When concentrating, I can focus my attention so that I become unaware of what's going on in the room around me.	○	○	○	○
After being interrupted or distracted, I can easily shift my attention back to what I was doing before.	○	○	○	○
When I need to concentrate and solve a problem, I have trouble focusing my attention	○	○	○	○
I can become interested in a new topic very quickly when I need to.	○	○	○	○
I can quickly switch from one task to another.	○	○	○	○
When trying to focus my attention on something, I have difficulty blocking out distracting thoughts.	○	○	○	○
It is easy for me to read or write while I'm also talking on the phone.	○	○	○	○
I have a hard time coming up with new ideas quickly.	○	○	○	○
When I am reading or studying, I am easily distracted if there are people talking in the same room.	○	○	○	○
When a distracting thought comes to mind, it is easy for me to shift my attention away from it.	○	○	○	○
It's very hard for me to concentrate on a difficult task when there are noises around.	○	○	○	○
It takes me a while to get really involved in a new task	○	○	○	○
When concentrating I ignore feelings of hunger or thirst.	○	○	○	○
It is easy for me to alternate between two different tasks.	○	○	○	○
It is hard for me to break from one way of thinking about something and look at it from another point of view.	○	○	○	○

	Almost never	Sometimes	Often	Always
I have trouble carrying on two conversations at once.	○	○	○	○
I have a hard time concentrating when I'm excited about something.	○	○	○	○
My concentration is going even if there is music in the room around me.	○	○	○	○

C.2 Questionnaire for Current Motivation [324]

Read each item carefully. Using the provided scale, please indicate your agreement with each item.

	Disagree						Agree
I'm afraid I will make a fool out of myself.	○	○	○	○	○	○	○
I probably won't manage to do this task.	○	○	○	○	○	○	○
I think I won't do well at the task.	○	○	○	○	○	○	○
I feel petrified by the demands of this task.	○	○	○	○	○	○	○
I'm really going to try as hard as I can on this task.	○	○	○	○	○	○	○
When I think about the task, I feel somewhat concerned.	○	○	○	○	○	○	○
I think I am up to the difficulty of this task.	○	○	○	○	○	○	○
This task is a real challenge for me.	○	○	○	○	○	○	○
It would be embarrassing to fail at this task.	○	○	○	○	○	○	○
I am eager to see how I will perform in the task.	○	○	○	○	○	○	○
I think everyone could do well on this task.	○	○	○	○	○	○	○
I feel under pressure to do this task well.	○	○	○	○	○	○	○
If I can do this task, I will feel proud of myself.	○	○	○	○	○	○	○

C.3 Situational Motivation Scale [121]

Why are you currently engaged in this activity?

	Corresponds not at all						Corresponds exactly
Because I am supposed to do it	○	○	○	○	○	○	○
Because this activity is fun	○	○	○	○	○	○	○
Because I am doing it for my own good	○	○	○	○	○	○	○
Because I think that this activity is interesting	○	○	○	○	○	○	○
I do this activity, but I am not sure it is a good thing to pursue it	○	○	○	○	○	○	○
By personal decision	○	○	○	○	○	○	○
Because I don't have any choice	○	○	○	○	○	○	○
There may be good reasons to do this activity, but personally I don't see any	○	○	○	○	○	○	○
I don't know; I don't see what this activity brings me	○	○	○	○	○	○	○

	Corresponds not at all						Corresponds exactly
Because I think that this activity is good for me	○	○	○	○	○	○	○
Because I feel that I have to do it	○	○	○	○	○	○	○
Because I think that this activity is pleasant	○	○	○	○	○	○	○
Because I feel good when doing this activity	○	○	○	○	○	○	○
Because it is something that I have to do	○	○	○	○	○	○	○
Because I believe that this activity is important for me	○	○	○	○	○	○	○
I do this activity but I am not sure if it is worth it	○	○	○	○	○	○	○

C.4 Immersive Tendencies Questionnaire [346]

	Never			Occasionally			Often
Do you ever become so involved in a television program or book that people have problems getting your attention?	○	○	○	○	○	○	○
Do you ever become so involved in a movie that you are not aware of things happening around you?	○	○	○	○	○	○	○
How often do you play arcade or video games? (OFTEN should be taken to mean every day or every two days, on average.)	○	○	○	○	○	○	○
Do you ever have dreams that are so real that you feel disoriented when you awake?	○	○	○	○	○	○	○
Do you easily become deeply involved in movies or TV dramas?	○	○	○	○	○	○	○
How frequently do you find yourself closely identifying with the characters in a story line?	○	○	○	○	○	○	○
Have you ever remained apprehensive or fearful long after watching a scary movie?	○	○	○	○	○	○	○
Do you ever become so involved in a daydream that you are not aware of things happening around you?	○	○	○	○	○	○	○
When playing sports, do you become so involved in the game that you lose track of time?	○	○	○	○	○	○	○
Do you ever become so involved in a video game that it is as if you are inside the game rather than moving a joystick and watching the screen?	○	○	○	○	○	○	○
Have you ever gotten scared by something happening on a TV show or in a movie?	○	○	○	○	○	○	○
Have you ever gotten excited during a chase or fight scene on TV or in the movies?	○	○	○	○	○	○	○
Do you ever become so involved in doing something that you lose all track of time?	○	○	○	○	○	○	○

	Not			Moderately			Very
How good are you at blocking out external distractions when you are involved in something?	○	○	○	○	○	○	○
How mentally alert do you feel at the present time?	○	○	○	○	○	○	○
How physically fit do you feel today?	○	○	○	○	○	○	○

C.5 Spatial Ability Self-Report Scale [304]

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I immediately notice when one of my friends makes even a small change in his or her physical appearance.	○	○	○	○	○
I can easily identify a three-dimensional shape drawn on paper.	○	○	○	○	○
I can fold and create new shapes from a square piece of paper in my mind.	○	○	○	○	○
I can disintegrate a three-dimensional object made of cubes in my mind.	○	○	○	○	○
I can compose additional drawings while solving a geometry problem.	○	○	○	○	○
I immediately forget the faces of people I have met.	○	○	○	○	○
I never get lost in streets that I have travelled before.	○	○	○	○	○
I can visualize the rotation of a figure drawn on paper around a certain point in my mind.	○	○	○	○	○
I can form a three-dimensional object in my mind when it is presented to me in a flattened form.	○	○	○	○	○
I can imagine the rotated versions of three-dimensional objects mentally.	○	○	○	○	○
I can find a shorter route to a location if I have been there before.	○	○	○	○	○
I have difficulty remembering the faces that I see in photographs.	○	○	○	○	○
I can solve a geometry problem without making a drawing.	○	○	○	○	○
By visualizing a three-dimensional object in my mind, I can cut and fold closed its cardboard model.	○	○	○	○	○
Looking at the street map of a region, I can find where I want to go.	○	○	○	○	○
If I see photographs of a building taken from different perspectives, I can visualize this three-dimensional structure in my mind.	○	○	○	○	○
When I want to create the shape of a fruit, vegetable or pie, I can easily imagine how I should cut.	○	○	○	○	○
I can visualize the shortest route on the streets that I travel.	○	○	○	○	○

C.6 International Wayfinding Anxiety Scale [172]

In the past the following scenarios caused me to feel...

	Not at all anxious				Very anxious
Leaving a store that I have been to for the first time and deciding which way to turn to get to a destination	○	○	○	○	○
Finding my way in an unfamiliar shopping mall, medical center, or large building complex	○	○	○	○	○
Pointing in the direction of a place outside that someone wants to get to and has asked for directions, when I am in a windowless room	○	○	○	○	○
Deciding which direction to walk in an unfamiliar city or town after coming out of a train/bus/metro station or parking garage	○	○	○	○	○
Finding my way to an appointment in an unfamiliar area of a city or town	○	○	○	○	○
Finding my way out of a complex arrangement of offices that I have visited for the first time	○	○	○	○	○
Finding my way back to a familiar area after realizing I have made a wrong turn and become lost while traveling	○	○	○	○	○
Trying a new route that I think will be a shortcut, without a map	○	○	○	○	○

C.7 International Wayfinding Strategy Scale [172]

In the past when travelling to a new location, I used the following strategies...

