Challenges in Conducting Theory-Informed Empirical Player Experience Research

Ph.D. Dissertation

Jan B. Vornhagen

ITU Copenhagen
Digital Design Department
Rued Langgaards vej 7
DK-2330 Copenhagen S

This thesis was in part funded by the European Union (ERC, THEORYCRAFT, 101043198). Views and opinions expressed are however those of the author only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Abstract

Empirical player experience (PX) research has seen increased calls for more and better theory-use. Theory drives research programs, allows to synthesize research and can inform design. These calls, however, rarely discuss the actual practice of doing theory-informed research. How do we work effectively with theory in PX research? What challenges do we face; and, should we not overcome them, what does it mean for empirical PX research to fail? In this thesis, I approach these questions by presenting five research projects and reflecting on their respective research processes, challenges, and failures.

Publication I is a systematic literature review of the reporting practices of Null Hypothesis Significance Testing at CHI PLAY. A method commonly used to test theories and widely used in empirical PX research. Based on the findings of Publication I, and a discussion of confirmatory, theory-testing methods, I highlight the role of exploratory methods in PX research.

Publication II is an exploratory mixed-methods study on art experiences in videogames. The study utilizes empirical aesthetics to understand what players consider art experiences. It highlights the importance of epistemic emotions — i.e., Interest and Insight — as well as a range of short and long term impacts the experiences had on players. As this paper employs a theory devised to explain and predict the art experience of people, I reflect on the process of bringing the theory over to player experience research by discussing how we used it in this paper and how it guided the study design and codebook-based analysis.

Publication III studies empowering and disempowering player experiences. While empowerment is a common, desirable goal for HCI and HCI games research, it has a multitude of conceptualizations and can overlap with other commonly used PX constructs such as emotional play experiences and basic need satisfaction. I reflect on the overlap of different PX constructs with empowerment and disempowerment and how theory guided the choice of behavioral measures and thematic analysis.

The last two projects I discuss have not been published. In this thesis I reflect on the process of each project, the challenges they encountered, and what to learn from them. The first project

sought to experimentally evoke emotional challenge and compare if Self-Determination Theory or an art appraisal theory better predict the enjoyment and appreciation. After conducting the study, results revealed mostly non-significant results, with further Bayesian analyses indicating that the desired experimental manipulation was unreliable, throwing into question what can be learned from the results. The second project is an ongoing exploratory experiment, investigating the player experience of rewards through the lens of Cognitive Evaluation Theory. A pre-study raised concerns about measuring the functional significance and reliability of the study manipulation, leading to a reassessment of the feasibility of the study design. I will reflect on how these two studies were designed highlighting difficulties in translating theories into both testable hypotheses and experimental manipulations. I further discuss issues when measuring concepts and when attempting to decide between the study being wrong, or the theory.

This thesis contributes to PX research in the following ways. First, this thesis highlights tangible opportunities and approaches to better theory-informed PX research. Second, it describes concrete barriers and pitfalls to creating informative theory-informed PX research. Third, I give a personal account on failure in my research practice, something that otherwise is scarcely talked about, in the hope of sparking more reflection and discussion about HCI and PX research practice.

Resumé

Empirisk forskning i spilleroplevelser har set øgede krav om mere og bedre teoribrug. Teori driver forskningsprogrammer, gør det muligt at syntetisere forskning og kan informere design. Disse opfordringer diskuterer dog sjældent den faktiske praksis med at lave teoriinformeret forskning. Hvordan arbejder vi effektivt med teori i PX-forskning? Hvilke udfordringer står vi over for; og hvis vi ikke overvinder dem, hvad betyder det så for empirisk PX-forskning at mislykkes? I denne afhandling nærmer jeg mig disse spørgsmål ved at præsentere fem forskningsprojekter og reflektere over deres respektive forskningsprocesser, udfordringer og fiaskoer.

Publikation I er en systematisk litteraturgennemgang af rapporteringspraksisserne for Null Hypothesis Significance Testing på CHI PLAY. En metode, der almindeligvis bruges til at teste teorier og er meget udbredt i empirisk PX-forskning. Baseret på resultaterne af publikation I og en diskussion af bekræftende, teoritestende metoder fremhæver jeg den rolle, som eksplorative metoder spiller i PX-forskning. Publikation II er et eksplorativt mixed-methods studie af kunstoplevelser i videospil. Undersøgelsen bruger empirisk æstetik til at forstå, hvad aktører betragter som kunstoplevelser. Det fremhæver vigtigheden af epistemiske følelser — det vil sige interesse og indsigt — samt en række kort- og langsigtede virkninger, som oplevelserne havde på spillerne.

Da denne artikel anvender en teori, der er udtænkt til at forklare og forudsige menneskers kunstoplevelser, reflekterer jeg over processen med at bringe teorien over til forskning i spilleroplevelser ved at diskutere, hvordan vi brugte den i denne artikel, og hvordan den styrede undersøgelsesdesignet og kodebogsbaseret analyse.

