Here Come the Test Results: An Analysis of Games in Experimental Design for Neuroscience Research

Madison McDougall

Follow this and additional works at: https://scholarship.claremont.edu/scripps_theses

Part of the Computational Neuroscience Commons

Recommended Citation
McDougall, Madison, "Here Come the Test Results: An Analysis of Games in Experimental Design for Neuroscience Research" (2024). Scripps Senior Theses. 2276. https://scholarship.claremont.edu/scripps_theses/2276

This Open Access Senior Thesis is brought to you for free and open access by the Scripps Student Scholarship at Scholarship @ Claremont. It has been accepted for inclusion in Scripps Senior Theses by an authorized administrator of Scholarship @ Claremont. For more information, please contact scholarship@claremont.edu.
Here Come the Test Results: An Analysis of Games in Experimental Design for Neuroscience Research

A Thesis Presented

by

Madison McDougall

To the Keck Science Department

of

Claremont McKenna, Scripps, and Pitzer Colleges

In Partial Fulfillment of

The Degree of Bachelor of Arts

Senior Thesis in Neuroscience

December 4, 2023
# Table of Contents

Abstract..............................................................................................................................................2

Introduction......................................................................................................................................2

Experimental Design in Cognitive Sciences..................................................................................4

Game Design as Experimental Design..........................................................................................9

How Games Are Being Used........................................................................................................18

Games as Treatments...................................................................................................................19

Big Data of Games........................................................................................................................23

Computers Playing Games...........................................................................................................28

How Games Should Be Used........................................................................................................33

When to Use a Game....................................................................................................................35

Self-made Games vs. Existing Games.........................................................................................36

Developing Your Own Game........................................................................................................37

Conclusion.......................................................................................................................................41

Acknowledgements.......................................................................................................................43

Literature Cited..............................................................................................................................44
Abstract

A large part of experimental design in neuroscience revolves around the tasks that are designed for participants to complete to understand the complex system that is the brain. However, as Alan Newell pointed out in his 1973 paper, “You can’t play 20 questions with nature and win,” as designed tasks often answer a binary question that limits their contribution to understanding the brain as a system or a “genuine slab of human behavior.” Games provide an apt solution to this problem, as their complexity embodies a naturalistic task while their well-defined rules create a robust experimental system. Video games specifically have been growing in popularity for cognitive research, and I have defined three different ways in which they are used: as treatments, as Big Data sources, and as sites for training computational models. Their greatest strengths lie in the accessibility, quality, and quantity of their data, and their manageable task complexity. However, most current studies fail to take advantage of the opportunities games provide for research, often due to a lack of understanding of the design choices made that contribute to the player experience. I outlined some principles for using games in neuroscience research, including deciding whether to use a commercially available game, a clone game, or a self-made game to produce meaningful data collection. When opting to create one’s own game, I recommended working with a game designer to gather an optimal quantity of data without impairing the quality of the composite task.

Introduction

With the rise of increased computational capabilities, modern neuroscientific research has a greater capacity to model and analyze brain data than ever before. However, experimental methodology has not evolved at the same rate as computational power. New experimental designs are required to facilitate further understanding of the brain and the behaviors it elicits.
Games such as chess have made for versatile research mediums in academia for over a hundred years, since the birth of the cognitive sciences, due to the complexity and multitude of tasks contained within their well-bounded environments. Modern developments in video games have made games more accessible than ever, with massive amounts of data generated by players in online spaces. The nature of the tasks in games and the structure and availability of their data make video games an ideal tool for neuroscientific research.

However, although many of the benefits of using games for research are beginning to be acknowledged, the majority of studies fail to take full advantage of what they have to offer. Using games implements a top-down approach to neuroscience research, a perspective that the field lacks, impeding the development of complete theories of cognition (or aspects thereof). To create the best complete processing models, models must remove vague descriptions in favor of computational relationships that are implementable and testable, and this necessitates controlled and parametrizable environments, an aspect that many studies lack. Additionally, many studies overlook the design choices involved in creating a game, and their studies using games often remove or stifle these choices, obstructing the purpose and value of the game. Using games in research is not something that can be done haphazardly or without knowledge of the medium you are using to conduct your research.

In this paper, I will outline common problems with current experimental design for cognitive and behavioral science and the solutions to incorporate in future research as outlined by Alan Newell, a computer scientist and cognitive psychological researcher whose teachings still influence the field today. I will then describe how video games are an optimal method for achieving Newell’s goals and illustrate to the reader the parallels between experimental design and game design. Then I will explain three major ways in which games are currently used for
cognitive research: as treatments, as data sources, and as sites for testing computational models. Finally, I create my own criteria for how the usage of games in research could be improved and my own analysis for how they can best be used.

**Experimental Design in Cognitive Sciences**

In 1973, Alan Newell identified several problems with the current state of cognitive and behavioral research (Newell 1973). His critiques were expressed after he was asked to review them at a symposium, where he identified a rising stagnation in the progress cognitive sciences were making. The ultimate goal in cognitive science is to understand why humans act like we do—whether it be understanding the mind through psychology or understanding the brain through neuroscience. Newell noticed that many of the empirical standards of his time were not efficient in pursuing these goals, and his critiques can be eloquently summed up by the same of his paper: “You can’t play 20 questions with nature and win.”

Newell’s problems with his colleagues’ work stemmed not from inaccuracies or malpractices in their research but rather from the scope of their studies and questions. Reflecting upon the work in the symposium, he was unimpressed by the rate of progress in the field. Psychological research at the time was driven by the investigation of phenomena, and dozens of papers might be written on each individual phenomenon, investigating the minutia of its variations. But what did this research actually contribute towards the greater goals of the field? Newell concluded, unfortunately, that the answer was *not enough*. Summing up all the hard, detailed work of his peers, it seemed as though any applicable understanding of the brain was not achievable for decades. The main limitation of this proper method of study was its reliance on the “construction of oppositions—usually binary ones,” which resulted from the procedures for empirical design at the time (Newell 1973). Neat, scholarly research encouraged a clean question
that could be tackled experimentally, receiving feedback at every step to minimize the risk of wasting one’s limited time and resources. While binary questions are relatively simple to formulate and test, to get tangible results and reveal one aspect of a phenomenon’s nature, they fail to cumulate to the brain entirely. As most students realize within the first week of a neuroscience course, the brain is more than the sum of its parts—so how could we hope to understand it by only looking at these parts discretely, with very little focus on the intricacies of how they interconnect to form a system as complex as the brain? In attempting to manage the scope of empirical research to investigations that were achievable and efficient, the scientific community essentially doomed itself to never making any significant process in understanding how the brain worked as a whole.