	Not at all true				Very true
I could visualize what was outside the building or complex in the direction I was heading inside the building	○	○	○	○	○
I asked for directions telling me whether to go east, west, north, or south at particular streets or landmarks	○	○	○	○	○
I visualized a map or layout of the area in my mind as I went	○	○	○	○	○
I appreciated the availability of someone (e.g., a receptionist) who could give me directions	○	○	○	○	○
I asked for directions telling me how many streets to pass before making each turn	○	○	○	○	○
I asked for directions telling me whether to turn right or left at particular streets or landmarks	○	○	○	○	○
Clearly visible signs pointing the way to different sections of the building or complex were important to me	○	○	○	○	○
As I went, I made a mental note of the mileage/distance I traveled on different roads	○	○	○	○	○
I found maps of the building or complex, with an arrow pointing to my present location, to be very helpful	○	○	○	○	○
I kept track of where I was in relation to the sun (or moon) in the sky as I went	○	○	○	○	○
I kept track of the relationship between where I was and the next place where I had to change direction	○	○	○	○	○

	Not at all true				Very true
I always kept in mind the direction from which I had entered the building or complex (e.g. north, south, east, or west side of the building)	○	○	○	○	○
I thought of my location in the building or complex in terms of north, south, east, and west	○	○	○	○	○
I kept track of the direction (north, south, east or west) in which I was going	○	○	○	○	○
Whenever I made a turn, I knew which direction I was facing	○	○	○	○	○
I kept track of where I was in relation to a reference point, such as the center of town, lake, river, or mountain	○	○	○	○	○
Clearly labeled room numbers and signs identifying parts of the building or complex were very helpful in finding my way	○	○	○	○	○

C.8 Sport Orientation Questionnaire [116]

Read each item carefully. Using the provided scale, please indicate your agreement with each item.

	Strongly disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Strongly agree
The best way to determine my ability is to set a goal and try to reach it.	○	○	○	○	○
Losing upsets me.	○	○	○	○	○
I am most competitive when I try to achieve personal goals.	○	○	○	○	○
I look forward to competing.	○	○	○	○	○
The only time I am satisfied is when I win.	○	○	○	○	○
I want to be successful in games.	○	○	○	○	○
I work hard to be successful in games.	○	○	○	○	○
I set goals for myself when I compete.	○	○	○	○	○
I want to be the best every time I compete.	○	○	○	○	○
Winning is important.	○	○	○	○	○
I am a determined competitor.	○	○	○	○	○
I hate to lose.	○	○	○	○	○
I enjoy competing against others.	○	○	○	○	○
I am a competitive person.	○	○	○	○	○
I try hardest when I have a specific goal.	○	○	○	○	○
I try my hardest to win.	○	○	○	○	○
The best test of my ability is competing against others.	○	○	○	○	○
Reaching personal performance goals is very important to me.	○	○	○	○	○
I perform my best when I am competing against an opponent.	○	○	○	○	○
My goal is to be the best gamer possible.	○	○	○	○	○

	Strongly disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Strongly agree
Scoring more points than my opponent is very important to me.	○	○	○	○	○
I thrive on competition.	○	○	○	○	○
Performing to the best of my ability is very important to me.	○	○	○	○	○
I look forward to the opportunity to test my skills in competition.	○	○	○	○	○
I have the most fun when I win.	○	○	○	○	○

C.9 Demographics for Manuscript A

Please answer the following questions.

What is your gender?

- Female
- Male
- Other
- Rather no say

What is your age?

Not at all	Slightly	Moderately	Quite a bit	Extremely

	Not at all	Slightly	Moderately	Quite a bit	Extremely
Are you an experienced computer user?	○	○	○	○	○

How many years have you been using computers for?

- Less than a year
- Between 1 and 2 years
- Between 3 and 5 years
- Between 5 and 10 years
- More than 10 years

How often (on average) do you use computers?

- Every day
- A few times per week
- Once per week

- A few times per month
- Once a month
- A few times per year
- Once per year
- Not at all

What is the diagonal size of your display? (Select the closest size or “I don’t know".)

- Less than 12 Inches
- 12 Inches
- 13 Inches
- 14 Inches
- 15 Inches
- 17 Inches
- 19 Inches
- 20 Inches
- 21 Inches
- 22 Inches
- 23 Inches
- 24 Inches
- 27 Inches
- More than 27 Inches
- I don’t know

Are you using a mouse or a trackpad?

- Mouse
- Trackpad
- Other

Do you use your mouse/trackpad with your right or left hand?

- Right
- Left

	Not at all	Slightly	Moderately	Quite a bit	Extremely
Are you experienced at video games?	○	○	○	○	○
Are you a gamer?	○	○	○	○	○
Are you experienced with keyboard and mouse input for games?	○	○	○	○	○

How many years have you been playing video games for?

- I don't play games
- Less than a year
- Between 1 and 2 years
- Between 3 and 5 years
- Between 5 and 10 years
- More than 10 years

How often (on average) do you play video games?

- Every day
- A few times per week
- Once per week
- A few times per month
- Once a month
- A few times per year
- Once per year
- Not at all

If you have played games more often in the past, how often were you playing at peak times?

- Every day
- A few times per week
- Once per week
- A few times per month
- Once a month
- A few times per year
- Once per year
- Not at all

Do you tend to play 2D or 3D games?

- Mostly 3D
- Mostly 3D, some 2D
- An equal amount of 2D and 3D
- Mostly 2D, some 3D
- Mostly 2D
- I don't play games

Which devices do you play games on? (Leave blank if none)

- Desktop Computers (e.g. Windows, OS X, Linux)
- Consoles (e.g. PlayStation, Xbox, Wii)
- Phones and Tablets (e.g. iOS and Android)
- Portable Gaming Devices (e.g. Sony PSP, Nintendo DS)

What genres do you enjoy playing? (Leave blank if none)

- Action
- Platform games
- First Person Shooter (FPS)
- Beat 'em up
- Adventure
- Role Playing Games (RPG)
- Massively Multiplayer Role Playing Games (MMORPG)
- Multiplayer Online Battle Arena Games (MOBA)
- Simulation
- Vehicle Simulation
- Strategy
- Music Games
- Puzzle Games
- Sport Games
- Casual Games
- Other

List your three favorite games (one game per line, leave blank if you are not a gamer).

Which of the following describes your playing style best?

- Socializer: I like to relate to other people when playing a game.
- Mastermind: I completely enjoy solving fiendish puzzles.
- Seeker: I really enjoy discovering curious and wonderful things when exploring virtual game worlds.
- Daredevil: I am all about the thrill of the chase when rushing in a video game and the excitement of risk taking gives me the edge.
- Survivor: I enjoy facing threatening situations in video games and enjoy escaping form threats barely alive.
- Achiever: I am extremely goal-oriented when playing and enjoy collecting and completing everything I find in a game.
- Conqueror: I like to be up against impossible odds and defeat my opponents with a crushing victory.

C.10 Demographics for Manuscript B

Please answer the following questions.

What is your gender?

- Female
- Male
- Other
- Rather no say

What is your age?

What is the diagonal size of your display? (Select the closest size or "I don't know".)

- Less than 12 Inches
- 12 Inches
- 13 Inches
- 14 Inches
- 15 Inches
- 17 Inches
- 19 Inches
- 20 Inches
- 21 Inches
- 22 Inches
- 23 Inches
- 24 Inches
- 27 Inches
- More than 27 Inches
- I don't know

Are you using a mouse or a trackpad?

- Mouse
- Trackpad
- Other

Do you use your mouse/trackpad with your right or left hand?

- Right
- Left

	Not at all	Slightly	Moderately	Quite a bit	Extremely
Are you experienced at playing video games?	○	○	○	○	○
Are you familiar with navigating 3D virtual environments?	○	○	○	○	○
Are you experienced at playing first-person shooter games?	○	○	○	○	○
Are you a gamer?	○	○	○	○	○
Are you experienced with using keyboard and mouse input simultaneously to control games?	○	○	○	○	○

How many years have you been playing video games for?