Publikation III studerer empowering og disempowering oplevelser. Mens empowerment er et almindeligt, ønskværdigt mål for HCI- og HCI-spilforskning, har det et væld af konceptualiseringer og kan overlappe med andre almindeligt anvendte PX-konstruktioner såsom følelsesmæssige legeoplevelser og grundlæggende behovstilfredsstillelse. Jeg reflekterer over overlappet mellem forskellige PX-konstruktioner med empowerment og disempowerment, og hvordan teorien styrede valget af adfærdsmæssige mål og tematisk analyse.

De sidste to projekter, jeg diskuterer, er ikke blevet offentliggjort. I dette speciale reflekterer jeg over processen i hvert projekt, de udfordringer, de stødte på, og hvad man kan lære af dem. Det første projekt forsøgte eksperimentelt at fremkalde følelsesmæssige udfordringer og sammenligne, om Self-Determination Theory eller en kunstvurderingsteori bedre forudsiger nydelsen og påskønnelsen. Efter at have udført undersøgelsen afslørede resultaterne for det meste ikke-signifikante resultater, hvor yderligere Bayesianske analyser indikerede, at den ønskede eksperimentelle manipulation var upålidelig, hvilket satte spørgsmålstegn ved, hvad der kan læres af resultaterne. Det andet projekt er et igangværende udforskende eksperiment, der undersøger spilleroplevelsen af belønninger gennem linsen af kognitiv evalueringsteori. En forundersøgelse gav anledning til bekymring med hensyn til måling af den funktionelle betydning og pålidelighed af undersøgelsesmanipulationen, hvilket førte til en revurdering af gennemførligheden af undersøgelsens design. Jeg vil reflektere over, hvordan disse to undersøgelser blev designet og fremhæve vanskeligheder med at oversætte teorier til både testbare hypoteser og eksperimentelle manipulationer. Jeg diskuterer yderligere problemstillinger, når jeg måler begreber, og når jeg forsøger at afgøre, om undersøgelsen er forkert, eller teorien.

Denne afhandling bidrager til PX-forskningen på følgende måder. For det første fremhæver denne afhandling håndgribelige muligheder og tilgange til bedre teoriinformeret PX-forskning. For det andet beskriver den konkrete barrierer og faldgruber for at skabe informativ teoriinformeret PX-forskning. For det tredje giver jeg en personlig redegørelse for fiasko i min forskningspraksis, noget der ellers næsten ikke tales om, i håb om at sætte gang i mere refleksion og diskussion om HCI og PX-forskningspraksis.

Contents

Thesis Details

Thesis Title:	Challenges in Conducting Theory-Informed Empirical Player Experience Research
Ph.D. Student:	Jan B. Vornhagen
Supervisors:	Prof. Elisa D. Mekler, IT University of Copenhagen

The main body of this thesis contains a summary and discussion of the following (shared) first-author publications.

Pub. I: Jan B. Vornhagen, April Tyack, and Elisa D. Mekler. 2020. Statistical Significance Testing at CHI PLAY: Challenges and Opportunities for More Transparency. In Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '20), 4–18. https://doi.org/10.1145/3410404.3414229

Pub. II: Julia A. Bopp, Jan B. Vornhagen, and Elisa D. Mekler. 2021. "My Soul Got a Little Bit Cleaner": Art Experience in Videogames. Proceedings of the ACM on Human-Computer Interaction 5, CHI PLAY: 237:1-237:19. https://doi.org/10.1145/3474664

Pub. III: Jan B. Vornhagen, Dan Bennett, Dooley Murphy, and Elisa D. Mekler. 2023. "I'm the leader and I'm going to save the world": Characterizing Empowering and Disempowering Game Experiences. Proceedings of the ACM on Human-Computer Interaction 7, CHI PLAY: 1330–1360.
https://doi.org/10.1145/3611071

This thesis has been submitted for assessment in partial fulfillment of the PhD degree. The thesis is based on the published scientific papers which are listed above. Parts of the papers are used directly or indirectly in the extended summary of the thesis. As part of the assessment, co-author statements have been made available to the assessment committee and are also available at the Faculty.

The thesis further references two unpublished project for which materials are available.

Proj. I: Jan B. Vornhagen and Elisa D. Mekler. 2021. The Origin of EC. Retrieved December 21, 2024 from https://osf.io/rnbse/?view_only=d59e36415fe3418c8d6ebaacd264de6f

Proj. II: Jan B. Vornhagen, Roosa Piitulainen, and Elisa D. Mekler. 2024. Tangible Rewards in Videogames. Retrieved December 21, 2024 from https://osf.io/zgb9d/?view_only=8e8099e913174e699dce9c827172d8fc

In addition to the main projects, the following publications have also been made. Name in bold indicates co-first-authorship.

