

# Towards Designing Games for Experimental Protocols Investigating Human-Based Phenomena

Phil Lopes  
phil.lopes@epfl.ch  
EPFL  
Lausanne, Switzerland

Ronan Boulic  
ronan.boulic@epfl.ch  
EPFL  
Lausanne, Switzerland

## ABSTRACT

Over the past few years scientific research has opened up to the idea of using digital games for human-based studies. Fields such as Neuroscience, Medical and Affective Computing are currently using games to study human-based phenomena. Even though a vast amount of work exists within the field, rarely is the subject of designing such games ever touched upon. In fact a common problem within the field is that the games themselves are often an afterthought, where certain gameplay limitations are never truly acknowledged and tend to be mostly ignored. Thus, this paper intends to provide some game design guidelines to the most common problems found in literature from work specifically using games for human physiological data-collection purposes. Furthermore, a brief description of the most popular physiological recording methods: Skin Conductance (SC), Heart-Rate Variability (HRV), Electromyogram (EMG), Electroencephalogram (EEG) and Functional Magnetic Resonance Imaging (fMRI); are provided as the game-play “limitations” of using such devices are an important factor to take into consideration in the game design process. As such, the objective of this paper is to provide awareness of specific game design limitations found in literature and analyse them from a game design perspective.

## CCS CONCEPTS

• **Human-centered computing** → *Interaction design theory, concepts and paradigms*; Virtual reality; Graphical user interfaces.

## KEYWORDS

Game Design, Physiology, Games for Research

### ACM Reference Format:

Phil Lopes and Ronan Boulic. 2020. Towards Designing Games for Experimental Protocols Investigating Human-Based Phenomena. In *International Conference on the Foundations of Digital Games (FDG '20)*, September 15–18, 2020, Bugibba, Malta. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3402942.3403012>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

FDG '20, September 15–18, 2020, Bugibba, Malta

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-8807-8/20/09...\$15.00

<https://doi.org/10.1145/3402942.3403012>

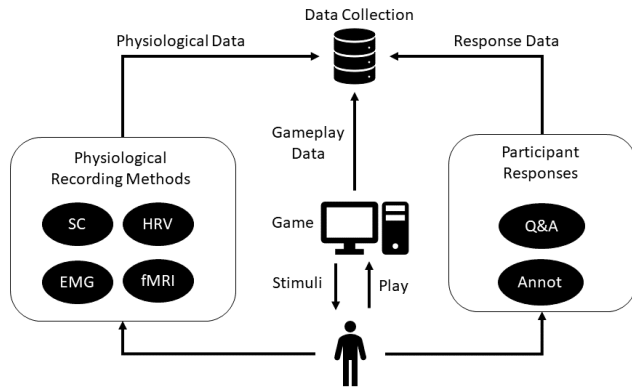
## 1 INTRODUCTION

Digital games are often thought as purely entertainment endeavours. Although this is true in most circumstances, the benefits of play, interactivity and engagement prevalent in the medium has slowly been used for the study of several factors in human related phenomena. Given the recent affordability of Virtual Reality (VR) thanks to hardware such as the *HTC Vive* (HTC, 2016) and the *Oculus Rift* (Oculus, 2016) also significantly contributed to this rise of interest in academia. Fields such as Neuroscience, Affective Computing, Psychology and even Medicine [24] for example, have been exploring the usage of games and VR for their particular research.

Research fields such as neuroscience and affective computing often rely on analysing human bodily responses to study the emotional impact of individuals towards an external stimuli (e.g. pictures, videos, movies and games). The interactivity and engaging capabilities of games are often what attracts these researchers to the medium as it forces the participant to directly interact with the experiment, which facilitates the stimulus of certain human conditions. The easiest example is to think of the inherent nature of games: players either win or lose influencing the end result; which often induces intense emotions during the course of play.

Although the nomenclature “Serious Games” (SG) has often been used to describe educational- or training-like games and is a popular topic for investigative purposes, they are often designed around some purpose where the end-goal of play is to benefit the player or society in some capacity. Although the type of games explored in this paper fits into this mould it intends to be more specific, where the usage and design of such games are purposefully constructed for the study of human behavioural patterns and scientific experimentation. As such, these games “serve the purpose” of an overarching research question and must work within the physical constraints of physiological human data collection. These methods present a series of interactive limitations, which include unusual postures, limited movement, a degree of sensor sensitivity, game-and-sensor synchronization, calibration and for EEG an extensive setup procedure. As such, throughout this paper the expression *Experimental Protocol Game* (EPG) will be used to describe these types of games for the sake of clarity.

The objective of this paper is to provide both a literary review of work exploring digital games that were specifically designed and developed for human-based data collection and research, while also providing an overview of the limitations and possible methods on overcoming them. Furthermore, given the rising popularity of digital games as experimental tasks for scientific study, this paper argues that it becomes increasingly necessary to establish, personalize and design games around the specific limitations and necessities of human-based data collection and research parameters.



**Figure 1: The typical EPG setup.** During the act of playing a game the physiological recording devices record participant physiology and gameplay data (e.g. Player Input, Game States). Ideally both data streams are synchronised to ensure that in-game events are accurately timestamped for physiological recordings to ease its analysis a posteriori. At the end of or during the intervals between gameplay, players are often tasked in either answering questionnaires, or in some cases annotate memorable events that occurred over the course of play.

Thus, this paper is organized as such: first the literary overview of the different research fields that have used games to study human-based phenomena is presented; the following section provides short descriptions of the most common devices used to capture human-based phenomena, where a larger portion is dedicated specifically to fMRI given its rarity within the digital game space; the last section is dedicated to the analysis of several prominent EPGs within literature, where the description of common design problems is observed and respective solutions on how to overcome them is given.

## 2 DIGITAL GAMES FOR SCIENCE

Digital games have been used to study human-based behavioural patterns since the early 1980s [61], where the field of psychology in particular have used games specifically to study concepts such as human memorization, knowledge transfer and even emotion stimulation [63]. However, according to Järvelä et al. [26] one of the main problems of using games as an experimental stimulus is its variability. Experimenters must be aware of all independent variables during the development and subsequently enforcement of their protocol. The reason is that there are certain aspects of play that are simply unavoidable such as for example the player skill, where certain individuals may finish or severely outperform others. Players also have agency within the virtual world, and thus the game must be sufficiently identical between individuals so that it is relevant to the variables intended for study.

Given such limitations one may ask: “Why should I use digital games instead of a more controlled experimental protocol?”. The advantage of using games is that it has the capability of engaging participants by providing a certain amount of control and agency

on the experimental task being performed, which in turn can lead to a more “genuine” or effective reaction [26]; or is capable of engrossing the player in performing tasks which would otherwise be uninteresting [1, 24]. In fact according to Järvelä et al. [26] the importance of keeping the variance of games is emphasized, as otherwise it removes all the positive qualities that games may offer, thus transforming the overall experiment into a standard experimental task, where participants are expected to recreate the exact specification described by the experimenter. As such, it is important to understand that using games for experimentation is somewhat of a balancing act, where the game must offer a degree of control for experimenters to study and force the variables to emerge during play, while also maintaining the interactive and agency aspects for players so that the benefits of play are existent (e.g. decision making, control). This also aligns with the suggestions offered by Washburn [63] on the problems and benefits of EPGs used for scientific study.