- I don't play games
- Less than a year
- Between 1 and 2 years
- Between 3 and 5 years
- Between 5 and 10 years
- More than 10 years

How often (on average) do you play video games?

- Every day
- A few times per week
- Once per week
- A few times per month
- Once a month
- A few times per year
- Once per year
- Not at all

If you have played games more often in the past, how often were you playing at peak times?

- Every day
- A few times per week
- Once per week
- A few times per month
- Once a month
- A few times per year
- Once per year
- Not at all

Do you tend to play 2D or 3D games?

- Mostly 3D
- Mostly 3D, some 2D
- An equal amount of 2D and 3D
- Mostly 2D, some 3D
- Mostly 2D
- I don't play games

Which devices do you play games on? (Leave blank if none)

- Desktop Computers (e.g. Windows, OS X, Linux)
- Consoles (e.g. PlayStation, Xbox, Wii)
- Phones and Tablets (e.g. iOS and Android)
- Portable Gaming Devices (e.g. Sony PSP, Nintendo DS)

What genres do you enjoy playing? (Leave blank if none)

- Action
- Platform games
- First Person Shooter (FPS)
- Beat 'em up
- Adventure
- Role Playing Games (RPG)
- Massively Multiplayer Role Playing Games (MMORPG)
- Multiplayer Online Battle Arena Games (MOBA)
- Simulation
- Vehicle Simulation
- Strategy
- Music Games
- Puzzle Games
- Sport Games
- Casual Games
- Other

List your three favorite games (one game per line, leave blank if you are not a gamer).

C.11 Demographics for Manuscript C

Please answer the following questions.

What is your gender?

- Female
- Male
- Other
- Rather not say

What is your age?

What is the diagonal size of your display? (Select the closest size or "I don't know".)

- Less than 12 Inches
- 12 Inches
- 13 Inches
- 14 Inches
- 15 Inches
- 17 Inches
- 19 Inches
- 20 Inches
- 21 Inches
- 22 Inches
- 23 Inches
- 24 Inches
- 27 Inches
- More than 27 Inches
- I don't know

Are you right handed or left handed?

- Right
- Left
- Ambidextrous

	Not at all	Slightly	Moderately	Quite a bit	Extremely
Are you a gamer?	○	○	○	○	○
Are you experienced at playing video games?	○	○	○	○	○

	Not at all	Slightly	Moderately	Quite a bit	Extremely
Do you ever become so involved in a video game that it is as if you are inside the game rather than moving a joystick or pressing buttons and watching the screen?	○	○	○	○	○

How many years have you been playing video games for?

- I don't play games
- Less than a year
- Between 1 and 2 years
- Between 3 and 5 years
- Between 5 and 10 years
- More than 10 years

How often (on average) do you play video games?

- Every day
- A few times per week
- Once per week
- A few times per month
- Once a month
- A few times per year
- Once per year
- Not at all

If you have played games more often in the past, how often were you playing at peak times?

- Every day
- A few times per week
- Once per week
- A few times per month
- Once a month
- A few times per year
- Once per year
- Not at all

What genres do you enjoy playing? (Leave blank if none)

- Action
- Platform games
- First Person Shooter (FPS)
- Beat 'em up

- Adventure
- Role Playing Games (RPG)
- Massively Multiplayer Role Playing Games (MMORPG)
- Multiplayer Online Battle Arena Games (MOBA)
- Battle Royale Games
- Simulation
- Vehicle Simulation
- Strategy
- Music Games
- Puzzle Games
- Sport Games
- Casual Games
- Other

Are you familiar with the game 'Super Hexagon'?

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

C.12 Demographics for Manuscript D

Please answer the following questions.

What is your gender?

- Man
- Woman
- Non-binary
- Prefer not to answer

What is your age?

How much do you self-identify as a gamer on the following scale?

Not at all Gamer

How familiar are you with side-scrolling platform games?

Not at all Very Familiar

How familiar are you with Super Mario games?

Not at all Very Familiar

How familiar are you with the game ‘Super Meat Boy’?

Not at all Very Familiar

How familiar are you with the game ‘Speedrunners’?

Not at all Very Familiar

How familiar are you with the game ‘Celeste’?

Not at all Very Familiar

How often (on average) do you play games?

- Every day
- A few times per week
- Once per week
- A few times per month
- Once a month
- A few times per year
- Once per year
- Not at all

If you have played games more often in the past, how often were you playing at peak times?

- Every day

- A few times per week
- Once per week
- A few times per month
- Once a month
- A few times per year
- Once per year
- Not at all

Appendix D

Questionnaires for Outcome Measures

D.1 NASA Task-Load Index (TLX) [125]

Mental Demand

How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, forgiving or exacting?

Low										High
○	○	○	○	○	○	○	○	○	○	○

Physical Demand

How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Low										High
○	○	○	○	○	○	○	○	○	○	○

Temporal Demand

How much time pressure did you feel due to the rate at which the task elements occurred? Was the pace slow and leisurely or rapid and frantic?

Low										High
○	○	○	○	○	○	○	○	○	○	○

Performance

How successful do you think you were in accomplishing the goals of the task set by the experiment (or yourself)? How satisfied were you with your performance in accomplishing these goals?

Low										High
○	○	○	○	○	○	○	○	○	○	○

Effort

How hard did you have to work (mentally and physically) to accomplish your level of performance?

	Low									High
	○	○	○	○	○	○	○	○	○	○	○

Frustration

How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Low										High
○	○	○	○	○	○	○	○	○	○	○

D.2 State-Trait Anxiety Inventory (STAI) [189]

A number of statements which people have used to describe themselves are given below. Read each statement and then choose based on how you feel right now.

	Not at all	Somewhat	Moderately	Very Much
I am tense	○	○	○	○
I feel content	○	○	○	○
I am worried	○	○	○	○
I feel calm	○	○	○	○
I am relaxed	○	○	○	○
I feel upset	○	○	○	○

D.3 Intrinsic Motivation Inventory (IMI) [193]

Reflect on your play experiences and rate your agreement with the following statements.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
After playing the game for a while, I felt pretty competent.	○	○	○	○	○
I put a lot of effort into this game.	○	○	○	○	○
I couldn't play this game very well.	○	○	○	○	○
This game did not hold my attention.	○	○	○	○	○
Playing the game was fun.	○	○	○	○	○
It was important to me to do well at this game.	○	○	○	○	○
I would describe this game as very interesting.	○	○	○	○	○
I felt pressured while playing the game.	○	○	○	○	○
I enjoyed this game very much.	○	○	○	○	○
I didn't try very hard at playing the game.	○	○	○	○	○
I tried very hard while playing the game.	○	○	○	○	○
I felt tense while playing the game.	○	○	○	○	○
I was anxious while playing the game.	○	○	○	○	○
While playing the game, I was thinking about how much I enjoyed it.	○	○	○	○	○

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I think I am pretty good at this game.	○	○	○	○	○
I am pretty skilled at the game.	○	○	○	○	○
I was very relaxed while playing the game.	○	○	○	○	○
I am satisfied with my performance at this game.	○	○	○	○	○

D.4 Flow Short Scale (FSS) [90]

Read each item carefully. Using the provided scale, please indicate your agreement with each item.

	Not at all						Very much
The right thoughts/movements occur of their own accord.	○	○	○	○	○	○	○
My thoughts/activities run fluidly and smoothly	○	○	○	○	○	○	○
I am worried about failing.	○	○	○	○	○	○	○
I am totally absorbed in what I am doing.	○	○	○	○	○	○	○
I feel just the right amount of challenge.	○	○	○	○	○	○	○
I feel that I have everything under control.	○	○	○	○	○	○	○
I do not notice time passing.	○	○	○	○	○	○	○
I am completely lost in thought.	○	○	○	○	○	○	○
I have no difficulty concentrating.	○	○	○	○	○	○	○
Something important to me is at stake here.	○	○	○	○	○	○	○
I must not make any mistakes here.	○	○	○	○	○	○	○
My mind is completely clear.	○	○	○	○	○	○	○
I know what I have to do each step of the way.	○	○	○	○	○	○	○

	Easy						Difficult
Compared to all other activities which I partake in, this one is...	○	○	○	○	○	○	○

	Low						High
I think my competence in this area is...	○	○	○	○	○	○	○

	Too low			Just right			Too high
For me personally, the current demands are...	○	○	○	○	○	○	○

D.5 Flow Scale Short (FKS) [324]

Think back to when you were playing the game. For each statement, indicate how much it describes the way you felt at that time.