Newell (1973) determines this lack of progress is due to the clarity of the system never being achieved through determinations about these binary oppositions, creating more avenues of investigation but no determinant answers to the questions asked. While it is not impossible that every minute question could eventually be answered, with the rate of questions being generated accelerating past the rate of questions being conclusively answered, that point seems far in the distant future. In fact, just as Newell predicted, fifty years down the line we are still on the level of investigating phenomena without substantial theories that can describe large parts of cognition completely or succinctly. So while the debate rages on as to whether investigating phenomena will ever cumulate into significant, it is perhaps more accurate to say that the current scientific research paradigms for cognitive neuroscience are not efficient in making significant progress toward understanding the brain.

So, what was to be done with the psychological fields? Thankfully for academics everywhere, Newell did not claim that the field was useless and all his colleagues’ research was
unsalvageable. Instead, he proposed a new paradigm for cognitive research that could be implemented in experimental research to develop robust and applicable theories of the mind. This paradigm emphasized implementing top-down frameworks as well as the typical bottom-up mechanisms in behavioral research, encouraging researchers to include “how?” and “why?” elements to their research questions rather than exclusively “what?” The first pillar of this new paradigm was regarding methods: “Know the method your subject is using to perform the task” and, with this in mind, “never average over methods” (Newell 1973). In essence, this pillar requires a researcher to put their data in a greater context—it is only useful to know the minutiae of a particular phenomenon if you have some idea of how those phenomena are being put together. Or, for the analogy-inclined, the ingredients list of a recipe is only useful if one has some knowledge of the process of combining those ingredients. The latter part of this pillar is a little more specific to a particular experiment. If you know that your participants are using different methods, do not include their data in the same averages. To go back to my previous analogy, if you have two bakers with the same list of ingredients but making two different dishes, do not compare the ratios of their ingredients because the final product is meant to be different. This pillar encourages researchers to acknowledge the greater framework of the mechanism they are investigating and put their findings in a wider context to show why their findings matter.

Newell’s next instruction is to “construct a complete processing model” (Newell 1973). A model is a simplification of a real-world system that serves as “a quantification of a hypothesis to investigate the hypothesis” (Trappenberg 2009). Model-based research is able to make scientific communication more precise while establishing causal relationships to form behavioral predictions from neural data (Blohm et al. 2020). This pillar provides two solutions to problems within the current typical empirical design: first, emphasizing the need to consider small-scale
phenomena within the context of the greater system; second, reducing unspecificity and unconstraint in existing partial models to create more actionable, definite models. Although perhaps an ambitious expectation in his time, computational models are far more practical with modern hardware and are able to provide and implement complex relationships similar to neural connections in the brain. Additionally, robust models may “provide a basis for identifying the system experimentally if a sufficiently large and diverse set of tasks is analyzed” (Newell 1973).

If one can define a model for a task, aspects of the original task may be shared by another unfamiliar task to which the same model can apply. More universally applicable models would be able to provide more information about general methods of the brain.

The final and arguably most important pillar of Newell’s paradigm is to “accept a single complex task and do all of it” (Newell 1973). This pillar addresses several problems Newell identified in cognitive research at the time. This instruction emphasizes placing phenomena in the context of the goals and methods of the task they contribute to. It removes the uncertainty that comes from stripping a task so much that all content and goals of the task are removed, which fails to preserve insights from the experiment in naturalistic behavior. Additionally, it ensures that whatever insights a researcher gleans significantly contribute toward a general understanding of human behavior. The goal of this pillar is to “demonstrate that one has a sufficient theory of a genuine slab of human behavior” to prove that a researcher has information that is valuable beyond the scope of the laboratory (Newell 1973).

The concerns Alan Newell expressed in 1973 revealed a potential crisis in the state of cognitive and behavioral research at the time. His proposed treatment for this diagnosis was a new paradigm for empirical research that focused on placing specific phenomena within the wider aspect of human (or animal) behavior that aimed to address the stagnation in new
cognitive insight. He encouraged researchers to develop a broader theory to go along with their empirical evidence, a more time-intensive and academically risky expectation that he deemed necessary for the relevance of the field. His insights are still cited today as a strong framework for cognitive research, and Newell continued to shape the field by implementing his methodology with his work over the next two decades.

But despite the acclaim for Newell’s paradigms, many studies still fail to meet his criteria, even fifty years later. Rather than focusing on understanding cognition as a whole, many studies have scope only marginally broader than those Newell critiqued, focusing on a task such as reading, basic motor control, or language formation and comprehension. Despite the acclaim Newell received, it seems that current experimental ideologies still favor relatively small-scope experiments accompanied by small-scope, precise models that are then tied into vague models that lack quantifiable relationships rather than the robustly defined yet generalizable models and theories Newell championed (Gobet 2017). This is not entirely unexpected, as Newell had extremely ambitious expectations for the field since researchers often overly rely on their data to generate a theory rather than using a theory to guide avenues of research. Finally, because academia may impede the scientific process, it is often “easier to generate papers based on experiments rather than on computational theories” (Gobet 2017).

In 2017, Fernand Gobet was asked to conduct a review of the papers in *Topics in Cognitive Science* through Newell’s lens, a process similar to what Newell himself did in 1973. Indeed, the current replication crisis harkens back to the slow rate of progress despite the flurries of papers that Newell pointed out (Gobet 2017). His findings confirmed that current studies did not adequately meet the expectations set by Newell in 1973, though each of the eight papers fell short in their own way. Though Gobet noted that half of the studies incorporated strategy
identification in their methodology, it was too often the sole goal of their study rather than an element to be considered. Overall, there was a shortage of computer modeling, and the primary sin was failing to compare the computational conclusions against human data to test its accuracy. Investigation of a task as a whole was also missing, though this can be attributed to this particular criteria befitting a research method rather than the scope of a single paper. Additionally, many of the papers focused too much on how the data were acquired and transformed, and minimally on their analysis and the conclusions drawn.

However, the unifying strength of these papers, often the source of their merits regarding Newell’s paradigm, was that they often used data from video game play. Online video games often record massive data of players’ activity and sometimes behavior, the scale of which benefits computational models. Video games are also inherently complex tasks, integrating motor processes, vision comprehension, and cognitive tasks all in one. The main flaw of these studies was their inability to fully take advantage of any of the multitude of facets of video games—a product of their complex manufactured interactions—that lend themselves to the kind of ambitious experimental design Newell favored.

**Game Design as Experimental Design**

Newell’s main critique revolved around current experimental design's failing to provide answers to meaningful questions, for their lack of ability to capture “genuine slab[s] of human behavior” (Newell 1973). It is understandable that researchers might fail to consistently meet Newell’s criteria due to the complexity of an experiment that would not only be able to capture human behavior on this scale but also measure it with enough detail at each level of analysis. Designing experiments requires a great level of thought and consideration that has been captured by innumerable papers on experimental design and assay and critiques thereof. The incorporation
of intelligent beings (usually humans) only increases the complexity due to the confounding variables that individual people introduce that make their behavior much more difficult to predict and control. Many behavioral experiments involving humans center on shaping the experience of the individual to extract a certain behavior from them, the components of which are then measured (Donchin 1995).