As such the following section discusses works using EPGs for different investigative purposes, where human data is collected specifically for studying or treating (medically speaking) specific human-based phenomena. In medical applications, the collected data are often used to measure the effectiveness a game can have on a specific treatment such as distracting patients from painful procedures [24], rehabilitation [1] or purposefully inducing traumatic events in a fully controllable environment for psychological treatments [25]. The field of Neuroscience and Affective Computing have often used games to evoke or study specific human behaviours, such as the effects of games on human physiology [61], emotion [12, 42] and learning [21]. Such studies have also been slowly transitioning into the field of games itself, where some researchers have been using such concepts to improve the medium itself and its entertainment value. A recent field of research termed *Experience-Driven Procedural Content Generation* (EDPCG) [68] consists of combining the field of Affective Computing with Digital Games to investigate how emotions can influence and even improve the player experience.

### 2.1 Experimental Protocol Games to Evoke Human Behaviours

It is often stated within the game design literature that games are a means of conveying an experience, where designers define rules, aesthetic properties and interactions which can provide impactful and meaningful experiences for the players engaging them [19, 55]. Similarly to other types of media such as film, painting and music, games are a form of emotional expression, where often the best games are the ones that were able to “connect” players on an emotional level. To quote Muriel and Crawford [45]: “[the players] might connect at an emotional or cognitive level (or possibly both); the important thing is that the connection is established and, in doing so, it makes the access to other points of view possible and facilitates an understanding of distant situations”. Thus, it is not surprising that designers tend to pay close attention to player emotions evoked during play, a particularly important process when play-testing a game [19, 55].

Affective Computing as described by Picard [48] consists of computational systems with the capability of recognizing human

emotions, and using this information to adapt and assist the human user during an assortment of computational tasks. For this to be possible it is first necessary to study individual emotions and have the ability to recognize said emotions through an autonomous computational process, which would subsequently allow systems to adjust based on the recognized emotion. Human emotions are often tied to bodily manifestations, for example anxiety tends to be accompanied by a higher heart-rate [54] and humans tend to sweat more during these events [65]. Thus, the majority of researchers in the field tend to rely specifically on physiological data for the emotion recognition task. However, to build a model capable of recognizing specific emotions from physiological signals, it is first necessary to collect a reasonable amount of data that can prove a strong statistical link between both. As such, given the potential of games in evoking human emotions, it is unsurprising that the field has seriously considered using games as a way of collecting such data.

One of the first notable papers to explore the impact of games on human bodily responses was in the work of Turner et al. [61]. The authors observed a substantial increase of player heart-rate when compared to the control condition, suggesting that the game (i.e. Space Invaders) had a substantial impact on participants. More recently the work of Whitehill et al. [64] used simple “brain-” games which feature several puzzle elements to measure low- and high-engagement from extracted facial expressions of each participant. Games can often be frustrating if players find it hard to control or too difficult [19], hence it is a good platform to evoke such emotions. In the work of Song et al. [56] in particular a game was constructed with voice activation control allowing players to navigate through a maze. As the player progresses further the voice activation becomes subsequently more unreliable resulting in a frustrating experience. The objective was to recognize frustration through speech and facial expression data. Researchers have also been using games to build extensive physiological datasets such as the work of Moghimi et al. [44], where the authors used a speedboat racing game to collect EEG and SC (see Section 3.1) signals of 30 participants. By applying an unsupervised learning approach the authors attempt to cluster over 743 features extracted from the latter signals into four emotional categories derived from the Russel Model [52] and eight discrete labels defined by Ekman [16]. The popular game *Wii Sports* (Nintendo, 2006) was used to build models capable of recognizing affective states [31].

Virtual Reality (VR) has players completely surrounded by a virtual world, interaction is heavily based on physical movement and players are deprived from “real-world senses”, fully transporting them into the virtual experience. Given these features it is no wonder that researchers are interested in using VR for the purposes of emotion elicitation. In particular, the work of Meuleman and Rudrauf [42] explores the potential of VR games and their ability to evoke a variety of emotions. The authors explore seven different games with diverging interaction methods and genres, where each game was tied to a dominant emotion which the authors suspected to be elicited by that specific game. Using a clustering method on physiological features the authors observed that emotions such as fear and joy provided the most accurate results. Chirico et al. [11] on the other hand explores the concept of “Awe”, where the authors claim that VR is the most effective way of inducing such emotion.

Awe is described as the perception of vastness and the need for accommodation. Gabana et al. [20] hypothesized within their work that VR can have substantial benefits on *Working Memory Performance* (WM) in comparison to standard “Desktop” games. Results from this work suggest that participants who have a low WM tended to perform better in VR. Furthermore, the authors also suggest that the emotional intensity of VR experiences tended to be higher in VR than on Desktop according to the participant self-reports.

Evoking certain human behaviours are often best achieved with games due to its interactivity. However, a common criticism in using games is that they often leads to noisy data and it is difficult to parse, especially if using physiology. Hence why most researchers tend to rely mostly on self-reporting questionnaires. Thus, this work suggests that certain considerations are necessary if using games for them to be effective. Such as if using physiology, it is necessary to think about the limitations of all devices used in the experiment, and also take into consideration the timing it takes for these specific bodily manifestations to appear. As such, this work attempts to provide game design considerations when creating an EPG.

## 2.2 Experimental Protocol Games for Clinical Treatment

Engagement is a regularly used term within the Digital Game space, often described as players being fully engrossed in the playing experience and actually losing one’s sense of time [13]. The engaging capabilities of digital games was explored by Hoffman et al. [24] specifically as a means of distracting burn victims during wound treatment, thus working as a non-pharmaceutical analgesic. The authors developed a Virtual Reality (VR) game called *SnowWorld* where patients could roam a snow and ice covered virtual level and throw snowballs at snowmen (i.e. the enemies). Data was effectively collected through self-reports, which reported a pain reduction of 35 – 50%, and fMRI scans showed reductions in pain-related brain activity during the VR experience. Validating the effectiveness of games through human-based data is a crucial aspect especially when dealing with clinical studies, thus this paper argues that taking into account such limitations is valuable even for such applications.

Games are also often used as motivators, such as transforming rehabilitation exercises into playful activities [1, 35]. It has been observed that patients who use rehabilitation games have a tendency to be more efficient and also have a higher degree of engagement, thus providing sufficient motivation to keep coming back to subsequent therapy sessions. Although the intent of such games is to provide playful activities with the ability to motivate participants, there is substantial benefit in also collecting this data providing information on the effectiveness of treatments or simply tracking a patients history (e.g. for rehabilitation). The work of Rivas et al. [51], utilizes the previous methods and applies features explored in affective computing to further understand patient motivation, by detecting tiredness, anxiety, pain and engagement. The hypothesis being that games and therapists can adapt such games increasing patient retention. Thus, the argument can be made that even for clinical applications the usage of EPGs can be important as the

data used can be re-applied for clinical applications and potentially further improve rehabilitation games.