	Not at all						Very much
I feel that I have everything under control.	○	○	○	○	○	○	○
I don't notice time passing.	○	○	○	○	○	○	○
I won't make any mistake here.	○	○	○	○	○	○	○
My mind is completely clear.	○	○	○	○	○	○	○
I know what I have to do each step of the way.	○	○	○	○	○	○	○
Something important to me is at stake here.	○	○	○	○	○	○	○
My thoughts/activities run fluidly and smoothly.	○	○	○	○	○	○	○
The right thoughts/movements occur of their own accord.	○	○	○	○	○	○	○
I feel just the right amount of challenge.	○	○	○	○	○	○	○
I have no difficulty concentrating.	○	○	○	○	○	○	○
I am completely lost in thought.	○	○	○	○	○	○	○
I am worried about failing.	○	○	○	○	○	○	○
I am totally absorbed in what I am doing.	○	○	○	○	○	○	○

D.6 Player Experience Inventory (PXI) [1]

Reflect on the task and rate your agreement with the following statements.

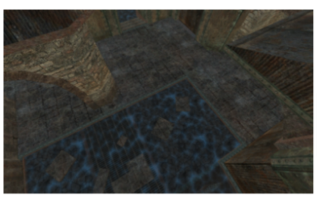
	Strongly disagree						Strongly agree
I appreciated the aesthetics of the task.	○	○	○	○	○	○	○
The challenges in the task were at the right level of difficulty for me.	○	○	○	○	○	○	○
I felt free to complete the task in my own way.	○	○	○	○	○	○	○
I wanted to explore how the task evolved.	○	○	○	○	○	○	○
I grasped the overall goal of the task.	○	○	○	○	○	○	○
The task was not too easy and not too hard to complete.	○	○	○	○	○	○	○
I felt like I had choices regarding how I wanted to complete this task.	○	○	○	○	○	○	○
I wanted to find out how the task progressed.	○	○	○	○	○	○	○
I understood the objectives of the task.	○	○	○	○	○	○	○
I felt a sense of mastery completing this task.	○	○	○	○	○	○	○
The task informed me of my progress in the task.	○	○	○	○	○	○	○
I was immersed in the task.	○	○	○	○	○	○	○
I felt capable while completing the task.	○	○	○	○	○	○	○
I could easily assess how I was performing in the task.	○	○	○	○	○	○	○
The goals of the task were clear to me.	○	○	○	○	○	○	○
I was no longer aware of my surroundings while I was completing the task.	○	○	○	○	○	○	○
I felt I was good at completing this task.	○	○	○	○	○	○	○
It was easy to know how to perform actions in the task.	○	○	○	○	○	○	○
The task was challenging but not too challenging.	○	○	○	○	○	○	○
I felt a sense of freedom about how I wanted to complete this task.	○	○	○	○	○	○	○
I enjoyed the way the task was styled.	○	○	○	○	○	○	○

	Strongly disagree						Strongly agree
Completing this task was valuable to me.	○	○	○	○	○	○	○
The actions to control the task were clear to me.	○	○	○	○	○	○	○
The task gave clear feedback on my progress towards the goals.	○	○	○	○	○	○	○
I liked the look and feel of the task.	○	○	○	○	○	○	○
I thought the task was easy to control.	○	○	○	○	○	○	○
I was fully focused on the task.	○	○	○	○	○	○	○
I felt eager to discover how the task continued.	○	○	○	○	○	○	○
Completing the task was meaningful to me.	○	○	○	○	○	○	○
The task felt relevant to me.	○	○	○	○	○	○	○

D.7 Manuscript A Spatial Knowledge Questions

Instructions

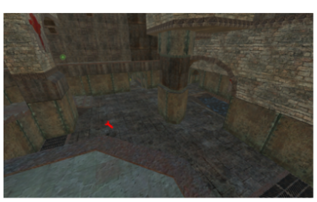
Choose the labelled location (A to F) on the map that corresponds to the location visible in the screenshot.



○ A ○ B ○ C ○ D ○ E ○ F

How confident are you in your answer to the previous question?

	Not at all	Slightly	Moderately	Quite a bit	Extremely
	○	○	○	○	○



○ A ○ B ○ C ○ D ○ E ○ F

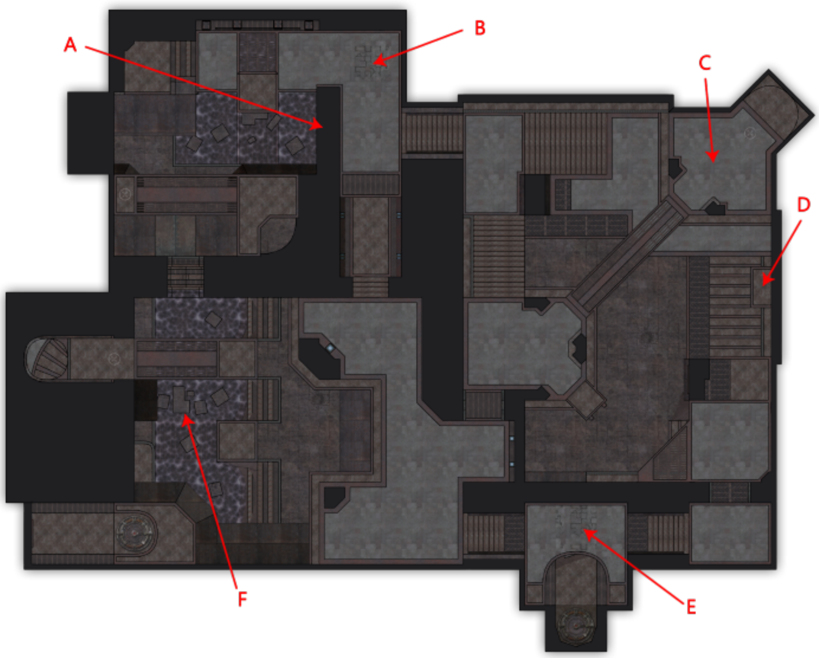
How confident are you in your answer to the previous question?

	Not at all	Slightly	Moderately	Quite a bit	Extremely
	○	○	○	○	○

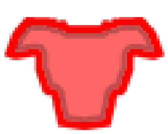
○ A ○ B ○ C ○ D ○ E ○ F

How confident are you in your answer to the previous question?

	Not at all	Slightly	Moderately	Quite a bit	Extremely
	○	○	○	○	○



For each question, please provide the best guess you can.

1. Where is the  **?**

○ A ○ B ○ C ○ D ○ E ○ F

1. b) How confident are you in your answer to the previous question (Question #1)?

	Not at all	Slightly	Moderately	Quite a bit	Extremely
	○	○	○	○	○

2. Where is the  **?**

○ A ○ B ○ C ○ D ○ E ○ F

2. b) How confident are you in your answer to the previous question (Question #2)?

	Not at all	Slightly	Moderately	Quite a bit	Extremely
	○	○	○	○	○

3. Where is the **?**

○ A ○ B ○ C ○ D ○ E ○ F

3. b) How confident are you in your answer to the previous question (Question #3)?

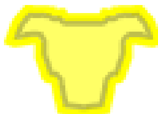
	Not at all	Slightly	Moderately	Quite a bit	Extremely
	○	○	○	○	○

4. Where is the  **?**

○ A ○ B ○ C ○ D ○ E ○ F

4. b) How confident are you in your answer to the previous question (Question #4)?