This intent closely mirrors, of all things, the practice of game design. In fact, it is not a far stretch to compare a lot of behavioral or cognitive tasks to games. Though there exist as many definitions for games as there are games, in the context of this paper, I shall use Salen and Zimmerman’s definition from Rules of Play: “A game is a system in which players engage in artificial conflict, defined by rules, that results in a quantifiable outcome” (Salen and Zimmerman 2003). In my comparison to an experiment, the “conflict” here is the task that the subject is being asked to perform, where the “rules” are the parameters as assigned by the researchers, and the “quantifiable outcome” is the data collected by the researchers. Subjects may be asked to compare sounds, remember a sequence of words, or track movement with their eyes, and then their performance is scored by whatever criteria the researchers have selected. This practice is strikingly similar to how one might be introduced to a video game, where one is given controls and rules and then given a goal, such as solving puzzles to beat a level, or collecting coins to achieve a high score, or surviving as long as possible against waves of enemies. Each of these experiences is not only defined by a clear set of parameters like an experiment may be, but an equally intensive amount of theory may go into how these parameters are selected and defined based on how they influence the player’s experience. Pulling from Rules of Play again, game design may be defined as: “the process by which a game designer creates a game, to be encountered by a player, from which meaningful play emerges” (Salen and
Zimmerman 2003). Play here is defined as an activity that is voluntary, limited, uncertain, unproductive, \textit{governed by rules}, and make-believe, or perhaps more comprehensibly as “free movement within a more rigid structure” (Caillois and Barash 2001; Salen and Zimmerman 2003). The key components of play and games are the presence of bounding rules and player choice so that a player has freedom to act but not infinite freedom. Although game design is less rigorously academically explored than experimental design, it is no less sophisticated, often taking years or decades of practice and experience to be consistently well implemented. Indeed, just as researchers still stumble with their experimental designs as suggested previously, game designers continue to stumble too, when the parameters and mechanics they implement fail to result in the desired player behavior (typically when mechanics are poorly understood by the player or inconsistently implemented by the designer) or player experience (usually one of entertainment or achievement, though relaxation or grueling agony are occasionally desired).

To further elucidate my point, I present a short example of phenomenal game design that precisely and elegantly shapes the player’s experience in a tutorial level. Tutorial levels are one of the most variable implementations in game design, as they require imparting the rules of the game in a clear yet simplistic manner that does not interfere with the naturalistic feel of a game and preserves its entertainment factor. Many games unfortunately still favor large walls of text that very explicitly state the precise controls and mechanics all at once which can be quite difficult to parse and unnatural to implement. This might be similar to experiments where human participants are given a long explanation of the task by the researcher and must then quickly figure out how to follow those instructions in the experiment. While dumping large amounts of information through text may be ideal for some games like detailed strategy games, it is disparate from how most people learn things in a natural environment. Therefore, its inclusion in
experimental design can impact participant behavior and the data collected from it, just as it disrupts the organic feel of gameplay. Regardless, a good tutorial level imparts upon the player the tools and understanding they need to play the game in a straightforward and unintrusive manner, a practice which is implemented superbly in the opening level of the 1994 SNES game *Super Metroid*.

*Super Metroid* is an action-adventure game in the *Metroid* series in which the player explores a large area, overcoming obstacles such as enemies and platforming (traversal of platformed spaces often requiring precise movement) challenges. The basic controls are therefore movement controls (moving left and right and jumping) and combat controls (aiming and firing a laser beam), though more complex mechanics and their controls are slowly introduced throughout the course of the game. For now, I will stick to how these basic controls are introduced to the player through the tutorial.
Following a brief cinematic that introduces the player to the narrative setting, Samus (the player character) spawns in on a floating platform (Fig. 1a). Below them to the left and right are other platforms, and the area is only bounded on the far left and right of the screen. Instinctively, most players would then try pressing the left or right buttons, which will result in them falling to a lower platform. This channel then alternates between platforms in the center and on the edges of the channel, and the player will alternate moving left and right to make their way down the
Thus the basics of horizontal movement have been imparted to the player in a few short seconds with no direct instruction. At the bottom of the channel, there is a protrusion from the right side, but the player’s movement is slightly impeded by two small steps before they approach (Fig. 1c). These steps are no more than a third of Samus’ height, which implies that they should not be insurmountable. A brief experiment with buttons as well as an assessment of possible actions will result in the player discovering Samus’ ability to jump, and they will likely then make their way toward the protrusion on the wall. As the player nears the protrusion, it reacts, folding into the wall with a slight warp sound, providing feedback to the player that suggests an interaction with the environment, encouraging them to proceed towards where the protrusion was (Fig. 1c). It is, of course, a door, and the player proceeds to the next module of the level. This has introduced more basic movement controls, as well as the setup for the design of the world. There are discrete chambers that the player will move through at their own pace through clearly marked entrances.

The next chamber starts with a short flight of stairs and another small step that the player must jump up, reinforcing the previously learned jump mechanic. Proceeding through another door at the end of the chamber, the player enters a new chamber that contains an S-shaped path, requiring the player to move right and left, falling between layers as they continue to get a feel for the movement in the game (Fig. 1d). The next chamber can be moved through straightforwardly, but then we pass by what looks like some fallen scientists, implying that something dangerous has been in the area—it is likely the player will encounter some combat soon (Fig. 1e). After passing through another straightforward passageway, this time unadorned with corpses, we enter another chamber that contains a small moving creature (Fig. 1f). This should immediately spark interest, as it is the only thing encountered thus far that has moved,
apart from the doors and our player character, each of which has been tied to some interactive element. However, upon approach, there does not seem to be any way to interact with the creature. The few moments of lack of clear progression likely result in more exploration from the player with the controls and possible actions, during which they will likely find the button to fire Samus’ laser beam if they have not already, the only basic action that has not been used thus far. Moments later, a much larger creature fades into view, and the silence is replaced with an uptempo score, signaling that it is time to fight! As the creature begins to move around the screen, the player quickly realizes they must be able to aim their beam to hit the creature and will be rewarded with a short flash over the creature if they are successfully able to hit it with the beam (Fig. 1g). Their newly learned movement skills will be tested as they must dodge the bursts of fire and the lashing tail of this creature, as their Energy stat decreases each time they are hit, introducing this parameter that they must maintain in order to continue playing. Eventually, the creature will disappear, and a self-destruct sequence and timer will pop up onscreen. It is now time to put the player’s movement skills to the test, as they must retraverse the area they moved through earlier before the timer runs out (Fig. 1h). The sequence ends when the player makes it back to the starting platform and successfully escapes the self-destruct in the following cinematic.

This short sequence typically takes no more than a minute or two to play through, and yet the player is taught everything they need to know about how to get started playing the game without a single explicit instruction, akin to how humans might learn other tasks in our life like riding a bike. It elegantly moves through the first two stages of basic game design philosophy—“teach, test, twist”—in which game mechanics are taught to the player and then tested under pressure (and will eventually be twisted by implementing them in new, unique
ways) in a matter of moments, in such a discrete manner that most players will not even notice. And yet, the design choices are there and contribute significantly to the player experience by the player’s ability to overlook their very existence. I hope this example has elucidated not only the implementation but also the sophistication of game design so as to make my comparison to experimental design not seem quite so ludicrous.