The inclusion of *Clinical Treatment* within this work showcases the importance of designing EPGs around the specific objectives and limitations of the research. Furthermore, validation for clinical applications requires an extensive evaluation process, which necessitates experimental data obtained from human participants. Such as the case of *SnowWorld* [24], where fMRI was used to observe the efficiency of their proposed method. Furthermore, when dealing with rehabilitation activities data are required specifically to keep track of patient progress and personalize their rehabilitation activities. Thus, this paper argues that applying game design knowledge for games that intend to provide some clinical benefit can potentially further improve the quality of said systems and the data extracted.

## 2.3 Experimental Protocol Games for Entertainment

The previous sections explored the concept of using games for a purpose of studying human-based phenomena for applications outside the digital game domain. This section discusses work within the literature that use EPGs specifically to improve the player game experience, hence improving the entertainment value of the medium.

In particular, authors have explored how physiological data can be viable in the game design process [46, 53]. In addition to this, some games have even explored the concept of using physiology during the actual act of play as a way of enhancing the game experience [6]. One such popular example is the game *Nevermind* (Flying Mollusk, 2015), which uses Heart-Rate Variability to alter certain aspects of the game itself.

Experience-Driven Games (EDG) [68] is a field of research that applies the concepts of Affective Computing for the medium of Digital Games. “Experience-Driven” entails that games should be reactive and capable of adapting themselves based on the player’s current experience. This is referred to as an “affective loop” where the game influences the player experience, which subsequently affects the adaptation method, and the latter in turn manipulates the game. As such, systems that explore this concept often rely on human-based data to either construct statistical models necessary for such systems to function, or for the purpose of evaluating the system and confirming it induces the correct experience. Thus, EPGs have also become a necessity for this type of research, at least for the small scale “in-lab” experiments and prototypes, where it is crucial to understand if the intended effects are working. Notable research exploring these concepts include the work of Chanel et al. [10], where they explored dynamic difficulty adaptation in the game of Tetris. By analysing the player’s skin conductance signal, the authors built a model capable of predicting the intensity of anxiety or boredom with a degree of accuracy to dynamically adjust the difficulty of a game according to the Flow theory [13]. Yannakakis et al. [67] used physiological data to investigate the ideal placement of player camera viewports within a 3D game. The authors explored specific concepts such as frustration, anxiety, challenge and relaxation to better understand how camera movement can impact the overall playing experience. Lopes et al. [36] used physiological data to evaluate a dynamic audio system capable of placing sound

dynamically within a 3D virtual environment. The system intended to place sound according to an emotional progression, and thus to evaluate the veracity of said system the authors use physiological data gathered from players for comparison.

## 3 DEVICE DESCRIPTION AND LIMITATIONS

To effectively use a type of physiology, it is the authors opinion that it is important to have an overall understanding on the limitations and specific characteristics of each physiological signal. One of the first aspects to consider is that physiological signals are vastly different from each others, as are their respective stimulus-response windows. The latter consists of the average time necessary for the human body to react to a stimulus. It is important to state that the topic of “how to process each signal” goes beyond the scope of this work, given that it is a post-experimental procedure. However, the suggestions discussed within this work intends to aid this process by offering game design solutions that can lessen the overall noise from physiological sensors and taking into account the stimulus-response window.

One of the most common problems of using EPGs is noise, as the majority of physiological recording devices are very sensitive to physical movement. This can lead to artefacts appearing in the signal data, which can ideally be filtered out in post-processing, or in the worst case scenario be discarded. The following section provides a brief overview of the most popular physiological devices being used within literature, where a larger portion of the section is given to fMRI. Considering the lack of resources regarding the functionalities of fMRI outside of the medical/neuroscientific field a more in-depth analysis is given. Table 1 presents a quick reference of the physiological devices explored within this work.

### 3.1 Skin Conductance and Heart-Rate Variability

Skin Conductance (SC) and Heart-Rate Variability (HRV) are one of the most common physiological modalities explored within literature [10, 15, 36, 49]. They are the most cost efficient thanks to the wide availability of multi-modal physiological devices such as the *BITalino* [3] or *Empatica E4* [40] which remain relatively inexpensive. Furthermore, the fact that they are easy to setup and are among the least intrusive “over-the-skin” devices available has definitely contributed to its extensive usage within literature.

Sympathetic responses for HRV (Heart-Rate Variability) have been shown to have a delay of approximately 1 to 2 seconds after stimulus, and a response can last up to 20 seconds [5]. For SC the response is similar, where it has been shown to typically respond one second after stimulus, while having a shorter duration of approximately 4-5 seconds [4]. In terms of sensor placement it specifically depends on the device. For example the *BITalino* (Plux) uses a more “traditional” method where electrodes are placed over key locations of the body, typically on the chest for HRV (Electrocardiogram) and fingers or feet for SC (Galvanic Skin Response). Such electrodes are often wired, effectively tethering players to the device restricting their physical movement. On the other hand devices such as the *Empatica E4* (Empatica) consist of a “bracelet”-like device which is placed on participants wrist and has the ability to record both SC and HRV. However they come at a cost, such as a lower sampling



**Table 1: The limitations and respective placement of the discussed physiological devices - Skin Conductance (SC), Heart-Rate Variability (HRV), Electromyogram (EMG), Electroencephalogram (EEG) and Functional Magnetic Resonance Imaging (fMRI). In the context of this table limitations refers specifically to the physical and posture constraints which can hinder gameplay.**

Data	Limitations	Placement
SC	Wires	Hands / Feet / Wrist
HRV	Wires	Chest / Wrist
EMG	Wires	Face / Arms / Legs / Wrist
EEG	Wires / Limited Movement	Scalp
fMRI	Posture / Limited Space	N/A

rate and less accuracy for SC given the non-ideal location of the GSR sensor.

### 3.2 Electromyography (EMG)

EMG is a device that detects muscle contractions. Electrodes are placed on key “muscle activation” locations, and depend on the specific muscles an experimenter wishes to record. For example, EMG can be placed on a participant’s face to record facial muscle movements, on the arms to detect contractions from biceps or even the lower arm for finger and hand movements.

Unlike SC and HRV, muscle contractions are usually voluntary in healthy individuals and not “automatic” responses like the former. According to [32] the average reaction time of healthy individuals while using EMG was approximately 200 ms, however this may vary substantially depending on the task at hand. This is especially true if the task requires participants to keep their muscles contracted for longer periods of time. When it comes to devices several options exists offering different sensor types. It can work by placing electrodes on key regions of the body (e.g. BITalino) or more recently with EMG specific bracelets such as the *Myo Armband* (Thalmic Labs). Even though the latter is easier to setup it is designed specifically for arm muscles, while the former provides significantly more options.