	Not at all	Slightly	Moderately	Quite a bit	Extremely
	○	○	○	○	○

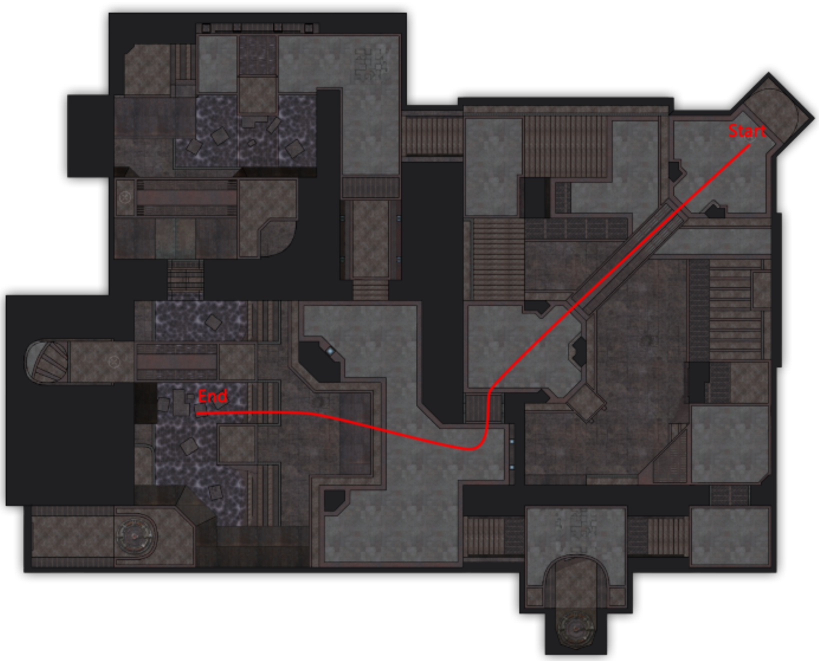
5. Where are the two  **?**

○ A ○ B ○ C ○ D ○ E ○ F

5. b) How confident are you in your answer to the previous question (Question #5)?

	Not at all	Slightly	Moderately	Quite a bit	Extremely
	○	○	○	○	○

Estimate the time it would take to travel this route



6. How much time (in seconds) would it take to directly traverse the route shown in the image above?

Instructions

Choose the labelled location (A to F) on the map that corresponds to the location visible in the screenshot.

○ A ○ B ○ C ○ D ○ E ○ F

How confident are you in your answer to the previous question?

	Not at all	Slightly	Moderately	Quite a bit	Extremely
	○	○	○	○	○

○ A ○ B ○ C ○ D ○ E ○ F

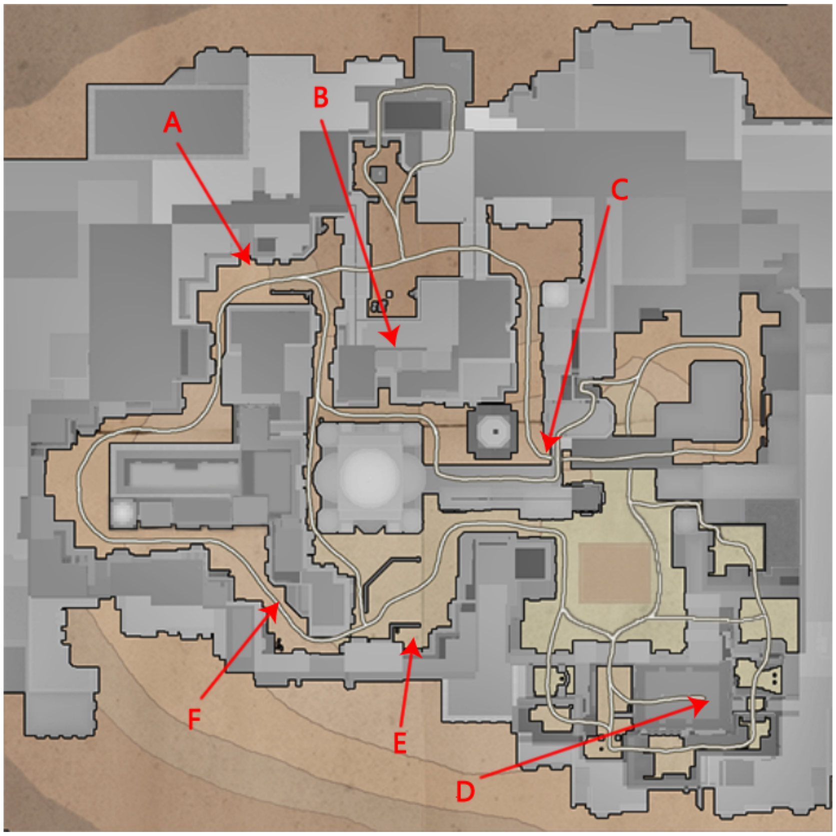
How confident are you in your answer to the previous question?

	Not at all	Slightly	Moderately	Quite a bit	Extremely
	○	○	○	○	○

○ A ○ B ○ C ○ D ○ E ○ F

How confident are you in your answer to the previous question?

	Not at all	Slightly	Moderately	Quite a bit	Extremely
	○	○	○	○	○



For each question, please provide the best guess you can.

1. Where is the  **?**

○ A ○ B ○ C ○ D ○ E ○ F

1. b) How confident are you in your answer to the previous question (Question #1)?

	Not at all	Slightly	Moderately	Quite a bit	Extremely
	○	○	○	○	○

2. Where is the  **?**

○ A ○ B ○ C ○ D ○ E ○ F

2. b) How confident are you in your answer to the previous question (Question #2)?

Not at all	Slightly	Moderately	Quite a bit	Extremely
○	○	○	○	○

3. Where is the  **?**

○ A ○ B ○ C ○ D ○ E ○ F

3. b) How confident are you in your answer to the previous question (Question #3)?

Not at all	Slightly	Moderately	Quite a bit	Extremely
○	○	○	○	○

4. Where is the  **?**

○ A ○ B ○ C ○ D ○ E ○ F

4. b) How confident are you in your answer to the previous question (Question #4)?

Not at all	Slightly	Moderately	Quite a bit	Extremely
○	○	○	○	○

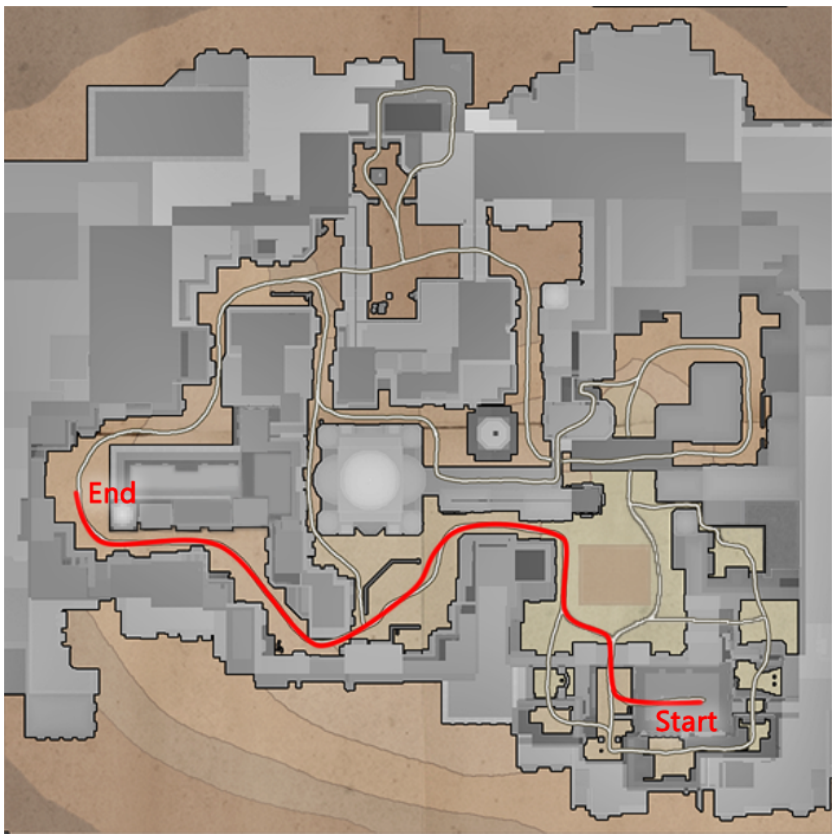
5. Where are the two  **?**

○ A ○ B ○ C ○ D ○ E ○ F

5. b) How confident are you in your answer to the previous question (Question #5)?