[1] Lahari Goswami, Amelia McNamara, Viktorija Paneva, **Jan B. Vornhagen**, and Erich Weichselgartner. 2022. A Cheat Sheet for a Transparent CHI paper. https://doi.org/10.17605/OSF.IO/YHWUQ

[2] Claire Reymond, Jan B. Vornhagen, Matthew Pelowski, Klaus Opwis, and Elisa D. Mekler. 2023. Images influencing images: How pictorial context affects the emotional interpretation of art photographs. Psychology of Aesthetics, Creativity, and the Arts: No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/aca0000523

[3] Raquel B. Robinson, **Jan B. Vornhagen**, Guo Freeman, Brendan Keogh, Regan L. Mandryk, and Brendan Sinclair. 2024. The Impact of Mass Game Industry Layoffs on the CHI PLAY Community. In Companion Proceedings of the 2024 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY Companion '24), 475–477. https://doi.org/10.1145/3665463.3678861

[4] Yvonne Jansen, Jan B. Vornhagen, Olga Iarygina, Kavous Salehzadeh Niksirat, Lonni Besançon, Pierre Dragicevic, Julien Gori, and Chat Wacharamanotham. 2024. The Many Ways of Being Transparent in Human-Computer Interaction Research. https://doi.org/10.31219/osf.io/2wze6

Glossary

List of Abbreviations

ACM Association of Computing Machinery

AESTHEMOS Aesthetic Emotions Scale

CET Cognitive Evaluation Theory

CHI Conference on Human Factors in Computing Systems

CHI PLAY Annual Symposium on Computer-Human Interaction in Play

GDC Game Developers Conference

HCI Human-Computer Interaction

IMI Intrinsic Motivation Invetory

NHST Null Hypothesis Significance Testing

OS Open Science

PENS Player Experience of Need Satisfaction

PX Player Experience

PXI Player Experience Inventory

QRPs Questionable Research Practices

SDT Self-Determination Theory

SIG Special Interest Group

TA Thematic Analysis

VIMAP Vienna Integrated Model of top-down and bottom-up processes in Art Perception

List of Figures

List of Tables

Acknowledgment

This thesis and the research therein could have never been done with the help of so many amazing people.

First, I want to thank my wonderful wife Maija. I can't put in words how much you helped me in these past years and I am so sorry for all the stress I put you through. Ich möchte meinen Eltern Petra und Volker danken, für all die Jahre an unerschütterlicher Unterstützung sowohl bei diesem, als auch bei all den anderen dummen Dingen die ich mir über die Jahre in den Kopf gesetzt habe. Vielen Dank an Yvonne, die meine Faszination für Psychologie geweckt und mich auf diesen Pfad gesetzt hat. And thank you Tammo, for all the helpful comments on this thesis and hours of conspiratorial lefty rants about neoliberalism, interspersed with co-op games.

Likewise, I could have not done any of this without my colleagues. Thank you Elisa. For giving me this chance and your support of all my work as well as all my Open Science flights of fancy. You always pushed me to do my best. Thank you also, to my colleagues and co-authors, Julia, Roosa, Dan, Feng, Dooley, and Raquel. Thank you for all your help in making our research projects reality and your contributions to this thesis.

I also want to thank my communities: all those people at ITU, who made my time in Copenhagen bearable, as well as a special big thank you to the ReproducibiliTea community and all the awesome people I had the honor of working with on the Steering Committee.

And Thank you April. You taught me more about actually doing academic work than anybody else. I wish you could have taught me more.

Thank you all. I could have not done any of this without any of you.

Jan B. Vornhagen
Copenhagen, 31.12.2024

Part I

Thesis

Chapter 1

Introduction

Player Experience (PX) research is a strand of Human-Computer Interaction (HCI) which studies how people experience games on a (socio-)psychological, behavioral, and physiological level [223]. The focus of PX lies not on the technological level, but *"the individual and personal experience of playing games"* [223, p.246]. To understand, measure, and foster player experience, PX research heavily builds on HCI and psychology [223].

In fact, theory is considered highly important in empirical PX research: It is used to lay the foundation and provides the tools to determine, measure, and build good PX [223]. The player experience is explained and structured via the theories we use, as they provide our lenses and determine the questions we ask of players [1]. For example Self-Determination Theory (SDT) explains motivation for humans in general [184], but has been used to conceptualize the player experience as "good" when it fosters intrinsic motivation by satisfying the players need for competence, autonomy, and relatedness [223]. Another theory used in PX [223] is Flow, which describes how people can get lost in any interesting experience [50] and instead lends itself to conceptualizing the player experience in terms of skill-challenge balance [see 49].

Moreover, theories aim to make recommendations for design, based on how they conceptualize PX. For example, SDT concepts, such as Need Satisfaction [e.g., 173], as well as, Cognitive Evaluation Theory (CET) [e.g., 170] have been offered to game design practitioners as a way to predict player reactions to their design choices [209].