Along with being extremely robust forms of experimental design, games, video games, in particular, contain a lot of other facets that facilitate experimental research, particularly within the framework of Newell’s research program. As already addressed, games serve particularly well Newell’s urging to “accept a complex task and do all of it,” as games are a complex task. Chess in particular has been extensively studied to create models for complex decision-making and to understand expertise over the last century (Connors et al. 2011). Studies involving chess drew their strength from both their ability to include concrete data regarding players’ decisions during the match, but also qualitative descriptions of the methods being used by players. This allowed researchers to formulate different models for how experts made decisions during the game compared to novices, and this difference in methods not only aligned with Newell’s recommendations for knowing subjects’ methods but also allowed for a basic understanding of how players’ methods changed as they gained greater knowledge of the task.

As beneficial as an originally analog game such as chess has been, digital games magnify the advantages of games by allowing the designers minute control of the mechanics in the game and through detailed and precise recording of actions taken during the game. The entire experience must be designed and manufactured, and the rules are regulated more consistently than analog games, making digital games a good source of bounded, simplified worlds that are still free enough to examine complex tasks within them. These benefits were recognized as early
as the 1980s with the creation of the first game explicitly for behavioral research, *Space Fortress*. Researchers at the Cognitive Psychological Laboratory at the University of Illinois recognized the potential of a video game for their research and developed their own due to difficulties with utilizing a ready-made game for research at the time (Donchin 1995). To be effective for research, a game (or an experiment) needs to permit the “monitor[ing] of … all variables that might be relevant in describing the way in which a subject executes the task” and “ensure that the raw data are retained well beyond the time of initial analysis,” both of which required access to a game’s source code which vendors were reluctant to part with at the time (Donchin 1995). While the first of these criteria still often requires the creation or modification of a game by the researchers, the second has become increasingly accessible in the modern era, as most games have an online component and details of the session (Gray 2017). These data can range from player activity across long stretches of time to keystrokes taken by individual players during a game session and can be accessed to varying degrees in large quantities. The scale of Big Data facilitates Newell’s goal of using large datasets to detect or test a computational processing model on real data and allows for the reconstruction and retention of game data post-experiment that Donchin et al. emphasized while developing *Space Fortress*. Additionally, these data can allow one to track player performance over timescales in the range of months or years rather than the minutes to hours typical of laboratory experiments (Huang et al. 2017). This provides potential for detailed longitudinal studies within individuals that contain data from thousands of individuals.

Another benefit of using games for research is the fact that games are intrinsically motivating. Unlike typical laboratory assays where participants are typically compensated for their time with extrinsic motivations such as course credit for students or monetary rewards,
games are designed to be enjoyable, and so people often choose to engage with them without extrinsic motivation (Howes 2017). This has the potential to increase the scale of their studies by attracting a larger participant pool and increasing retention for laboratory longitudinal studies that may require tens of hours in the lab (Gray 2017). Researchers may also increase the size of their study sample by committing fewer financial resources to subject compensation, recruiting more participants through their enjoyment of the activity, and increasing the engagement of participants due to their interest in the task being studied—though there are potential ethical concerns regarding the potential for labor exploitation by minimizing participant compensation. Beyond laboratory experiments, the enjoyment of games contributes to the large amount of data from online games played by self-identified gamers (whose competitive online matches result in match data being stored for self-comparison and self-analysis) as well as the general public (whose primary interaction with games is likely through their smartphone, the limiting processing power requiring much of the data to be hosted in servers). The massive draw of games, the swathes of data they can generate, and their inherent ability to capture a complex yet manageably scaled task make them the ideal vector for studying human behavior.

**How Games Are Being Used**

Games have been used in cognitive behavioral research for at least a century, but the specific ways they are used have evolved and developed variations over time. To facilitate my analysis, I have identified three ways games are typically used to study behavior or cognitive processes. The first uses a specific game as a treatment— they presume that a game is or could be capable of developing a certain skill or otherwise affecting the player and seek to measure that effect. The more ambitious of this variety may inquire as to why or how the game accomplishes this, but often this binary detection alone is enough for these studies. The second type involves
analyzing the Big Data generated by online games to develop or test theories of cognition. The final type involves creating a computational model to play a game and then comparing the model to real human behavior. Each of these strategies has certain benefits and potential pitfalls that indicate their success in adhering to Newell’s guidelines for experimental design in cognitive research.

Games as Treatments

Using a game as an experimental treatment implies that the practice of developing a certain skill or process can be embedded into an entertaining game. This is an example of gamification, which is here defined as: a process that aims to increase the intrinsic and extrinsic motivation of an activity through the implementation of game design elements (Manzano-Leon et al. 2021; Mazarakis and Bräuer 2023). These elements include mechanics, dynamics, and aesthetics and have been shown to increase engagement with an otherwise uninteresting or laborious task (Manzano-Leon et al. 2021). Understandably, gamification has been implemented particularly within educational environments from elementary through university levels with marked success (Mazarakis and Bräuer 2023). Gamification has also been utilized to promote fitness and healthy lifestyles in games such as *Wii Fit* (a console game that facilitates and monitors workout routines) or *Beat Saber* (a virtual-reality rhythm game that facilitates continuous physical exercise) and in remote geriatric care to increase the consistency of and reduce the barrier towards physical therapy. Even step-tracking apps in many smartphones and smartwatches are forms of gamification in everyday life, increasing the motivation for these difficult tasks with progress bars or badges or check marks for completion.

Treating a game as an experimental treatment therefore expands upon these ideas by embedding the potential to increase or retrieve a skill within an entertaining game. *Tetris* is one
game that has been studied extensively for treatment effects, such as spatial skills, relief for flashback trauma, or engineering skills (Sibert et al. 2017). But there is also a growing market for games that are designed to improve cognitive abilities, so-called “brain-training” games (Simons et al. 2016). The efficacy of these games is a subject of much debate that is beyond the scope of this paper, but regardless there is a continued interest in the ability of games to improve skills and cognition.