	Not at all	Slightly	Moderately	Quite a bit	Extremely
	○	○	○	○	○

Estimate the time it would take to travel this route



6. How much time (in seconds) would it take to directly traverse the route shown in the image above?

D.8 Debrief for Manuscript A

Thank you for participating in our study!

What did you think this study was about?

Did you use a strategy to intentionally learn the map layout?

- Yes
- No

If you did use a strategy, describe it as best as you are able.

If you did use a strategy, when did you start to use it?

- I didn’t use a strategy
- Right from the beginning of the experiment
- When navigating the first map to find the flags
- When navigating the second map to find the flags
- When answering questions regarding the locations of rooms or landmarks on the two maps
- During the first map of the landmark navigation task
- During the second map of the landmark navigation task

Did you intentionally use or avoid teleporters?

- I used them intentionally
- I used them unintentionally
- I avoided them intentionally
- I avoided them unintentionally

If your decision to use teleporters was intentional what led you to make it? If this was unintentional, why might you have avoided or used them?

Did you intentionally use or avoid the jump pads/bounce pads?

- I used them intentionally
- I used them unintentionally
- I avoided them intentionally
- I avoided them unintentionally

If your decision to use jump pads was intentional what led you to make it? If this was unintentional, why might you have avoided or used them?

Do you have any additional comments?

D.9 Debrief for Manuscript B

Do you make use of any intentional strategies to remember the specific routes you had trained on previously?

- Yes
- No

If you did, can you briefly describe what you did to help you remember the routes?

Do you make use of any intentional strategies to remember the locations of the landmarks?

- Yes
- No

If you did, can you briefly describe what you did to help you remember the location of landmarks?

Appendix E

Additional Analyses, Results, and Figures

When writing the manuscripts, there were times when data or analyses did not make it into the final manuscript. Additionally, my committee was interested in reporting on gender differences for each of the manuscripts. This appendix contains these cut analyses or data, as well as some new analyses.

E.1 Manuscript A

E.1.1 Study 1

Figure E.1: The map used for the Furious Heights environment from *Quake Live*. In the experiment, icons indicating the location of items were included.

Individual Differences and Covariates

The statistical analyses of the manuscripts each included covariates, to account for individual differences between participants. Essentially the "correct" for these differences (e.g., gaming experience) so that they are less likely to be the reason for between-group differences. A significant covariate means that the dependent measure was significantly affected by the covariate (and if the covariate were not included in the analyses, I may have ended up with misleading results).

There were no significant covariates.

	Training						Transfer					
	Completion Time			Distance Travelled			Completion Time			Distance Travelled		
Covariate	$F_{1,22}$	p	η_p^2	$F_{1,22}$	p	η_p^2	$F_{1,22}$	p	η_p^2	$F_{1,22}$	p	η_p^2
Focus (ITQ)	4.21	.052	.161	0.35	.562	.016	0.39	.541	.017	1.62	.216	.069
Games (ITQ)	1.53	.229	.065	0.03	.856	.002	0.75	.395	.033	0.24	.626	.011
Orientation Strategy	0.23	.639	.010	0.59	.449	.026	0.14	.714	.006	0.13	.726	.006
Gaming Experience	1.24	.278	.053	<0.01	.979	<.001	2.13	.158	.088	0.19	.666	.009
Gender	1.18	.289	.051	0.45	.509	.020	1.37	.255	.058	0.58	.456	.025

Table E.1: Between-subject covariates for Manuscript A, Study 1's dependent measures.

Gender Differences

To investigate whether navigation performance differed between participants' genders, I performed separate ANOVAs with gender as the independent variable for each measure in both sessions (training and transfer). I also performed separate ANCOVAs, including gaming experience as a covariate. This was done as it seemed probable that differences in performance would be due to gaming expertise rather than gender, it may simply have been that the women who completed our study were less experienced with games overall.

We asked participants about their gender via the prompt: "What is your gender?" And the following options: "Female", "Male", "Other", or "Rather not say". For no assistance, we had 5 female and 5 male participants, for moderate assistance we had 6 female and 4 male participants, and for strong assistance, we had 4 female and 6 male participants.

The results (Table E.2) show that if gaming experience is not included as a covariate, then there is a significant main effect of gender on the transfer session's completion time, where female participants completed the transfer session faster than male participants. Those gender differences disappeared if gaming experience was included as a covariate in the analysis.

					With Gaming Experience as a Covariate					
	Training		Transfer		Training			Transfer		
Measure	$F_{1,28}$	p	$F_{1,28}$	p	$F_{1,27}$	p	η_p^2	$F_{1,27}$	p	η_p^2
Completion Time	2.82	.104	**7.61**	**.010**	0.29	.594	.011	1.74	.198	.061
Distance Travelled	0.15	.698	0.22	.643	0.02	.899	.001	0.27	.607	.010

Table E.2: Gender differences for Study 1 of Manuscript A.

Debrief Responses

I will not provide a full qualitative analysis of the responses here, only share some highlights, as well as the counts of the responses to each question asked.

When asked if they used a strategy to intentionally learn the map layout, only five participants said "yes" (25 said "no").

When asked about teleporter use, 22 participants said they used them intentionally, six said they used them unintentionally, and two said they avoided them unintentionally.

When asked about jump pad use, 16 participants said they used them intentionally, eight said they used them unintentionally, two said they avoided them intentionally, and four said they avoided them unintentionally.

Responses for the five participants participants who intentionally used a strategy included:

- "I tried to remember what was outside and inside but that's about it. I'm terrible with maps and sense of direction."
- "I was trying to remember the location of the objects by using other objects as references."
- "I don't know that I would call it a strategy – primarily just a combination of memory and visualization, which clearly failed miserably."
- "I attempted to memorize any landmarks"
- "i tried to remember where each item was and which direction i took"

E.1.2 Study 2

Individual Differences and Covariates

Gaming experience was a significant covariate for completion time during both the training and transfer sessions

	Training						**Transfer**					
	Completion Time			Distance Travelled			Completion Time			Distance Travelled		
Covariate	$F_{1,36}$	p	η_p^2	$F_{1,36}$	p	η_p^2	$F_{1,36}$	p	η_p^2	$F_{1,36}$	p	η_p^2
Focus (ITQ)	2.92	.096	.075	0.23	.635	.006	0.19	.668	.005	0.06	.804	.002
Games (ITQ)	0.08	.779	.002	<0.01	.962	<.001	0.06	.804	.002	0.73	.398	.020
Orientation Strategy	2.78	.104	.072	0.49	.489	.013	1.39	.246	.037	1.11	.298	.030
Gaming Experience	**8.25**	**.007**	**.186**	0.69	.411	.019	**8.37**	**.006**	**.189**	.046	.831	.001
Gender	1.71	.199	.045	0.10	.748	.003	1.92	.175	.051	.583	.450	.016

Table E.3: Results of ANOVAs and ANCOVAs investigating gender differences in Study 1 of Manuscript A.

Gender Differences

To investigate whether navigation performance differed between participants' genders, I performed separate ANOVAs with gender as the independent variable for each measure in both sessions (training and transfer). Like with Study 1, I also performed separate ANCOVAs, including gaming experience as a covariate.

We asked participants about their gender via the prompt: "What is your gender?" And the following options: "Female", "Male", "Other", or "Rather not say". For no assistance, we had 5 female and 9 male participants, for moderate assistance we had 6 female and 10 male participants, and for strong assistance, we had 4 female and 10 male participants.