As theories have many uses, *how* they are used has long been a point of interest in HCI [124]. Theories are argued to drive *"research programs, resulting in insights and enabling new discoveries"*[176, p. 15] as they provide *"the means to analyze and predict the performance of users carrying out tasks for specific kinds of computer interfaces and systems"* [176, p. 16]. According to Halverson

[95], theories can make sense of a domain by providing descriptive power — i.e., they allow us to make sense of and describe phenomena — rhetorical power — i.e., they allow to talk about phenomena in a clear and persuasive manner — inferential power — i.e., provide guidance and insight into a phenomenon — and application power — i.e., inform design. They can guide research by highlighting certain concepts (and toning down others), opening up diverging avenues for research [95]. Outlasvirta & Hornbæk [149] argue that theory can direct design towards desirable outcomes by fostering thought experiments. Kostakos [115] argues that building and iterating of theory can foster motor themes — i.e., central topics we accumulate robust knowledge about [123].

Despite these arguments that theory is useful, a common claim by HCI scholars is that the field does not utilize theory well [13, 40, 39]. One reason brought forward for this, may be that cognitive theories about human behavior are too general [96, 176], or deemed not suitable for the iterative design of practitioners [39, 150]. Consequently both HCI [e.g., 13, 15, 78, 124, 149] and PX [e.g., 9, 59, 209] scholars have increasingly called for more and better engagement with theory.

This being said, there are few studies investigating the actual use of theory in HCI. Hekler et al. [96] found that HCI researchers use theory to inform design, guide evaluation, and define their target user, but also found that that HCI researchers tended to only use some of the constructs theories provide, while overgeneralizing their claims. Similarly, Tyack & Mekler [208, 209] found that HCI games researchers primarily use only selected concepts of SDT, while further pointing out that fort most studies that mention SDT or use SDT based measures the theory is *"ultimately inconsequential to the research"* [209, p.40:42].

Moreover, recommendations for better theory use [e.g., 16, 40, 149, 203] primarily focus on design research. i.e., give recommendations for how to use theory to design better technological solutions. These recommendations are not addressed towards empirical research, which may seek to inform design [48], but does not produce design artifacts itself.

Those recommendations that do address empirical PX research [9, 209] remain at a rather high level, however, and leave open how they would work in practice. There are no accounts of the process of using theory in empirical PX research. What does it look like to design a theory-informed study? How does theory guide our analysis and shape our discussions? And what do we do when we face challenges or even fail when using theory?

This last question is pertinent, as failure is typically considered undesirable in research [100], yet is nominally core to many research areas. For example, falsification, i.e., failing to support a theory, being the core of the hypothetico-deductive model of science [164]. In HCI, Hornbæk [98] has argued that failing to find proof justifying our beliefs can be highly informative, as it pushes the field towards more interesting questions. In design research [157] and creative practice [196],

failure is an important step on the way to a successful project.

This thesis presents different ways theory may productively inform empirical PX research across five projects. It discusses the general practice of theory testing in empirical PX research, as well as utilizing theory to inform both exploratory and confirmatory work. It provides accounts of how studies were prepared, designed, and analyzed, focusing on theory-informed decisions made during this process. Crucially, it also identifies instances where the application of theory has "failed" in some way.

1.1 Contributions

This thesis contributes to empirical player experience research by discussing theory-use in empirical PX research.

- In this thesis I describe ways to work with theory in four different research projects. I first discuss theory-testing in empirical PX research and argue for more theory-informed exploratory work. While past HCI works argue for more theory-use [e.g., 13, 15, 78, 124, 149], there are no accounts on how this looks like in practice. I discuss the use and shortcomings of theory-testing methods in empirical PX research, as well as other theory-uses in empirical PX research, which is more compatible with concurrent PX research.
- I reflect on my research projects to identify and discuss several barriers and challenges for theory-informed empirical PX research. So far, no works address the difficulties of trying to work theory-informed in empirical PX research. I discuss barriers in regards to translation (e.g., translating theoretical tenets into PX predictions), to measurements (e.g., a lack of validated questionnaires for the game setting), and to experimental design (e.g., implementing theory-based manipulation in experimental games). Based on the identified barriers, I list open questions and goals for future research.
- Lastly, I reflect on failure of theory-use in the research process. While done prior in regards to design research[e.g., 79, 100, 157] no works to date have addressed failure in the empirical PX research process. I discuss the difficulties that arise when encountering issues during the research process, when we seek to turn these projects into scientific output, and what we as a research community could do to learn from failure.