The best studies that investigate games as a treatment typically dive deeper into this concept than those that simply detect a significant treatment effect. Instead of posing binary questions akin to those Newell despised, studies might investigate how games are able to achieve improvement in cognitive abilities and specifically what aspects they can affect. One such study focused on investigating the benefits of playing Portal 2, a 3D puzzle-solving and platformer game that is widely esteemed in the gaming community for its innovative puzzle design and engaging narrative (Shute et al. 2015). This study used a popular brain-training app, Lumosity, as its control, a conservative choice designed to minimize the possibility of overstating the effects of Portal 2. Researchers randomly assigned participants to a particular game and then measured their performance with various cognitive tasks before and after completing a certain threshold of game experience.
Portal 2 is a puzzle-solving game whose central mechanic is the ability to travel between two linked portals that the player develops the ability to place and move around the environment. The challenges are sorted into discrete levels (called “tests” as the game fittingly takes place within a research laboratory) and the player must use these portals and resulting various quirks of physics (Fig. 2) to overcome obstacles and traverse from the entrance of the level to the exit. Owing to the nature of the game, the researchers measured the player’s development of three skills: problem-solving, spatial skills, and persistence. Problem-solving speaks for itself due to the puzzle-based elements of the game. Spatial skills are required for the placement of portals and the prediction of how they will aid the player’s traversal of the level. Persistence reflects the “principle of challenge” in game design which requires that a task be difficult enough that the reward of completion feels valuable but not so difficult as to discourage the player, so its growth would indicate a “desire to exhibit high standards of performance in the face of frustration” (Shute et al. 2015).
"Lumosity" is an app that employs gamification techniques to engage users in several brain-training exercises. It includes 52 different 2D games that each aim to improve a variety of abilities including memory, attention, processing speed, mental flexibility, spatial orientation, logical reasoning, and problem-solving skills (Shute et al. 2015). It explicitly claims to be able to improve these abilities and provides feedback for users on their progress, while "Portal 2" may seemingly improve them simply as a byproduct of the gameplay experience.

Interestingly, participants assigned to play "Portal 2" outperformed those engaging with "Lumosity" in every measure investigated. The results of this study not only embolden the argument that video games are not a degenerative waste of time but might actually improve the cognitive skills of the player (in some cases), but also participate in the investigation of the efficacy of brain-training programs. The applications to a wider narrative and discussion are what make this games-as-treatment study particularly noteworthy, although it falls short of significant contributions to cognitive theories or many of the other goals proposed by Newell.

Some of the drawbacks of this study that persist throughout many other studies that use games as treatment include its reliance on laboratory sessions. Since the actual data of this study is acquired through external measures conducted before and after play sessions, it evades the advantages of large-scale, remote data collection. Additionally, it fails to experiment with the mechanics within the game that lead to the measured skill gain, which would require a more involved procedure including modifying the game levels, instead treating the game as a sort of "black box" whose contents remain inscrutable (Gray 2017). But its greatest weakness is its reliance on the construction of a binary opposition—can video games increase cognitive skills?—though, to its credit, it does expand on the nature of the binary to the best of its ability. This simple question does not allow for an exploration into how the game achieves this skill.
increase or fails to, unnecessarily limiting what can be learned about human cognition from their research. Ultimately, the limitations of this type of study do not permit it to take advantage of all that games have to offer, failing to acknowledge or investigate the elements of a game that lead to a significant treatment effect and, by extension, the cognitive processes that are affected.

*Big Data of Games*

The abundance of online games has led to new sources of naturally occurring datasets on the scale of Big Data. Studies that utilize this resource employ many of the strategies that are common in Big Data but also have the capability to anticipate human behavior, and potentially make inferences about underlying biological processes. One of the major advantages of online game data is that it tracks player activity over large swaths of time, often in the scale of months or years depending on the game and player. This makes it an ideal candidate for longitudinal research that tracks changes in an individual’s behavior over a long period of time. These timescales allow researchers to model longer cognitive processes and developments, one of the most common being skill acquisition and expertise gain, similar to de Groot’s work with chess (Stafford and Dewar 2014; Connors et al. 2011). However, instead of comparing experts and novices, online game data allows one to track an individual’s development as they progress from novice to intermediate to expert, aiding the development of complex models of the process.

This ability to acquire longitudinal data over these extended time frames once again relies on the player’s continued interest in the game—it is very difficult for someone to repeat or practice an activity if they have no motivation to do so, and games provide that motivation inherently (Boot et al. 2017). It is also one of the only sources of data that equally tracks novice and expert information—the quality or quantity of the data regarding a game session or an individual's play activity does not vary with their skill level.
Difficulties do arise when it comes to accessing this data. Although a lot of data is collected by the game developers, typically for diagnostic measures, and portions of players’ individual data may be available to them, they are not typically publicly available. Developers are often reluctant to provide access to private data or are legally obstructed from doing so due to valid concerns about players’ privacy. One method for overcoming this is having data collection specifically built into the game so that it can be sent directly to the researchers, but this would require an explicit statement in the game description and often requires the development of a game specifically for this purpose rather than retroactive inclusion.

Axon is a game developed by PRELOADED on behalf of the Wellcome Trust specifically for this purpose. By implementing data collection directly into the game code, data collection is unobtrusive to the player and anonymized to be ideal for research (Stafford and Haasnoot 2017). The game was designed to be accessible to a broad audience, to provide opportunities for players to learn a little bit about neurons and neuroscience, and to collect data for analysis by researchers (Stafford and Haasnoot 2017). Gameplay centers around growing the axon of a neuron by rapidly clicking a series of nodes within a range that gradually decreases, decaying more quickly as the play session continues. This simplistic design makes the game quick to pick up and comprehensible to a wide audience. This design choice accompanied by the appeals to citizen science by marketing the game as a research tool based on neuroscience principles positioned the game to reach a wide audience of players despite its niche standing.
Fig 3. Adapted from Stafford and Haasnoot 2017. Improvements in performance with practice for those who do not take breaks (“no gap”) and those who have long breaks, either overnight (“sleep”) or during the daytime (“wake”). Uncategorized players not shown. Black line shows median for all players and 95% (dashed line with large dots) and 98% (dashed line with small dots) confidence limits based on samples the size of the smallest of “no gap,” “sleep,” and “wake.”

Axon data have been utilized by a variety of studies, one of which used them to test sleep consolidation’s effect on skill acquisition over long time periods (Stafford and Haasnoot 2017). Much of the work in this study surrounded the transformation of Big Data into something usable and analyzable, mostly by using timezones and gaps in play to estimate when sleep was occurring between play sessions. These improvements could then be compared to gaps in play where no sleep was estimated to have taken place and sessions of prolonged continuous play. These naturally generated game data were uniquely insightful as sleep consolidation is particularly difficult to recreate naturalistically in a lab setting (Stafford and Haasnoot 2017). While the study did provide starting places for understanding why some people may gain a certain skill much faster than others, it failed to recreate the benefits of sleep consolidation outside of the laboratory, which have been supported by various studies over the years (Fig 3). Though this could possibly have arisen due to flaws in their method of data analysis, two more
interesting possibilities exist. One is that the lack of experimental control interfered with the ability to detect sleep consolidation. This lack of control is one of the principal detractors of using data collected outside a laboratory setting, as it is impossible to know what activities a participant engages in beyond the game sessions. The other devalues the game itself—that the gameplay task was too simple to benefit from sleep consolidation due to the focus of developers on *developing a tool for research rather than a game* (Stafford and Haasnoot 2017). The game itself was not well-designed as the choices the player could make within the game system were too limited, making the game too simple and not requiring complex learning to play or improve at the game. This neglect of the spirit of game design is unfortunately often present in games made for research, as I will continue to explore in subsequent sections. However, the strengths of this particular study lie in the researchers’ ability to predict human behavior over data from over a million players and make inferences about the underlying cognitive processes.