The results (Table E.4) show that if gaming experience is not included as a covariate, then there is a significant main effect of gender on the training session's completion time as well as the transfer session's completion time; men completed the sessions faster than women. Those gender differences disappeared if gaming experience was included as a covariate in the analysis.

					With Gaming Experience as a Covariate					
	Training		**Transfer**		**Training**			**Transfer**		
Measure	$F_{1,42}$	p	$F_{1,42}$	p	$F_{1,41}$	p	η_p^2	$F_{1,41}$	p	η_p^2
Completion Time	**5.31**	**.026**	**9.26**	**.004**	1.04	.314	.025	3.15	.083	.071
Distance Travelled	0.62	.436	0.26	.612	0.03	.867	.001	0.52	.477	.012

Table E.4: Results of ANOVAs and ANCOVAs investigating gender differences in Study 2 of Manuscript A.

Debrief Responses

Like Study A, I will not provide a full qualitative analysis of the responses here, only share the counts of the responses to each question asked.

- When asked if they used a strategy to intentionally learn the map layout, only 11 participants (25%) said “yes”.
- When asked about teleporter use, 40 participants said they used them intentionally, 3 said they used them unintentionally, and 1 said they avoided them intentionally.
- When asked about jump pad use, 33 participants said they used them intentionally, 9 said they used them unintentionally, 1 said they avoided them intentionally, and 1 said they avoided them unintentionally.

E.2 Manuscript B

E.2.1 Study 1

Individual Differences and Covariates

Gaming experience was a significant covariate for completion time in the training and transfer sessions. Wayfinding anxiety was a significant covariate for completion time and distance travelled in the training session, and gaming immersive tendencies was a significant covariate for distance travelled in the transfer session.

	Training						**Transfer**					
	Completion Time			Distance Travelled			Completion Time			Distance Travelled		
Covariate	$F_{1,37}$	p	η_p^2	$F_{1,37}$	p	η_p^2	$F_{1,37}$	p	η_p^2	$F_{1,37}$	p	η_p^2
Wayfinding Anxiety	**4.89**	**.033**	**.117**	**4.94**	**.032**	**.118**	1.49	.230	.039	0.79	.380	.021
Orientation Strategy	0.45	.505	.012	<0.01	.989	<.001	0.24	.628	.006	<0.01	.965	<.001
Gaming (ITQ)	<0.01	.983	<.001	0.31	.584	.008	1.76	.193	.045	**8.51**	**.006**	**.187**
Gaming Experience	**10.8**	**.002**	**.225**	0.49	.487	.013	**10.5**	**.002**	**.221**	0.02	.878	.001

Table E.5: Individual differences and covariates for Manuscript B, Study 1.

Gender Differences

To investigate whether navigation performance differed between participants’ genders, I performed separate ANOVAs with gender as the independent variable for each measure in both sessions (training and transfer). I also performed separate ANCOVAs, including gaming experience as a covariate. This was done as it seemed probable that differences in performance would be due to gaming expertise rather than gender, it may simply have been that female participants were less experienced at games overall.

We asked participants about their gender via the prompt: “What is your gender?” And the following options: “Female”, “Male”, “Other”, or “Rather not say”. For map assistance, we had 5 female and 9 male participants, for position assistance we had 6 female and 10 male participants, and for trail assistance, we had 4 female and 10 male participants.

We found a significant main effect of gender on the transfer session’s completion time when gaming experience was not included as a covariate; men completed the transfer session faster than women.

					With Gaming Experience as a Covariate					
	Training		**Transfer**		**Training**			**Transfer**		
Measure	$F_{1,42}$	p	$F_{1,42}$	p	$F_{1,41}$	p	η_p^2	$F_{1,41}$	p	η_p^2
Completion Time	3.16	.083	**5.63**	**.022**	0.49	.490	.012	1.49	.230	.035
Distance Travelled	0.26	.611	0.78	.381	<0.01	.984	<.001	1.65	.207	.039

Table E.6: Results of ANOVA and ANCOVAs investigating gender differences in Study 1 of Manuscript B.

E.2.2 Study 2

Individual Differences and Covariates

Visual memory was a significant covariate for transfer completion time. Gaming experience was a significant covariate for completion time in the training, transfer, and retention sessions, as well as completion time in the retention session.

	Training						Transfer						Retention					
	Completion Time			Distance Travelled			Completion Time			Distance Travelled			Completion Time			Distance Travelled		
Covariate	$F_{1,73}$	p	η_p^2	$F_{1,73}$	p	η_p^2	$F_{1,74}$	p	η_p^2	$F_{1,72}$	p	η_p^2	$F_{1,65}$	p	η_p^2	$F_{1,63}$	p	η_p^2
Visual Memory	1.43	.236	.02	1.95	.167	.03	**6.69**	**.012**	**.08**	2.00	.161	.03	2.11	.151	.03	0.84	.837	.00
Gaming (ITQ)	0.23	.632	.10	1.34	.728	.00	0.07	.792	.00	0.21	.645	.00	1.11	.296	.02	0.31	.577	.00
Gaming Experience	**7.98**	**.006**	**.00**	0.12	.728	.02	**13.9**	**<.001**	**.16**	2.61	.108	.04	**27.4**	**<.001**	**.30**	**7.53**	**.008**	**.11**

Table E.7: Individual differences and covariates for Manuscript B, Study 2.

Gender Differences

To investigate whether navigation performance differed between participants' genders, I performed a separate RM-ANOVA with gender as the independent variable for each measure for the training session. For the retention and transfer sessions, I also performed separate ANOVAs for each measure and each session. Like in Study 1, I also did the analyses a second time with gaming experience as a covariate, to see if any gender differences were actually due to gaming experience instead of gender.

We asked participants about their gender via the prompt: "What is your gender?" And the following options: "Female", "Male", "Other", or "Rather not say". For map assistance, we had 10 female and 14 male participants, for trail assistance we had 15 female and 13 male participants, and for rails assistance, we had 12 female and 15 male participants.

For Training Completion Time and Distance RM-ANOVA. For Retention and Transfer, separate ANOVAs.

							With Gaming Experience as a Covariate								
	Training		Transfer		Retention		Training			Transfer			Retention		
Measure	$F_{1,68}$	p	$F_{1,77}$	p	$F_{1,42}$	p	$F_{1,67}$	p	η_p^2	$F_{1,76}$	p	η_p^2	$F_{1,41}$	p	η_p^2
Completion Time	2.33	.131	**8.40**	**.005**	**6.74**	**.012**	1.43	.235	.019	3.84	.054	.048	1.59	.212	.023
Distance Travelled	0.01	.940	0.04	.842	1.14	.289	0.02	.893	<.001	<0.01	.450	.008	0.03	.863	<.001

Table E.8: Caption

Between-Subjects Completion Time

To explore the between-subjects differences in completion time for the training and retention session, I performed an RM-ANCOVA with Completion Time as the repeated-measures factor, Session (training and retention) as the within-subjects factor, and Assistance as the between-subjects factor. Instead of including every day of the training session, only the last day was included, as well as the retention session. For covariates, the same covariates that were used in the performance measure analyses within the manuscript were used

There was a main effect of Session on Completion Time ($F_{1,65} = 133$, $p < .001$, $\eta_p^2 = .672$); participants completed the navigation tasks more slowly in the retention session. There was also a significant interaction between Session and Assistance ($F_{2,65} = 38.9$, $p < .001$, $\eta_p^2 = .545$). Post-hoc pairwise comparisons (using Bonferroni corrections) show that with no assistance, there was no change in performance between the last training day and the retention session ($p > .999$). However, for Trail and Rail assistance, participants were slower at navigating the environment in the retention session compared to the last day of training ($p < .001$). However, in the retention session, all groups performed similarly to each other ($p \geq .666$).

E.2.3 Strategy Use

In Study 2 of Manuscript B, I asked participants two questions relating to intentional strategy use: "do you make use of any intentional strategies to remember the specific routes you had trained on previously?", and "Do you make use of any intentional strategies to remember the location of the landmarks?" 11 participants (13.8%) used an intentional strategy to remember the routes, and 28 participants (35%) used an intentional strategy to remember the landmark locations.