1.2 Outline of the Thesis

In Chapter 2, I start by providing background on the role of theory in HCI and discuss the recent calls for more theory use in HCI and PX research. In Chapter 3, I discuss the role of hypothesis testing when working with theories, followed by a discussion of the shortcomings of the approach, and the findings of *Pub. I*. I close that chapter by discussing exploratory research. In Chapter 4, I present an exploratory, theory-inspired work on the aesthetic videogame experience *Pub. II*. I embed the results within the PX literature, discuss how theory informed the work and how we adopted the theory from empirical aesthetics to PX research. In Chapter 5, I present the results of *Pub. III*, which describe the phenomenon of empowering and disempowering gaming experiences. I relate the results of a thematic analysis to other PX concepts — such as attribution theory and need frustration — and point out difficulties when trying to conceptualize phenomena. In Chapter 6, I present a hitherto unpublished research project on the experience of discomfort in videogames. I discuss the conceptualization and design of this experimental study. I reflect on the derivation of the hypotheses, the process from design to pre-study to final study, and the difficulties of creating the stimulus, as well as how the study was ultimately deemed to have failed. In Chapter 7, I present a currently on-going project which utilizes an exploratory experimental approach to study how different types of rewards impact the player experience. I discuss CET and the predictions we can derive from this theory in regards to the player experience. I discuss the concept of theory mapping and the issue of verbal theories. I conclude by presenting results from a pre-study, and discussing the difficulty of measuring phenomena in PX research. I close this thesis in Chapter 8 by reflecting on my experiences when working with theory in PX research. I point out pitfalls and open questions and aim to provide some guidance towards better theory-informed empirical PX research.

Chapter 2

Theory in Psychology, HCI, & PX

2.1 Theory in Psychology

Empirical science is dependent on theory. It is by way of theory that we perceive the world [164], organize and share knowledge [176], and determine what puzzles to solve [116]. Theory helps us reliably predict phenomena [80], enables us to organize and integrate data [85], and guides us towards important questions [116]. In general, theories have two roles in psychology [80]: Explanation — i.e., why do people do something — and prediction — i.e., what will people do in a certain situation.

Theory is a ubiquitous word in the behavioral sciences that can have several meanings [2]. Consequently, to narrow down the concept, this thesis will utilize the concept of *explanatory* theories by Borsboom et al. [31], who define it as a set of linked propositions, united under a general explanatory principle, that explain phenomena. Phenomena in turn are *"stable and general features of the world that scientists seek to explain"*[31, p.758]. They are generalizations based on direct observations — data. Consequently, theory explains phenomena which predicts data, while data generalizes to phenomena, from which we can abduct theory [31]. This relation between theory, phenomena, and data is illustrated in Figure 2.1.

The primary way psychologists bring the three levels of Data, Phenomena, and Theory together is the hypothetico-deductive model [31] in which theories are used to derive hypotheses which are then tested by collecting data. This data is then analyzed — most commonly using Null Hypothesis Significance Testing (NHST) [60, 86] — to either support or falsify the theory [80, 164].

While the hypothetico-deductive approach as it is performed today is vulnerable to both issues in how its performed [136] — which I will further discuss in Chapter 3 — and has been criticized for lacking in its ability to build theory [31], it does lead to psychology — and other behavioral sciences — to heavily depend on theory when doing research: To perform a psychological study, researchers need a theory from which to derive hypotheses they can test.

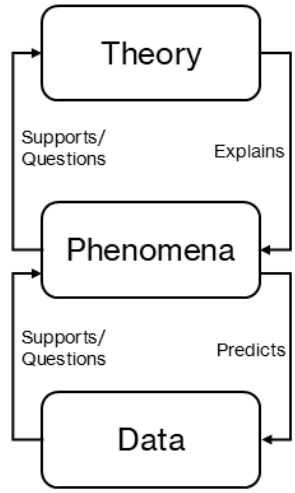

Fig. 2.1: Illustration of the relationship between theory, phenomena, and data.
Adapted from [31, 145]

2.2 Theory and Theory-Use in HCI

Theory-use — and the discussion thereof — has a long history in HCI. Early on, computer scientists saw the need to adopt cognitive theories to understand how users would experience newly developed soft- and hardware [39]. From its start, HCI was closely intertwined with technological progress [137, 176], and as technology became more ubiquitous, HCI researchers adopted theories from more and more disciplines to better address technology use in different contexts [176].

In HCI, theory can be used to *"establish common terminology, explain and contextualise research findings, predict outcomes, inform design practice and generate questions for further study"* [209, p.40:3]. While psychology forefronts theories' ability to explain and predict behavior, HCI additionally values theory for its ability to inform design [e.g., 149, 176]. Several HCI scholars have identified a range of uses for theory with different degrees of overlap. For example, Gregor [92] described goals and uses of theories in information systems as :

(1) descriptive — describing phenomena and their relationships ("What" questions)

(2) explanatory — explaining how effects come to be ("Why" questions)

(3) predictive — predicting what will happen in a given situation

(4) prescriptive — saying how a certain effect can be achieved

Bederson and Shneiderman [17] utilize essentially the same four categories, but additionally describe a use of theory for HCI practitioners by being:

(5) generative — enabling practitioners to create new things

Rogers [175, 176] further expands the list by adding theory being:

(6) informative — importing knowledge from another field

(7) ethnographic — being descriptive theory grounded in other disciplinary traditions

(8) conceptual — offering high-level frameworks for design

(9) critical — offering critique based on cultural and aesthetic concerns

Halverson [95] points out that, regadless of the specific use of a theory, they can have descriptive, rhetorical, inferential and application power. i.e., theories (to different degrees) allow us to describe, talk about, make decisions on, and design for specific phenomena.