### 3.3 Electroencephalography (EEG)

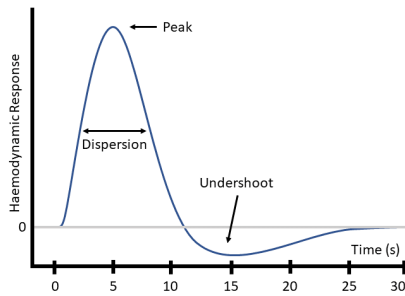
EEG is a physiological device capable of detecting electrical activity originating from the human brain. Its usage was often used exclusively in the field of Neuroscience, but given the wider and cost effective availability of these devices it has seen prominence in other fields of research. Typically electrodes are placed on key locations of the scalp, and can vary in number between 4 to 256 in total. Given the quantity and the specificity of electrode placement, one of the main downsides of EEG is the laborious setup process, which can often take significant chunks of experiment time.

The stimulus-response window for EEG - often referred to as Event-Related Potential (ERP) - is significantly more complex than previous methods. First experimenters have to work with a wider range of channels, which is dependent on the number of electrodes used. Secondly an ERP can vary substantially based on the type of stimulus. Specifically an ERP consists of small voltages originating from the brain in response to specific events and can be either positive or negative (i.e. polarity). It is common for researchers to reference specific types of ERPs based on its polarity and the response-time after stimulus. For example, a P100 and N200 denotes

a positive and negative ERP with a response-time of approximately 100 and 200 ms, respectively. The work of Sur and Sinha [58] offer substantial insight on common patterns found in literature between different ERPs and stimuli. However, notable ERPs suggested within research that use games to study human-based phenomena with EEGs has included the P300 [18, 62] and the P200 [66]. Although, it is important to emphasize that this does not mean that other ERPs are not viable for games and can even change based on certain factors present in the game itself, which is something to consider during designing an EPG. Lastly, it is important to consider the physical movement limitations as participants will be tethered by several electrodes. Thus, it is important to ensure that participants can easily play without having to look at the controllers to avoid head movement, which also means that games using head-tracking may not be the most ideal for this type of device.

### 3.4 Functional Magnetic Resonance Imaging (fMRI)

In both the medical and research fields fMRI scanners are a frequently used imaging tool allowing physicians and experimenters to observe human brain functions. It does this by exploiting the magnetic properties of the hydrogen nuclei, given that the most common atom within a human body is water. Although several processes are necessary to construct the final neurological image, the underlying process comes from detecting the signal strength variance of the hydrogen nuclei which varies depending on its surroundings, given that heat influences its behaviour. Therefore providing a means of discriminating between the different sections of the brain for image reconstruction. fMRI also offers the ability to visualize the changes of blood flow in the brain by using a method called blood-oxygen-level dependent (BOLD) imaging [8]. The process consists in analysing the different magnetic properties of the haemoglobin, which in turn is responsible for transporting oxygen in the human body. The haemoglobin when “oxygenated” presents different magnetic properties than when it is not, thus allowing the ability to detect the small variations of oxygenation. Considering that neural activity demands different requirements of oxygen depending on which neurons are currently being activated, this process allows us to observe - albeit indirectly - neural activity. The body is constantly prioritizing and re-adjusting its blood flow by providing the necessary oxygen and other nutrients to body parts that are currently undergoing stress (the Haemodynamic Response). Considering that BOLD works by analysing the different oxygenation levels of haemoglobin, this method therefore



**Figure 2: The typical Haemodynamic Response (HR) as described by [23]. Peak response occurs at approximately 4-6 seconds after stimulus (oxygen compensation), and is subsequently followed by an undershoot that can last approximately 10-30 seconds before returning to the baseline value.**

allows for the observation of blood flow activity present in the brain, referred to as the Haemodynamic Response Function (HRF). More precisely, the HRF consists of the phenomena where an increased blood flow is observed in certain locations of the human brain. Given that activated neurons will require a higher amount of oxygen and nutrients it can be determined that locations with larger blood flow concentration correspond to a higher neural activity [8, 22]. Hence, the HRF can be considered the fMRI's "version" of the stimulus-response window (see Fig. 2). This response consists of a peak occurring at approximately 4 to 6 seconds after stimulus. It is then often followed by an undershoot that can last up to 30 seconds until stabilizing [22]. For the interested reader a more in-depth description of fMRI can be found in the work of Buxton [8].

fMRI has often been used to study the influence of games in neuroscientific literature [21], however the act of playing a game was rarely done during the actual fMRI scan. Studies tended to usually focus on the indirect influence of games (e.g. transfer learning, improved focus), where they have participants playing games outside the scanner and then observe their effects on the brain based on other simpler tasks conducted during the actual scan [21]. However, this trend has shifted over the years where researchers have become more interested in studying the influence of game playing during the actual scan [2, 28]. Furthermore, the latest development of fMRI compatible VR goggles such as the *VisualSystem* by NordicNeuroLab has also further facilitated this process.

One advantage of fMRI is its non-intrusiveness as it does not require physical sensors be applied on the body or an extensive setup process (e.g. EEG). Thus, for experimenters a longer time can be spent on the experimental task easing the burden on participants. However, if using digital games in an fMRI study there are a few considerations that need to be taken into account. The first consideration is to understand that players are playing a game lying-down, which can lead to potential side-effects such as motion sickness [39, 59]. The other consideration is the limited physical movement, as participants are not allowed to move their upper torso or head during a scan. The reason being that head movements

effect the data obtained from the scanner, which may lead to artefacts or even unusable data. Thus, during the design of a game for fMRI these considerations must be seriously considered.

## 4 APPROACHING GAME DESIGN FOR EPGs

This section provides several considerations necessary when designing EPGs. However, it is important to note that the suggestions offered within this work are not the de-facto standard. In fact a lot of the suggestions are open ended with solutions being very specific to the study. Thus, a lot of the considerations are based on available literature applying EPGs and common concepts within game design. The objective of this work consists solely of pointing out the specific difficulties that most commonly appear when designing EPGs, and potential methods to overcoming said problems.

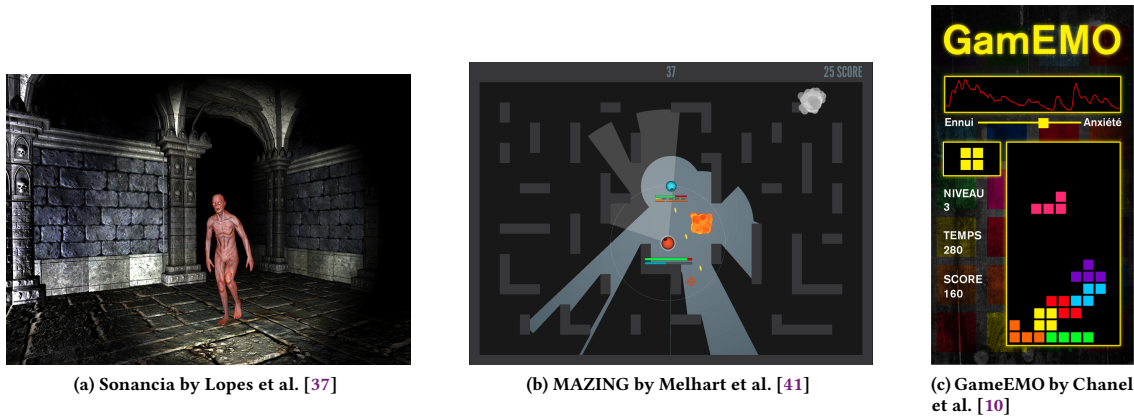
### 4.1 Adapting or Starting Fresh?