---

**Fig. 4.** Adapted from Thompson et al. 2017. The components of a perception action cycle (PAC). The whole PAC begins and ends with an attention shift. Its duration encompasses First-Action-Latency, Inter-Action-Latencies (multiplied by Action-Count: the number of actions, n), and the Latency-to-End.
Fig. 5. Adapted from Thompson et al. 2017. Parallels between eye tracking and StarCraft 2 fixation data reflected in a schematic of eye (top row) and screen (bottom row) fixation detection. Raw gaze/screen co-ordinates (left, orange marks) indicate where an observer is looking in two-dimensional space.

An alternative to developing one’s own game is sourcing data from an existing game. While difficult for reasons already stated, it is not out of reach. Calls to citizen science are once again effective in crowdsourcing data, as one study was able to collect replay data (a log of all actions taken during a game) submitted by over three thousand players by posting on various game forums for StarCraft 2, a 2010 real-time strategy game with a robust community (Thompson et al. 2017). Rather than using general play activity data like the Axon study did, this study used high-resolution data from individual games to infer motor chunking and action latencies in a complex cognitive-motor task (StarCraft 2). Their analysis aimed to identify perception-action cycles (PACs) and the action latencies within by taking advantage of the necessity to open windows to view different information and perform certain actions during gameplay (Fig 4). These PACs were “analogous to the gross saccade-fixation-saccade pattern of eye movements,” and new map windows are accessed “often on the same scale as eye
movements themselves” (Fig 5). These findings are particularly hopeful as they are comparable to human behavior, emphasizing their generalizability as well as adherence to Newell’s encouragement to compare model data to human behavior. It also pertains to Newell’s requirement to “accept a complex task and do all of it,” as this is one level of analysis in a series of studies with the same data relating to the greater complex task of playing the game. Additionally, the researchers propose that their strategy for tracking attention shifts can be used to identify action sequences in other sources of data, in line with Newell’s conditions for creating computer models that can be generalized across multiple tasks.

Each of these studies takes advantage of the popularity of games to maximize the amount of data they access, often several orders of magnitude larger than that acquired in the lab, but they differ in the level of analysis they investigate, from gameplay patterns over long time periods to in-game actions within a single match. They also vary in the specificity of the models created, with the latter study creating a much more computationally actualizable model than the former. They each aimed to encompass “genuine slabs of human behavior,” but arguably fell short of a complete understanding of one. Where these types of studies fall short is typically in their ability to compare to live human data, and to understand the strategies that are undertaken by the player. These require some laboratory experiments and data, and both of these studies could be augmented by gathering some data by interviewing players and having them play matches within the laboratory environment.

*Computers Playing Games*

Studies in this final category contribute chiefly to fulfilling Newell’s expectation of creating a complete processing model for complex tasks by implementing some form of machine learning to play games. Games are particularly useful in this regard as sufficiently complex, yet
still computationally manageable, tasks that incorporate several smaller tasks such as visual processing, motor skills, decision making, and problem-solving involving developing an understanding of the parameters of the task, providing ample resources to develop computational models that “do all of [a complex task]” (Newell 1973). These computational models are particularly insightful when they can be compared to humans completing the task or when their methods are comparable to how people describe their own strategies for the task. Though training computers to play games is not a new phenomenon, recent implementations have provided uniquely interesting insights by more closely scrutinizing how the computer’s performance compares to human behavior.

Although developing a computational model does not necessitate developing one’s own game, some games (particularly newer ones) are complex enough that creating a simplified clone of a game is not unreasonable for this purpose. One study was intrigued by the interpersonal collaboration required by the game *Overcooked* and picked it as a medium to study how individuals anticipate a collaborator’s behavior to coordinate their actions toward a common goal (Wang et al. 2020). *Overcooked* is a relatively fast-paced collaborative multiplayer game where players work together to prepare meals in a kitchen where obstacles may prevent or impede an individual’s ability to complete the entirety of the task on their own. Thus, players must be able to coordinate their behavior to prepare the meal together, often by verbal communication but chiefly by being able to intuitively divide the task amongst themselves and continually adapt to each other’s needs.
The researchers identified a complex task in *Overcooked* that is an ideal choice for fulfilling Newell’s paradigms. Multi-agent coordination is a task that can be broken down into many smaller tasks at multiple levels of analysis that can be examined in concert, such as “coordinat[ing] both their high-level plans (e.g., what sub-task they should work on) and their low-level actions (e.g., avoiding collisions)” (Wang et al. 2020). They contribute to a greater theory of mind with their development of Bayesian Delegation, a process that “enables agents to rapidly infer the sub-tasks of other agents” using Bayesian inference (Wang et al. 2020). Along with their full model that implemented Bayesian Delegation and joint planning, they created three variants of the model that lesioned each or both of these components to rigorously test their model (Fig 6a). Most importantly, they also brought in human subjects to complete the same tasks as the model they had prepared to compare how humans behaved differently or the same as their model to test for the presence of their developed mechanism (Fig 6b). Although the...
research team admits that thus far their model has only been successful in these relatively simple scenarios, they hope to be able to expand their model to longer, more complex tasks.

Along with this study’s proximity to Newell’s ideal experimental design, it also benefits by acknowledging the value and design in the game utilized. The researchers went beyond just finding a game that was simple to modify or easy to collect data from but were able to identify and deconstruct the game mechanics that contribute to the player experience and the complex task that is accomplished. They identified some of the key design elements of *Overcooked* gameplay: level design, multi-linear tasks, and player coordination, and were then able to manipulate them to test certain aspects of their model. The level design element of the barrier between one player and the other, or lack thereof, is an obstacle that limits player choice to varying degrees (Fig 6a), which affects the possible actions that each player can take and simplifies the inference a player makes regarding the other player’s task. The researchers also deconstructed the recipe tasks into ones that differently combined multiple linear tasks (Tomato, Tomato + Lettuce, and Salad), which may be performed separately but eventually have a limiting combining step. This methodology has an advantage over simply running their model on different *Overcooked* levels and noticing that it performed differently across levels, with no understanding as to why the model’s performance varied. Their understanding of and respect for the game augmented their research and allowed them to dive deeper into the questions they could ask and answer.

Computer models also have the potential to tie directly into real neural data. Researchers at CalTech developed a Deep-Q-Network (DQN) capable of playing three old-school Atari games: *Pong, Enduro*, and *Space Invaders* (Cross et al. 2021). Unfortunately, the specifics of these games are not particularly relevant to this study; however, the researchers were able to
integrate fMRI data from human subjects playing the same games to generate predictions from the DQN. The final outputs were able to consistently accurately predict the voxel responses across several brain regions, and intermediate layers of the network could be used to predict choices at key decision points during gameplay. Although this study does not take advantage of many assets of video games, it does connect gameplay analysis to neurological data and construct a complete computational processing model capable of predicting human behavior to some degree, attributes that could be applied to other studies to bolster their significance.