Overall, while the exact uses of theories in HCI differ, they are deemed highly useful as tools that describe and explain phenomena, and allow HCI scholars to design and create artifacts to evoke them.

Yet, despite this apparent ubiquity and proposed usefulness of theory, HCI scholars have argued that theory is not used enough [13, 40, 39]. Particularly cognitive theories have been suspected to be too general to tackle the problems HCI usually seeks to solve [96, 176] and too unwieldy to be handled by practitioners [39]. HCI practitioners alleged inability to use theory has been described as an issue of translation, i.e., the difficulty of moving knowledge between scientific discovery and practice [48].

Translating general theory into something usable in design practice is a core goal of HCI [203], and many HCI scholars have proposed some method to create an intermediate concept [16, 101] between theory and design practice. For example, Sutcliffe [203] introduced *"generic models"* — theory-derived claims optimized to be reusable in different design projects. Similarly, Dalsgaard & Dindler [52] propose bridging concepts which take both theory and tangible design examples to form design articulations. Carroll & Rosson [40] introduced *"design rationale"*, a technique to explicitly couple theory with design decisions. A more encompassing approach to create theories catered to HCI is the concept of *"Generative Theories of Interaction* by Beaudouin-Lafon, Bødker & Mackay [15]: They define generative theories of interaction as theories which are (1) grounded in behavioral theories; (2) can be used to analytically describe a new technology, critique it based on needs and context, and can constructively inspire new ideas; (3) have concepts and generative principles that make it actionable.

Notably, these practices are primarily focused on making behavioral theories easier for designers to implement into their practice. Design practice, however, only makes up a subset of HCI [48, 231]. Those areas of HCI interested in the study of the user and evaluation of design [96, 112], not in creating new artifacts are scarcely catered to.

2.3 Theory-Use in PX research

One area that is more focused on users than design practice is empirical PX research, which seeks to understand how and why certain game systems create a certain experience for players [62, 223]. As such, PX research puts less focus on design practice and is instead more interested in the evaluation of experiences and game design [e.g., 1]. While theory has a wide variety of meaning and uses in HCI, theory-use in empirical PX is primarily discussed in terms deliminating a PX construct — i.e., defining what makes a good player experience.

To facilitate this, PX research is heavily built [223] on both general psychological theories — e.g., SDT [184] or Flow [50] — and game specific translations — e.g., the Player Experience of Need Satisfaction (PENS) [185] or GameFlow [204]. These theories are then used as tools to, for example evaluate how certain game elements impact the player experience — e.g., the effect of power-ups on players [57] — or explain the preferences of players for one game over another — e.g., using developmental psychology to understand player preferences [24].

This being said, like with mainline HCI, theory-use in games research has been criticized. Work by Tyack & Mekler [208, 209] found that use of SDT [184] in HCI games research was generally superficial, while Ballou [9] argued a lack of engagement with theory being a cause for psychological games research not making much progress towards understanding the effect of games on players.

Yet, recommendations for better theory-use in empirical PX research is rare. PX textbooks mention theory as a central element of our practice, but they rarely talk about how to use it in our research [e.g., 223]. The aforementioned work by Tyack & Mekler [209] argue, based on Beck & Stolterman [16], for deeper embedding of theory at all stages of the research process (e.g., for derivation of hypotheses, choice of methods, etc) and engaging in "talk-back" (i.e., relate findings back to the theory that informed the study [see also 211]), but likewise does not go into detail as to how this looks in practice. Ballou [9] noted, that much psychological games research is not well suited to conduct precise theory tests, and recommends to instead focus on research practices further down the "derivation chain" [130, 187] (see Chapter 3) — such as creating descriptions of phenomena, findings boundary conditions, and conduct exploratory experiments — in order to eventually build up robust theory covering PX, but also does not explain how this looks in practice. Those recommendations that are more practice oriented are, however, primarily focused on facilitating the hypothetico-deductive method — such as hypothesis formalization which outlines the process of operationalizing hypotheses [108] and the *Touchstone2* software that aids experimental design [65] — which is only a singular way to engage with theory in empirical PX research.

2.4 Summary

Overall, theory is deemed useful for PX research, but our theory-use is considered lackluster. While there are recommendations for improving how we engage with theory, it is unclear how this would look in practice, as most recommendations are high-level suggestions for research practices. In this thesis, I want to provide personal accounts of my own engagement with theory-informed empirical PX research from four research projects.