The usage of open-source or commercial accessible games have been a staple within the community, where games are often reappropriated for scientific study [33, 60, 61]. It is clear that there are several advantages in using established digital games such as player familiarity and a more "polished experience" rather than a rough prototype [26]. In fact picking the "right" game is a crucial aspect as it is necessary to ensure the conditions being studied are effectively met. However, even taking the necessary precautions it can become quite difficult to adapt a professionally developed game to the very specific needs of the experimenter as they are notoriously "closed-source".

Several solutions have been explored such as modding [57] and even using open-source variations of other popular games [9], although this does restrict the experiment and research to a limited set of potential games. Furthermore, studies may require a certain amount of personalization to the point where using an established game can become more labour intensive than just developing a small "to-the-point" game. For simpler games it may actually be possible for experimenters to simply recreate a popular game themselves such as Tetris [10] or even Pong [34].

The reality is that building a game from scratch is a costly process, and often times the expertise to build a high-fidelity game similar to commercial products are rarely available in academia. However, in the past few years powerful game engines such as *Unity* (Unity Technologies, 2005) and *Unreal Engine* (Epic Games, 1998) were made publicly available and free for non-commercial products, which has significantly facilitated the game making process for academics and amateurs alike. This, in conjunction with the large amount of resources available on-line, such as content (e.g. animations, models, code), tutorials and public forums provides a significant boost for custom game creation. Some notable examples of games built from scratch includes Lopes' et al.'s *Sonancia* [37] and Maelhart's et al.'s *MAZING* [41].

This paper argues that as these game development tools become more powerful and easy to use, it makes sense to think about constructing games around the experimental limitations, instead of adapting them and severely hindering its experience due to the non-ideal circumstances. This is especially true for EPGs, where physiological sensors can hinder player interaction - as they often require limited movement - and can force them into unfamiliar



**Figure 3: Examples of different EPGs created for Human-Based studies within literature, where 3a and 3b are original creations, while 3c a recreation of popular game Tetris.**

playing postures (e.g. fMRI). Furthermore, a game designed around the different stimulus-response windows would provide “cleaner” data making its analysis easier.

## 4.2 Design Around Stimulus-Response

As previously discussed, the manifestation of bodily reactions occurs differently depending on the physiology being observed. Thus, when designing an EPG it can be important to understand these manifestations and seriously consider them into the game design.

**4.2.1 “Free-form” Play.** The most straight-forward approach and also the most common, usually consists of providing a “more natural” gaming experience (in the traditional sense), where participants continuously play a game, without the experimenters entirely considering the physiological limitations. More precisely, a participant is let “loose” during a game-playing session, allowing them to play for an extensive stretch of time. Usually these sessions are free-form allowing the participants ample time to reach the objective. The advantages of using a free-form approach is that it allows participants to immerse themselves in the experience without any interruption whatsoever. The downside however is that a large uninterrupted signal is obtained making it harder to process and pin-point the exact points of interest in the signal. If dealing with thousands of data points it can be extremely difficult to process without either making some assumptions, or using player feedback/annotations. The latter is necessary to extract “sections of interest” as dealing with an entire signal is often inefficient and noisy.

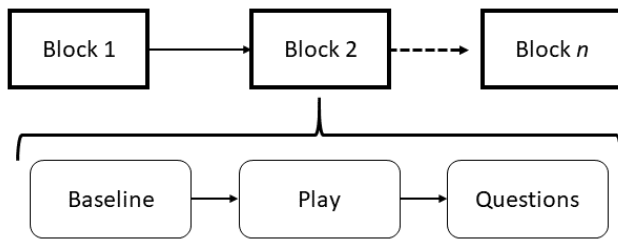
It is also important to consider that if using solely questionnaires with such a protocol, it can be difficult to effectively correlate specific signal features to the user responses. Previous works have attempted to extract meaningful statistical features from signals of entire gameplay sessions, which are used for comparing the effects of different experimental conditions [15, 17]. However, finding better ways of contextualizing the signal can be advantageous. It can help to isolate the intervals of interest from an entire session,

allowing experimenters to analyse only the sections where meaningful events occurred lessening the noise by including everything in-between.

Thus, it becomes necessary to come up with other solutions for user response. One approach is the classic “thinking aloud” method allowing participants to state their current feeling at specific points during play [14]. Alternatively, another method consists of “real-time” annotation, where participants (or 3rd parties) review the entire game-play segment and annotate it based on perceived emotion [12, 38]. A disadvantage of the latter is that it will take the same amount of time to annotate the gameplay as it was to play it which will always double the total experimental time (if relying on 1st party annotation). Additionally, players are recollecting their experience a posteriori, meaning it will not be exactly identical as first-hand experience.

**4.2.2 “Block” Play.** Another solution is to divide play in multiple experimental “blocks”, where “bite-size” levels/sections of the game are constructed and played with the intent of inducing a specific condition where the signal can peak and subside through the subsequent baseline – similarly inspired by block fMRI experiments [22]. A block would consist of three phases: Baseline, Play and Questions (see Fig. 4). The first being a baseline scenario which is the same for all blocks and persists for each block, instead of just at the very beginning. The second being a short play session where the timing is determined by the physiology being investigated. Lastly followed by a quick questionnaire ideally taking no longer than a few seconds and implemented in-game to avoid breaking the player experience too much.

By necessity, if using fMRI this type of protocol is usually considered [22]. As by dividing the task by blocks allows the experimenters to analyse the full range of the HRF while avoiding overlap due to successive stimuli (see Fig. 2). Hence, that is why the majority of works using this style tends to be exclusively studies using EPGs with fMRI [2, 28]. Although this type of design may be more suited for fMRI it can be advantageous to other types of physiologies. One of the most common difficulties of human-based data is actually processing the physiological signal effectively [10, 11, 34, 36]. Given



**Figure 4: Example of a “Block”-style experimental design for EPGs. A block consists of 3 scenarios: Baseline, Play and Questions, where the playing portion consists of short segment of play for facilitating the processing of physiological signals.**

that play occurs over the course of small segments, such a protocol allows experimenters to take the signal from each block and even have the ability to compare between them to analyse the effect of different game-playing conditions. However, for statistical relevance it means that participants will need to play and even repeat several blocks, which are randomized per participant.

### 4.3 Beware of Unintended Side-Effects

Games are multimedia experiences relying on visuals, sound, narrative and interaction. Given all these different facets it is important to make sure that the selected media does not provoke any unintended stimuli so as to not add noise or obfuscate the actual study in question. Thus, the design and piloting stage is often crucial in the construction of an EPG. For example when using VR a common adverse side-effect is motion sickness. Unfortunately, such effects can significantly impact players which in turn influences physiology, as they start to react towards these effects rather than to the intended ones [30]. Although there are no solutions to solve the problem of motion sickness in VR yet, some tools do exist to measure the intensity of its effect. The SSQ [29] is one way to gauge motion sickness intensity, allowing experimenters to measure the adverse effects during the piloting and play-testing stage so that the risk can be mitigated.