For route memory strategies, 13 participants said that they used landmarks as a wayfinding aid to help them with the routes. For example, *"I made a note of places that were close to each other such as the tank and the bridge. Those mental notes allowed me to find certain locations faster."*, or *"I kept in mind certain objects that I encounter in some routes, such as the tank and the bridge."*

For landmark memory strategies, 10 participants said that they used the location or proximity of other landmarks to remember the location of other landmarks. For example, *"I tried to remember what was 'close' to another thing. Like the bookshelf and desk were not too far apart and such."* 8 participants considered the visual characteristics of the surrounding scenery, e.g., "I tried to recall how the areas surrounding the landmarks looked like, such as whether it was an indoor room or had trees around it." 6 participants considered the elevation of the landmark, e.g., *"I kind of assigned them levels. I knew which level the door, the tank, the shelf, etc. were on. So I just had to find my way to that level and then look for things that reminded me of it."* 2 participants attempted to construct a mental map of the environment, e.g., *"I imagined how the map would look in a 2D representation, but I wish I had the time to study all the various shortcuts, and to explore the inside of the other houses, it would be funny!"*

Note that I did not perform any sort of analysis on the written responses for this appendix. I simply picked out one example of each type of response.

E.3 Manuscript C

E.3.1 Covariates

In Manuscript C, I determined if there were performance differences between the groups via a RM-ANCOVA. I did not, however, report on the effects of the covariates. Figure E.9 shows all between-subjects effects, including those of the covariates. It is apparent that the first session's performance significantly predicted later performance in the task, as did a participant's competitiveness and win orientation.

	df	F	p	η_p^2
Interval Group	**4**	**3.65**	**.009**	**.147**
Session 1 Performance	**1**	**269**	**<.001**	**.760**
Gaming Experience	1	1.40	.240	.016
Focus (ITQ)	1	2.93	.091	.033
Involvement (ITQ)	1	1.44	.233	.017
Attentional Control	1	0.10	.759	.001
Competetiveness (SOQ)	**1**	**5.92**	**.017**	**.065**
Goal Orientation (SOQ)	1	0.20	.658	.002
Win Orientation (SOQ)	**1**	**4.18**	**.044**	**.047**
Anxiety (QCM)	1	0.70	.405	.008
Challenge (QCM)	1	1.72	.194	.020
Probability (QCM)	1	3.43	.068	.039
Amotivation (SIMS)	1	0.77	.381	.009
External Regulation (SIMS)	1	2.06	.154	.024
Identified Regulation (SIMS)	1	0.35	.557	.004
Intrinsic Motivation (SIMS)	1	2.78	.099	.032
Residual	85			

Table E.9: Between Subjects Effects for The RM-ANCOVA used in Manuscript C.

E.3.2 Gender Differences

To explore gender differences, I performed an RM-ANCOVA with Performance (average life time) per Session (2, 3, 4, and Retention) as the repeated-measures factor, and Gender as the between-subject factor. Performance in Session 1 was used as a covariate, to compensate for individual differences in initial performance. I found no effect of Gender on Performance ($F_{1,102} = 1.56$, $p = .215$, $\eta_p^2 = .015$).

Because the initial performance may have been affected by gender, I performed a second RM-ANCOVA with the only difference being that Session 1 performance was included as a repeated measure instead of as a covariate. I did find a main effect of Gender on Performance with this approach ($F_{1,103} = 22.8$, $p < .001$, $\eta_p^2 = .181$). Male participants performed better than female participants.

We asked participants about their gender via the prompt: "What is your gender?" And the following options: "Female", "Male", "Other", or "Rather not say". See Table E.10 for genders in each Interval Group.

Gender	3-Second	2-Minute	5-Minute	10-Minute	1-Day
Female	12	9	12	6	13
Male	8	13	9	15	8

Table E.10: Counts of the participants' genders in each Interval Group.

E.4 Manuscript D

E.4.1 Additional Figures

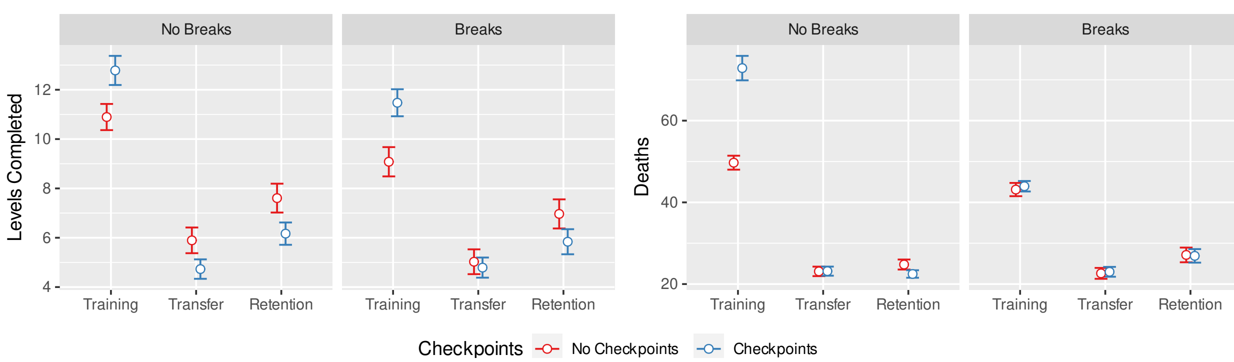


Figure E.2: Raw means for the performance measures of Manuscript D. Error bars are standard error.

In the manuscript, the plot with the performance measures contained the estimated marginal means for performance instead of the raw means. This was done because including the covariates (particularly total playtime) resulted in estimated marginal means much different than the raw means.

E.4.2 Gender Differences

To explore gender differences, I performed an RM-ANCOVA with Levels Completed as the repeated-measures factor, Session (training, transfer, and retention) as the within-subjects factor, and Gender as the between-subjects factor. I also included Attentional Control and Gaming Expertise as covariates. I removed one non-binary participant from the analyses. There was a significant main effect of Gender on Levels Completed ($F_{1,114} = 6.68$, $p = .011$, $\eta_p^2 = .055$), where men completed significantly more levels than women.

Participants were asked about their gender via the prompt: "What is your gender?" with the choices: "Man", "Woman", "Non-binary", or "Prefer not to say". We had 27 men and 10 women complete the no checkpoints, no breaks version, 27 men and 10 women complete the no checkpoints with breaks version, 24 men and 13 women complete the no breaks with checkpoints version, and 23 men and 15 women complete the version with checkpoints and breaks.

E.4.3 Between-Subjects Completion Time

To explore the between-subjects differences in completion time for the training and retention session, I performed an RM-ANCOVA with Levels Completed as the repeated-measures factor, Session (training and retention) as the within-subjects factor, and Checkpoints and Breaks as between-subjects factors. For covariates, the same covariates that were used in the performance measure analyses within the manuscript were used.

There was a main effect of Session on Completion Time ($F_{1,108} = 688$, $p < .001$, $\eta_p^2 = .864$); participants completed fewer levels in the retention session compared to the training session. There was no interaction between Session and Breaks ($F_{1,108} = 1.10$, $p = .296$, $\eta_p^2 = .010$); the efficacy of Breaks did not change between the two sessions. The was a significant interaction between Session and Checkpoints ($F_{1,108} = 74.0$, $p < .001$, $\eta_p^2 = .407$). Post-hoc pairwise comparisons (using Bonferroni corrections) show that during training, participants were able to complete about 3.8 more levels on average with Checkpoints enabled, compared to those who trained without Checkpoints ($p < .001$). If participants trained with Checkpoints, then on the retention session they completed 6.6 fewer routes on average, compared to the training session ($p < .001$). If they trained without checkpoints, then they completed only 2.4 fewer routes than during training ($p < .001$) (but they also performed worse during training). Regardless of whether someone trained with or without Checkpoints made little difference once checkpoints were removed (average difference of 0.4 levels completed, $p > .999$).