With this, I provide accounts of both successful and (so far) unsuccessful research projects; something that is rare in research [cf. 79, 212], and has never been done in empirical PX research. In fact, researchers have criticized that there is an expectation for research to be a success narrative [100]. As Yarkoni noted, it *"is an unfortunate cultural norm within psychology (and, to be fair, many other fields) to demand that every research contribution end on a wholly positive or 'constructive' note."*[233, p.14]. Moreover, acknowledging issues in our research is potentially detrimental to our chances to get published [97].

Yet, particularly in design research practice, some authors have discussed the benefits of reporting challenges and failure in their work. For example, Gaver et al. [79] discuss the failure of a technological intervention, particularly in regards to how they came to the conclusion that their artifact failed, and what knowledge they gained in the process. Similarly, in an auto-ethnography, Howell et al. [100] reflected on their research process and explicitly engaged with failure to open the discourse in design research towards acknowledging challenges as a way to *"reflect on and re-understand the specific situated roles we played in our past projects"* [100, p.42:3].

Still, accounts from inside the research process remain rare, despite that for empirical PX research they could help articulating the opportunities and challenges of theory-use, and subsequently, how our engagement with theory could be improved. In the rest of this thesis, I want to look at the practice of working with theory in empirical PX research. In the next chapters, I will first look at how empirical PX research primarily approaches theory — i.e., the hypothetico-deductive method using NHST — and discuss how useful it is for our field. I will then discuss four research projects at different stages (published, discontinued, and in progress), in regards to how they used theory and what challenges they faced.

Chapter 3

Theory Testing

As mentioned in the previous chapter, empirical Player Experience research is heavily built on psychological theories and methods [58, 112, 203, 223]. Consequently, to test theory, empirical PX research closely emulates the de-facto default psychological approach to theory testing [8, 60, 112]: Null Hypothesis Significance Testing (NHST). While NHST is widely used [60, 86], it has been increasingly criticized for being methodologically muddled [86], easy to use incorrectly [200, 222], and unsuited to test all but the most precise hypotheses [205].

In this chapter, I will first discuss the benefits and issues of NHST. I will then discuss *Pub. I*, a systematic literature review of NHST reporting at CHI PLAY — a central venue for empirical PX research. Based on the findings of *Pub. I*, I will argue that empirical PX research is often not aimed at testing theories, and consequently, using a method to test theory may be suboptimal for achieving our goals. Lastly, I will discuss exploratory research beyond NHST as a more useful approach to studying PX.

3.1 Null Hypothesis Significance Testing

NHST is a widespread method to perform inferential statistics, i.e., statistics to discern if a limited dataset generalizes to a population [60]. It is a central pillar of the *hypothetico-deductive* model of theory-testing as it is performed in psychology [86]: Researchers generate a hypothesis to test a theoretical claim, they design a study, collect data, and then use NHST to decide if the data supports or falsifies the hypothesis and therefore the theory [136]. As studies using NHST are commonly used to test hypotheses, they are considered *confirmatory* research [63].

In practice, using NHST entails first asking a research question, which is then "operationalized" by turning it into a statistical hypothesis: An if statement that directly corresponds to a measurement. For example, the research question "Do players enjoy challenge in videogames" can (among many other options) be operationalized to "If players enjoy challenge in videogames, then players in the high challenge condition will report higher enjoyment than players in the low challenge condition". This statistical hypothesis has two competing outcomes: Either, there is no effect of challenge on enjoyment — which commonly constituted the *null hypothesis (H_0)* — or there is an effect — which constitutes the *alternative hypothesis (H_1)* [139]. To decide between these competing hypothesis, researchers perform statistical tests — e.g., a t-test or an ANOVA [71] — which produce a *p-Value*. This p-Value represents the probability of the Data if the H_0 is true. If the p-Value is lower than the significance threshold, the H_0 is discarded in favor of H_1. If $p >= \alpha$, then the H_0 is retained [139].

NHST can be suitable to test theories as it allows for nominal control of two types of errors. The *Type I* error — falsely rejecting the null — and the *Type II* error — falsely *not* rejecting the null [60, 139]. By, for example, setting the Type I error rate to $\alpha = .05$ and Type II error rate to $\beta = .2$, we control that *in the long run* — i.e., when performing the same study over and over again on newly sampled data — we will falsely reject H_0 5 out of 100 times, and will falsely not reject H_0 20 out of a 100 times.

This is potentially powerful when testing theories, but has two major drawbacks in practice. First, as error rates are controlled in the long run the evidence of a single result is weak [60, 61, 63]: Controlling the Type I error at the usual α level of 0.05 means that with the given result, we would falsely accept the H_1 in question 5 times in 100 hypothetical studies, *not* that we have a less than 5% probability the H_1 is false [60]. Moreover, finding H_0 to be false does not also mean that H_1 is true [80]. In fact, it is only by our study design that H_1 should be the only other alternative. In practice, this usually means a study may find that there is not "no difference" (i.e., the H_0 is rejected) therefore there has to be a difference (i.e., H_1 is accepted). This has long been criticized as "weak" theorizing as this only ever requires to reject a single point (i.e., 0) [129].