### 4.4 Simplicity or Complexity

As suggested by Järvelä et al. [26] a balance between experimental control and “player freedom” has to be achieved to take full advantage of an EPG. In gameplay terms “freedom” is an abstract concept, but for the purpose of this work it refers more specifically to game complexity. More precisely it is a reflection of the amount of game state possibilities, where a complex game presents an exponential amount of game states due to its rules and available player actions, while a less complex game has a limited set of game states. Thus, a less complex game is easier to control experimentally than a complex one as the number of variables increases with complexity. The reason why an experimenter may want a more complex game can arise for several reasons. If looking at the emotion-elicitation perspective for example, certain emotions may only be elicited if a degree of investment from the player exists (e.g. Sadness). Furthermore, it also depends on the degree of influence an

experimenter has over the game without raising player suspicion (e.g. “rigging” game mechanics to favour specific outcomes) which may be necessary for certain studies.

For example a game such as *Tetris* (Alexey Pajitnov, 1984) has a lower complexity given that players are limited to a small action space such as rotating and displacing Tetris pieces. On the other hand, a game such as *Super Mario World* (Nintendo, 1990) has a higher complexity as players have several more actions available comparatively to Tetris, and have an indirect control of the pacing (if no timer is available). The best example where complexity can be relevant is in the field of Dynamic Difficulty Adjustment where Tetris is easier to control as the overall difficulty consists of only one variable (i.e. the falling speed of Tetrominoes) [10], while for Super Mario it can depend of a variety of factors such as platform positions, enemy types and concentrations, etc. [27].

### 4.5 Addressing Failure or “Losing at the Game”

Traditionally when playing a game, the player can either win or lose. A game presented in the context of an experiment is no different, as this risk is often what allows certain human phenomena to emerge for the specific studies being conducted. On the other hand, some experiments may not even require a failure state as the intent is not to provide a competitive setting but to explore how the interaction itself can be beneficial in some specific way (e.g. The Snow World Game [24]). For such cases the complete removal of failure states makes perfect sense.

For experiments that require this risk of failure it becomes necessary to think about the consequences of such risk. More precisely, for a task such as data collection a certain degree of consistency is required among participants, thus abruptly ending play due to failure may not be the best overall solution. Previous research have dealt with failure by either instantly resetting the game [10] or by placing the player back into the start location [37]. However, this tactic may backfire once the players start to understand that failure comes with a minor risk, which may result in players purposefully failing to achieve certain goals. Thus, eliminating the intensity associated with the factor of risk, which in turn may not stimulate the appropriate reactions sought after. Thus, researchers have attempted to find ways of inducing some competitive element in games such as manipulating scoreboards which can be made up of false or realistic scores [7]. A common tactic used within the field of neuroscience in particular, is to provide the illusion of real-world consequence due to poor performance (or failure in this case) [2]. Experimenters deceive participants by claiming that their overall pay will be reduced for each failure at the start of the experiment, even though the full amount is eventually given by the end. Lastly, modifying game-play elements to disempower players (e.g. remove power-ups, move slower) is an alternative solution to penalize players while still keeping the game going.

### 4.6 Rewarding Players – Going Beyond Point-Based Systems

Incentives such as point-based systems or collectibles (e.g. Coins) are a common way of rewarding players for accomplishing certain objectives within the game. These types of reward schemes are a pretty standard practice within Serious Games [43, 50] and



also EPGs [10, 34, 44]. This paper argues that EPGs should learn from more entertainment based games, where player incentives go beyond scores or collectibles. Often referenced as “meaningful gamification” [47], is the idea that games should go beyond the superficial layer of point- or collectible-based systems, by providing a wider array of options or interactions for players to engage. For example, methods that empower players in some capacity as an alternative, allowing them to either overcome previously difficult obstacles or providing a temporary power boost. Such concepts could lead to powerful reactions from players by simply changing the current “status-quo” via certain mechanics. The opposite can also be done, such as weakening the player through in-game mechanics where the expectation would be to stimulate negative reactions.

#### 4.7 Dealing with Player Movement

Limited movement is a direct consequence of using physiological recording devices as they were never intentionally designed with participant movement in mind. Thus, it is important to keep in mind the devices being used, as players may either be covered in very sensitive wires and/or physically constrained to an abnormal playing position. The simplest method is constraining play to one hand only, while attaching wires to the non-dominant hand [10, 34]. Other options also include using more “friendly movement” devices such as bracelets, but at the cost of a less accurate signal. For motion based controllers such as VR the best solution is to attach dangling wires to the participant’s skin using some adhesive solutions. Furthermore, if possible avoid placing sensors directly on locations where a lot of the movement will take place (e.g. place SC sensors on feet instead of hands). Alternatively, experimenters can also look at using custom controllers or input that does not require any strong physical movements (e.g. eye-tracking).

#### 4.8 Dealing with Player Skill and Tutorials

To the best of our knowledge, very little literature exists on how to overcome the divergence between player skill. The most common method is to simply filter participants based on playtime questionnaires before experimentation starts [28]. This may be necessary if skill is an important factor for the overall study. However it is worth noting that all data collected will be constrained to that specific population type. If skill is not a factor, it is still worth considering in the game design as experimenters may want a similar experience for all participants. Interestingly, the literature rarely suggests using difficulty adjustment parameters to factor this problem. Ideally, experimenters would pre-screen each participant and adjust the difficulty accordingly to that individuals skill a priori. It can indeed be time consuming, as it does depend specifically on the complexity of the game and thus fine tuning its parameters. Other solutions can include making the game more accessible by reducing complexity, such as limiting the game to one or two inputs and mechanics.

Even though most experimenters keep in mind that a training phase is necessary [10, 39, 44], this paper advocates that it is still important to mention that a training stage or tutorial is essential for the majority of experiments. Unless experimenters want to study a players first experience with the game, there is no other practical reason to forgo a simple training phase. This is more efficient than

simply explaining the rules and mechanics, as players will have the opportunity to familiarize themselves with the different concepts of the game and controls.

## 5 CONCLUSIONS

Games can be a powerful method of inducing emotional reactions or motivating individuals for the study of certain human-based phenomena. Although games have been used extensively within research, it is rarely designed around the limitations of statistical analysis, the experimental protocol or the devices used for recording this data. From a game design perspective the solutions presented in this paper are straightforward, however they are rarely applied in practice and current research scenarios. Thus, this paper offers an initial insight and overview of aspects that experimenters should keep in mind when attempting to use games for experimentation while dealing with physiological, posture and experimental protocol limitations.

## ACKNOWLEDGMENTS

This research is supported by the SNF grant CRSII5\_180319 / 1. The authors would also like to thank Dr. Joana Leitão for the invaluable feedback and insight over the course of this paper.