When done well, computational models for video games have the greatest potential to benefit from everything games have to offer and to fall in line with Newell’s paradigms for cognitive research. By nature, they work towards generating a complex processing model to capture a complex task like game playing, but they also have the potential to fulfill other aspects. A player’s chosen methods for a task might be inferred by constructing several models capable of capturing different playstyles and comparing them to the actions taken by the player. Computational models can also connect to human data to make generalizations about human cognitive processes and even have the potential to take advantage of Big Data from games to compare how computer models improve at a game compared to human players, though neither of the ones examined here are able to do so. Of each of the three methods I have outlined for using video games in cognitive research, this method of creating computational models capable of playing games tends to be the most conducive to insightful research in terms of the quality of data gained, by balancing task complexity and parameter control. The data have the least uncertainty of the three methods outlined and can be examined at multiple levels of analysis to provide a high quantity of information without sacrificing the quality.
How Games Should Be Used

Despite being used for behavioral research for over a hundred years, games are still greatly undervalued by cognitive researchers and neuroscientists, especially considering advantages like accessibility and online data collection provided by modern video games. Games are an excellent medium for behavioral research as they embody complex tasks, containing a wide array of cognitive skills, motor skills, and communication skills. There are a multitude of games that contain different types of tasks and facilitate different skills, so the options for areas of study are nearly endless. Games and their data are also becoming increasingly accessible, becoming playable at home or on a smartphone, and with detailed game metrics uploaded to cloud servers for easy consolidation. Their enjoyment (which often coincides with a well-designed game) also results in a large, engaged player base capable of generating huge masses of data. Additionally, having access to the source code of a game allows researchers to fine-tune the parameters with just as much as, if not more than, precision in a laboratory setting, which may have a larger degree of error than in a computer-based experiment.

And yet, often they are relegated to unimpressive, uninteresting experiments that reduce them to a simple treatment group that fail to explore the intricacies of the task at a variety of levels of analysis. Studies that use games as treatments often fail to expand their investigation beyond a handful of binary questions which cumulate to very little insight into human cognition and behavior.

Beyond the limitations of treatment groups, games seem to still only be used for very niche avenues of investigation. Big Data are overwhelmingly used to study skill acquisition, which, although longitudinal game data are uniquely useful for, is only a single way that available Big Data can be used. Many online games track keystrokes or each action taken during
a game for replay purposes, and although they are not stored as indefinitely as ranking scores or play activity, are still a large source of naturally occurring data. Although it is understandable that ease of access is a barrier for many researchers, surmounting the obstacles to obtaining this kind of data or picking from games whose data are not as easily accessible, either by working with game companies directly or developing a game specifically for this purpose, would drastically improve the diversity of game studies and the types of games they look at.

Additionally, there are still startlingly few well-defined computational models, though it seems in recent years these have started to gain popularity. Computational models are one of the best ways to robustly define the relationship between different variables in a system and are currently consistently accepted as a reputable analog for the brain. Many of the studies I have explored have lacked complex processing models and would benefit from their inclusion, even if they do not aim to contribute to the development of a unified theory of cognition (Gobet 2017).

Finally, there are aspects of games that have yet to be thoroughly explored, namely the social and collaboration aspects (Chabris 2017). Multiplayer games are incredibly popular, and a lot of online game data is sourced from online multiplayer games, yet little attention has been paid to the interactions that take place during multiplayer games. Multiplayer games may be collaborative, competitive, or some combination thereof between players, and can occur remotely (online-play) or in-person (couch co-op or split-screen multiplayer). These interactions have an as-yet untapped potential for research questions that ought to be explored in the future.

However, it is disrespectful of me to criticize past works so extensively without proposing some recommendations for using video games for neuroscientific research. Thus I will create a brief outline for deciding when to use a video game in your research and how to go about selecting one, choosing how to study it, and potentially developing one if necessary.
When to Use a Game

Games are useful for cognitive research in that they are “designed to produce behavior approaching the complexity of the real world by being intuitive to players,” and thus they can be used to study inductive biases (previously-acquired knowledge or skills of learners) (Allen et al. 2023). These biases are represented in the brain and thus studying them can provide a lot of information about how the brain learns and makes decisions. Any study that wants to study the brain in a more naturalistic setting, a more complex environment than a laboratory can provide, could benefit from using a game as their research task. They also provide a great framework for Newell’s requirement of a manageably complex task, and the resulting hypothesis can therefore often be expected to represent or contribute to an understanding of a “genuine slab of human behavior.”

Additionally, the enjoyment of games attracts a wide audience across racial, socioeconomic, gender, and age demographics, creating a vast and diverse potential player base (Allen et al. 2023). The intrinsic entertainment of the task increases the amount of collected data while also preserving the diversity of participants both for laboratory studies and Big Data analyses. The massive online player bases for games facilitate the generation of massive amounts of data and a significant community to interact with and source data from. Many participants may find that participating in a study where they get to play games is more interesting to them or worth their time, increasing the number of interested participants and potentially decreasing the expected scale of their compensation. This makes games an ideal candidate for when you want to survey a large audience or attract a large number of participants.

Finally, online game data is also a great source of longitudinal data, which makes them ideal for studies that want to track certain metrics over a long period of time. These data avoid
many of the issues with retention and continued compensation that laboratory longitudinal studies struggle with and are potentially sourced from a much larger number of individuals.

*Self-made Games vs. Existing Games*

Having opted to use a video game for research, the first major choice is selecting between using a commercially available game or developing one’s own game. Generally speaking, using a ready-made game saves a lot of time and good selection ensures a well-designed game with well-defined parameters that maximizes player engagement (and likely already has an active market), while developing a new game requires a significant investment of resources and has the potential to be poorly made if not done right. However, using a ready-made game also has a variety of limitations that should be taken into consideration.

The earliest video game developed as a research tool, *Space Fortress*, arose due to the limitations of ready-made games that made them unsuitable for research. Although technology has improved vastly since the 1980s, many of the obstacles researchers encountered persist today. Donchin and colleagues determined that:

“A game is a useful research tool if, and only if, the investigator can exercise systematic control over the game’s parameters. Furthermore, unless the game can yield very detailed measures of performance, as well as capture the actual game for replay, the research will be impoverished” (Donchin 1995).

These attributes are not commonly available in commercial games and, as was true in the 1980s and is still true today, developers are reluctant to share source code with consumers or researchers, making the first point virtually non-existent in commercial games (Donchin 1995). To the second point, while a lot of this type of detailed data is collected, it is not always publicly available and developers may be unwilling to share it privately with developers, or there are legal
or ethical concerns with doing so. Additionally, there is very little control over how a game is presented to players (Donchin 1995). Commercial games may include aesthetic components that interfere with or obstruct the desired cognitive process, degrading the data quality. Sourcing the data generated outside a laboratory setting also means that the researcher has no control over the environment of the gameplay or knowledge of the participants’ prior experiences relevant to the task.