## REFERENCES

- [1] Gazihan Alankus, Amanda Lazar, Matt May, and Caitlin Kelleher. 2010. Towards customizable games for stroke rehabilitation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2113–2122.
- [2] Dominik R Bach, Marc Guitart-Masip, Pau A Packard, Júlia Miró, Mercè Falip, Lluís Fuentemilla, and Raymond J Dolan. 2014. Human hippocampus arbitrates approach-avoidance conflict. *Current Biology* 24, 5 (2014), 541–547.
- [3] Diana Batista, Hugo Silva, and Ana Fred. 2017. Experimental characterization and analysis of the BiTolino platforms against a reference device. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2418–2421.
- [4] Mathias Benedek and Christian Kaernbach. 2010. Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology* 47, 4 (2010), 647–658.
- [5] Gary G Berntson, J Thomas Bigger Jr, Dwain L Eckberg, Paul Grossman, Peter G Kaufmann, Marek Malik, Haikady N Nagaraja, Stephen W Porges, J Philip Saul, Peter H Stone, et al. 1997. Heart rate variability: origins, methods, and interpretive caveats. *Psychophysiology* 34, 6 (1997), 623–648.
- [6] Boyan Bontchev. 2016. Adaptation in affective video games: A literature review. *Cybernetics and Information Technologies* 16, 3 (2016), 3–34.
- [7] 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*. ACM, 115–120.
- [8] Richard B Buxton. 2009. *Introduction to functional magnetic resonance imaging: principles and techniques*. Cambridge university press.
- [9] Luigi Cardamone, Georgios N Yannakakis, Julian Togelius, and Pier Luca Lanzi. 2011. Evolving interesting maps for a first person shooter. In *European Conference on the Applications of Evolutionary Computation*. Springer, 63–72.
- [10] Guillaume Chanel, Cyril Rebetez, Mireille Bétrancourt, and Thierry Pun. 2011. Emotion assessment from physiological signals for adaptation of game difficulty. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 41, 6 (2011), 1052–1063.
- [11] Alice Chirico, David B Yaden, Giuseppe Riva, and Andrea Gaggioli. 2016. The potential of virtual reality for the investigation of awe. *Frontiers in psychology* 7 (2016), 1766.
- [12] Andrea Clerico, Cindy Chamberland, Mark Parent, Pierre-Emmanuel Michon, Sebastien Tremblay, Tiago H Falk, Jean-Christophe Gagnon, and Philip Jackson. 2016. Biometrics and classifier fusion to predict the fun-factor in video gaming. In *2016 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 1–8.
- [13] Mihaly Csikszentmihalyi. 1997. *Finding flow: The psychology of engagement with everyday life*. Basic Books.
- [14] Heather Desurvire and Magy Seif El-Nasr. 2013. Methods for game user research: Studying player behavior to enhance game design. *IEEE computer graphics and applications* 33, 4 (2013), 82–87.

- [15] Anders Drachen, Lennart E Nacke, Georgios Yannakakis, and Anja Lee Pedersen. 2010. Correlation between heart rate, electrodermal activity and player experience in first-person shooter games. In *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games*. ACM, 49–54.
- [16] Paul Ekman. 1992. An argument for basic emotions. *Cognition & emotion* 6, 3-4 (1992), 169–200.
- [17] Anna Felnhöfer, Oswald D Kothgassner, Mareike Schmidt, Anna-Katharina Heinze, Leon Beutl, Helmut Hlavacs, and Ilse Kryspin-Exner. 2015. Is virtual reality emotionally arousing? Investigating five emotion inducing virtual park scenarios. *International journal of human-computer studies* 82 (2015), 48–56.
- [18] Andrea Finke, Alexander Lenhardt, and Helge Ritter. 2009. The MindGame: a P300-based brain-computer interface game. *Neural Networks* 22, 9 (2009), 1329–1333.
- [19] Tracy Fullerton, Chris Swain, and Steven Hoffman. 2004. *Game design workshop: Designing, prototyping, & playtesting games*. CRC Press.
- [20] D. Gabana, L. Tokarchuk, E. Hannon, and H. Gunes. 2017. Effects of valence and arousal on working memory performance in virtual reality gaming. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. 36–41.
- [21] C Shawn Green and Daphne Bavelier. 2012. Learning, attentional control, and action video games. *Current biology* 22, 6 (2012), R197–R206.
- [22] R Henson. 2007. Efficient experimental design for fMRI. *Statistical parametric mapping: The analysis of functional brain images* (2007), 193–210.
- [23] R Henson and K Friston. 2007. Convolution models for fMRI. *Statistical parametric mapping: The analysis of functional brain images* (2007), 178–192.
- [24] Hunter G Hoffman, Gloria T Chambers, Walter J Meyer III, Lisa L Arceneaux, William J Russell, Eric J Seibel, Todd L Richards, Sam R Sharar, and David R Patterson. 2011. Virtual reality as an adjunctive non-pharmacologic analgesic for acute burn pain during medical procedures. *Annals of Behavioral Medicine* 41, 2 (2011), 183–191.
- [25] Christoffer Holmgård, Georgios N Yannakakis, Karen-Inge Karstoft, and Henrik Steen Andersen. 2013. Stress detection for ptsd via the startlemaart game. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. IEEE, 523–528.
- [26] Simon Järvelä, Inger Ekman, J Matias Kivikangas, and Niklas Ravaja. 2014. A practical guide to using digital games as an experiment stimulus. *Transactions of the Digital Games Research Association* 1, 2 (2014).
- [27] Kostas Karpouzis, Georgios N Yannakakis, Noor Shaker, and Stylianos Asteriadis. 2015. The platformer experience dataset. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 712–718.
- [28] Jari Kätsyri, Riitta Hari, Niklas Ravaja, and Lauri Nummenmaa. 2013. The opponent matters: elevated fMRI reward responses to winning against a human versus a computer opponent during interactive video game playing. *Cerebral Cortex* 23, 12 (2013), 2829–2839.
- [29] Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lilienthal. 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology* 3, 3 (1993), 203–220.
- [30] Young Youn Kim, Hyun Ju Kim, Eun Nam Kim, Hee Dong Ko, and Hyun Taek Kim. 2005. Characteristic changes in the physiological components of cybersickness. *Psychophysiology* 42, 5 (2005), 616–625.
- [31] A. Kleinsmith, N. Bianchi-Berthouze, and A. Steed. 2011. Automatic Recognition of Non-Acted Affective Postures. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 41, 4 (Aug 2011), 1027–1038. <https://doi.org/10.1109/TSMCB.2010.2103557>
- [32] Do-Young Kwon, Byung Kyu Park, Ji Won Kim, Gwang-Moon Eom, Junghwa Hong, Seong-Beom Koh, and Kun-Woo Park. 2014. Quantitative electromyographic analysis of reaction time to external auditory stimuli in drug-naïve Parkinson's disease. *Parkinson's Disease* 2014 (2014).
- [33] Joshua Lewis, Patrick Trinh, and David Kirsh. 2011. A corpus analysis of strategy video game play in starcraft: Brood war. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 33.
- [34] Changchun Liu, Pramila Agrawal, Nilanjan Sarkar, and Shuo Chen. 2009. Dynamic difficulty adjustment in computer games through real-time anxiety-based affective feedback. *International Journal of Human-Computer Interaction* 25, 6 (2009), 506–529.
- [35] Keith Lohse, Navid Shirzad, Alida Verster, Nicola Hodges, and HF Machiel Van der Loos. 2013. Video games and rehabilitation: using design principles to enhance engagement in physical therapy. *Journal of Neurologic Physical Therapy* 37, 4 (2013), 166–175.
- [36] Phil Lopes. 2017. *Generating Multifaceted Content in Games: A Study on Levels and Sound*. PhD Dissertation. Institute of Digital Games of the University of Malta.
- [37] Phil Lopes, Antonios Liapis, and Georgios N Yannakakis. 2015. Targeting horror via level and soundscape generation. In *Eleventh Artificial Intelligence and Interactive Digital Entertainment Conference*.
- [38] Phil Lopes, Georgios N Yannakakis, and Antonios Liapis. 2017. RankTrace: Relative and unbounded affect annotation. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 158–163.
- [39] Julien Marengo, Phil Lopes, and Ronan Boulic. 2019. On the Influence of the Supine Posture on Simulation Sickness in Virtual Reality. In *2019 IEEE Conference on Games (CoG)*. IEEE, 1–8.
- [40] Cameron McCarthy, Nikhilesh Pradhan, Calum Redpath, and Andy Adler. 2016. Validation of the Empatica E4 wristband. In *2016 IEEE EMBS International Student Conference (ISC)*. IEEE, 1–4.
- [41] David Melhart, Georgios N Yannakakis, and Antonios Liapis. 2020. I Feel I Feel You: A Theory of Mind Experiment in Games. *arXiv preprint arXiv:2001.08656* (2020).
- [42] Ben Meuleman and David Rudrauf. 2018. Induction and profiling of strong multi-componential emotions in virtual reality. *IEEE Transactions on Affective Computing* (2018).
- [43] David R Michael and Sandra L Chen. 2005. *Serious games: Games that educate, train, and inform*. Muska & Lipman/Premier-Trade.
- [44] M. Moghimi, R. Stone, and P. Rotshtein. 2017. Affective Recognition in Dynamic and Interactive Virtual Environments. *IEEE Transactions on Affective Computing* (2017), 1–1.
- [45] Daniel Muriel and Garry Crawford. 2018. *Video games as culture: considering the role and importance of video games in contemporary society*. Routledge.
- [46] Lennart Erik Nacke, Michael Kalyn, Calvin Lough, and Regan Lee Mandryk. 2011. Biofeedback Game Design: Using Direct and Indirect Physiological Control to Enhance Game Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) (CHI '11). Association for Computing Machinery, New York, NY, USA, 103–112.
- [47] Scott Nicholson. 2015. A recipe for meaningful gamification. In *Gamification in education and business*. Springer, 1–20.
- [48] Rosalind W Picard. 2000. *Affective computing*. MIT press.
- [49] Pramila Rani, Nilanjan Sarkar, and Changchun Liu. 2005. Maintaining optimal challenge in computer games through real-time physiological feedback. In *Proceedings of the 11th international conference on human computer interaction*, Vol. 58. 22–27.
- [50] Ute Ritterfeld, Michael Cody, and Peter Vorderer. 2009. *Serious games: Mechanisms and effects*. Routledge.
- [51] J. J. Rivas, F. Orihuela-Espina, L. E. Sucar, A. Williams, and N. Bianchi-Berthouze. 2019. Automatic Recognition of Multiple Affective States in Virtual Rehabilitation by Exploiting the Dependency Relationships. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. 1–7.
- [52] James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161.
- [53] José Luis González Sánchez, Francisco Luis Gutiérrez Vela, Francisco Montero Simarro, and Natalia Padilla-Zea. 2012. Playability: analysing user experience in video games. *Behaviour & Information Technology* 31, 10 (2012), 1033–1054.
- [54] G Sartory, S Rachman, and S Grey. 1977. An investigation of the relation between reported fear and heart rate. *Behaviour Research and Therapy* (1977).
- [55] Jesse Schell. 2019. *The Art of Game Design: A book of lenses*. AK Peters/CRC Press.
- [56] M. Song, Z. Yang, A. Baird, E. Parada-Cabaleiro, Z. Zhang, Z. Zhao, and B. Schuller. 2019. Audiovisual Analysis for Recognising Frustration during Game-Play: Introducing the Multimodal Game Frustration Database. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. 517–523.
- [57] Frithjof Staude-Müller, Thomas Bliesener, and Stefanie Luthman. 2008. Hostile and hardened? An experimental study on (de-) sensitization to violence and suffering through playing video games. *Swiss journal of psychology* 67, 1 (2008), 41–50.
- [58] Shravani Sur and VK Sinha. 2009. Event-related potential: An overview. *Industrial psychiatry journal* 18, 1 (2009), 70.
- [59] Nana Tian, Romain Clement, Phil Lopes, and Ronan Boulic. 2020. On the Effect of the Vertical Axis Alignment on Cybersickness and Game Experience in a Lying-Down Posture. In *2020 IEEE Conference on Games (CoG)*. IEEE, 1–8.
- [60] Julian Togelius, Sergey Karakovsky, and Robin Baumgarten. 2010. The 2009 mario ai competition. In *IEEE Congress on Evolutionary Computation*. IEEE, 1–8.
- [61] J Rick Turner, Douglas Carroll, and Hazel Courtney. 1983. Cardiac and Metabolic Responses to "Space Invaders": An Instance of Metabolically-Exaggerated Cardiac Adjustment? *Psychophysiology* 20, 5 (1983), 544–549.
- [62] Qiang Wang, Olga Sourina, and Minh Khoa Nguyen. 2010. Eeg-based "serious" games design for medical applications. In *2010 International Conference on Cyberworlds*. IEEE, 270–276.
- [63] David A Washburn. 2003. The games psychologists play (and the data they provide). *Behavior Research Methods, Instruments, & Computers* 35, 2 (2003), 185–193.
- [64] J. Whitehill, Z. Serpell, Y. Lin, A. Foster, and J. R. Movellan. 2014. The Faces of Engagement: Automatic Recognition of Student Engagement from Facial Expressions. *IEEE Transactions on Affective Computing* 5, 1 (Jan 2014), 86–98.
- [65] Leanne M Williams, Pritha Das, Belinda Liddell, Gloria Olivieri, Anthony Peduto, Michael J Brammer, and Evian Gordon. 2005. BOLD, sweat and fears: fMRI and skin conductance distinguish facial fear signals. *Neuroreport* 16, 1 (2005), 49–52.
- [66] Sijing Wu, Cho Kin Cheng, Jing Feng, Lisa D'Angelo, Claude Alain, and Ian Spence. 2012. Playing a first-person shooter video game induces neuroplastic

- change. *Journal of cognitive neuroscience* 24, 6 (2012), 1286–1293.
- [67] Georgios N Yannakakis, Héctor P Martínez, and Arnav Jhala. 2010. Towards affective camera control in games. *User Modeling and User-Adapted Interaction* 20, 4 (2010), 313–340.
- [68] Georgios N Yannakakis and Julian Togelius. 2011. Experience-driven procedural content generation. *IEEE Transactions on Affective Computing* 2, 3 (2011), 147–161.