Occasionally, it may be possible to clone or modify an existing game for use in the lab if data accessibility is poor or if the game requires simplification for study. This is a valid option achievable through the implementation of various stealth assessment measures (basically just collecting data from players as they play) for when research may require more precise control over parameters and access to data but does not warrant a significant investment in creating a game (Shute et al. 2016).

*Developing Your Own Game*

When confronted with these obstacles, the alternative is to develop one’s own game. A game must meet several criteria for the researcher, which are similar to those required by any laboratory experiment (Donchin 1995). To make for a suitable research tool, a game must 1) “monitor, and record, all variables that might be relevant in describing the way in which the subject executes the task”; 2) “monitor, and control for, differences between individuals in their approach”; 3) be “sensitive to … ‘subject options’ … [which] leave much room for the subject to pick and choose among interpretations and strategies”; and 4) “ensure that the raw data are retained well beyond the time of the original analysis” (Donchin 1995). If it is not feasible to achieve these goals with a commercially available game or a clone of such a game, then developing a self-made game is the only other option for using a game in research (Fig 7).
Fig. 7. Flow chart summarizing key decisions in deciding whether to use a clone game, pre-made game, or self-made game.

A researcher must also choose how they are going to use their game, either in laboratory experiments where they control all aspects of the experiment or by making their game publicly available and collecting data remotely. Public games have the advantage of potentially obtaining data of the nature and scale of commercial games but at the cost of investing time and resources into marketing, polishing and maintenance, and tools for remote data collection. Games for use in the lab minimize these concerns and provide the researcher with direct control over how the game is experienced, but the experiment will be constrained by the laboratory environment in terms of time and the organicness of the task.

Beyond these basic experimental criteria, the rest of the game depends on the nature of the research question and the experiment. Game design is not incredibly dissimilar from experimental design in terms of setting parameters, goals, and, potentially, performance metrics.
for the player/participant. It is important to identify what kind of cognitive metrics you might want to study or what kind of experience or want the user to have depending on whether you are opting for a stealth assessment of specific skills or a general assessment of participant behavior (Allen et al. 2023).

As many games involve learning or improving in some regard, it is also important to include feedback methods to aid the player in their discovery (or recognize that the absence of feedback affects the natural learning process in games) (Deveau et al. 2015). Feedback is most effective when there are multiple senses involved, such as a flash and sound effect when a player lands a hit on an enemy. If you choose to omit feedback, know that this is a choice and should be treated as a manipulated variable in most cases.

Researchers may also wish to “streamline” their games as much as possible to get the most information about the parameters of interest (Rafferty et al. 2014). While this is a valid concern and a possible reason for deciding not to use a commercially available game, one common mistake is stripping away so much of the game experience that the “game” is virtually indistinguishable from a standard behavioral laboratory task. “Optimal” game design is not synonymous with making every single aspect of the game serve a specific, maximally efficient role. Often this interferes with the organic feeling of gameplay or results in the removal of key game design components—particularly the challenge-reward loop by oversimplifying the challenge or by devaluing the reward. In fact, when discussing the optimally designed game that is supposed to serve as an example of good game design for neuroscience research, the researchers state that the behavior exhibited by the players did not match the behavior desired by the researchers with how they designed the task—the goals the researchers had set out for the player did not match the goals the player identified. In this case, players focused more on
understanding the rules of the game than on achieving the highest possible score immediately, which underlines both the importance of a streamlined onboarding process as well as game design that captures the skill or process one is aiming to study (Rafferty et al. 2014). Since “participants do seem to be acting based on a different reward function than that given by the point structure,” the researchers even suggest tying the participants’ monetary compensation into the score to redirect player focus on that aspect of the task, which underscores their neglect for the design of their game beyond statistical optimization (Rafferty et al. 2014). Self-made “games” that subscribe to this type of design philosophy disrespect the practice of game design as a whole and misunderstand the process and skill involved in creating fun, interesting games. This is not to say that simplifying a game is a poor decision—often it is a reasonable choice for laboratory experiments to narrow the scope of a game to fit the time limit and setting. However, doing so requires an understanding and appreciation for the construction and design choices of a game.

This last point in particular is why I encourage researchers seeking to develop their own game to work with a developer on their game. Much like a student writing their first paper or a scientist designing their first experiment, there is a lot that the inexperienced practitioner can get wrong in their first iteration, often requiring the aid of someone more experienced to guide them and continued practice to improve. Even someone who studies a particular activity or subject extensively may still be missing critical information that can only be obtained by practicing the activity or subject itself (Chabris 2017). For a research game, an individual developer or an indie studio would make the ideal partner as they are flexible and are typically well-versed in many aspects of game design rather than specializing in a specific medium. Working with, or at least
consulting with, an experienced game designer is critical to creating a polished, useful research tool.

Conclusion

Developing a suitable research question is an important step in all research fields, particularly in how one will investigate and answer their question. Fifty years ago, Alan Newell critiqued the cognitive scientific fields for the quality of their research questions and their methodology of investigating them, and several of the problems he identified then persist in neuroscientific research today. As I have examined, video games provide a unique solution to these problems by encouraging the exploration of more complex tasks at multiple levels of analysis and by being a suitable medium for developing computational models. Although there are several ways video games can be used for research, studies that create computational models tend to maximize the advantages of video games as a research medium and the amount of information one can gain about human behavior and the underlying systems. When selecting a game, there are various aspects of the game and its source that a researcher must consider depending on the nature of their research. Depending primarily on data accessibility and parameter control, a researcher may opt to alter a preexisting game or develop their own game from scratch. However, if the researcher does not have an understanding of game design or the construction of their chosen game, their data can be negatively impacted due to their inability to create the desired behavior or satisfiably measure their independent variable. Rather than splitting their expertise, I suggest that, when working with video games as a research medium, researchers would benefit from working alongside or consulting a game designer, whether it be an independent developer or an indie studio. Respect for and understanding of games and their
design allows for the greatest potential for video games to augment neuroscientific research and address the dearth of interesting research questions.
Acknowledgements

Foremost, thanks to my readers for guiding my research, their insightful feedback, and overall support over the course of this project. I would not have even known where to start with this process if it were not for your kindness and guidance. Thank you to my dad, who inspired my love of science and encouraged me to think big for my future. Thank you to my mum, who was always there for me and supported me when I felt like nobody could. Thank you to my co-researcher (my cat, Tigerlily) who stayed up late with me during research sessions and never judged my sleeping habits, and my friends whose support kept me going throughout this process. And a massive thank you to video games, who always kept me thinking, well, I have to know what happens next, what is over that hill, what secrets might I uncover, when I otherwise felt like shuffling off this mortal coil.
Literature Cited