Moreover, this puts heavy emphasis on replications — i.e., running the exact or conceptually same study over and over again to rule out false positives [192, 238]. Replications, however, are seldom done [99], as they are complex [e.g., 143, 192], and often a thankless task that may not result in publications [177, 178].

Second, while NHST allows for the nominal control of error levels, it is susceptible to both accidental and intentional inflation of said error level; to the point that scholars argued that it is not only possible, but trivially easy to make any test significant [200]. During the research process, a lot of decisions — such as how exactly to prepare the data — are left up to the researcher. These *researcher degrees of freedom* are numerous and often ambiguous, as there is usually not a

single correct answer to any of the decisions a researcher has to make during the study design, the data collection, the analysis and the write up of the final paper [200, 222].

In practice, a researcher may take a non-significant result, remove some outliers, or split the data by participant gender, rerun the test, and suddenly have a significant result [194, 200]. Utilizing degrees of freedom to increase the chances of getting a significant result have been coined *Questionable Research Practices (QRPs)*[105], and were the target of many of the methodological improvements proposed by the Open Science (OS) movement in the wake of *Replication Crisis* [for an overview, see 159].

The main starting points to reduce QRPs and improve the use of NHST was the promotion of *rigor* with which to perform and report statistical tests, and *transparency* of how the analysis was done. Simmons et al. [200] proposed 6 requirements for authors to reduce QRPs — such as a minimum sample size of 20 data points per condition, a preemptively decided stopping rule, and mandatory reporting of all observations, exclusion criteria, variables, and analyses. Wicherts et al. [222] extended this into a checklist of 34 degrees of freedoms that authors should report, separated into hypothesizing, design, collection, analysis, and reporting phases. Most recently, Aczel et al. [3] provided a 12 and 36 item, consensus-based based transparency checklist to foster transparency and therefore aid reproduction of results and facilitate research synthesis.

Furthermore, to facilitate the transparency of NHST — and research as a whole — researchers were increasingly encouraged to share their data, analysis scripts, and research artifacts, in order to allow for reproduction of the results and easier replication [136, 134]. This was made significantly easier by the rise of *FAIR* repositories [225] — repositories that make data findable, accessible, interoperable, and reusable, such as the *Open Science Framework*, or *Zenodo* — and heavily promoted by, for example, the TOP guidelines [141]. Moreover, practices like preregistration [142] — in which researcher degrees of freedom are codified before the study is conducted — registered reports [188] — in which a project is peer-reviewed before data collection, and accepted in principle for publication — and open peer-review [182] — in which the peer review process is made comprehensible from the outside — were pushed for and established.

3.2 NHST at CHI PLAY

Like with theory-use, the use of NHST has long been criticized in HCI [e.g., 37]. Since the early 2010, the CHI community has formally organized to promote better statistical practices, for example via the *RepliCHI* Special Interest Group (SIG) [227], panel [229], and workshops [230, 228], and the *Transparent Statistics in HCI* SIG [113]. Likewise, several works have pointed out methodological issues around the use of NHST [22, 37, 44, 112] and the lack of transparency

and artifact sharing [5, 45, 44, 216] in the community. Moreover, several works have attempted to translate the psychology focused OS practices and principles over to HCI [e.g., 23, 21, 104], and facilitate their adoption [e.g., 63, 88, 217]. Still, statistical significance is often the primary way in HCI to validate our research claims [44].

Despite these calls for improved methodology around NHST and more transparency, how much and how well it is used in empirical PX research has remained mostly speculative. Consequently, to understand how empirical PX researchers use and report their research, we conducted a systematic literature review — "Statistical Significance Testing at CHI PLAY: Challenges and Opportunities for More Transparency" (*Pub. I*). In it, we reviewed the entire full paper corpus published at Annual Symposium on Computer-Human Interaction in Play (CHI PLAY) — a central conference for HCI games research[1] — investigating how it compares to the NHST reporting best practice laid out by Wicherts et al. [222].

The goal of this work was two-fold: First, to take stock of how we report on our studies in HCI games research and investigate the uptake of OS and transparent reporting practices. Second, to create a resource communicating transparent statistical reporting to both authors and reviewers, facilitating a shared expectation of what should be included when reporting inferential tests.

Using the PRISMA-P protocol [195], *Pub. I* investigated the entire CHI PLAY corpus. Of all full papers published at the conference between 2014 and 2019 ($n = 246$), slightly less than half ($n = 119$) used NHST. Our analysis revealed that inferential quantitative analyses at CHI PLAY were exclusively done utilizing NHST, indicating a similar monopoly on quantitative analysis as in psychology [60, 86]. In regards to the reporting, many studies did not include information such as a power analysis or other forms of sample size justifications (94.1%), did not report assumption tests (53.8%) — required for the validity of parametric tests — or reported incomplete test statistics (31.1%). The results of the analyzed papers often lacked effect sizes (37%), and rarely utilized non-dichotomous reporting measures (6.7%). Moreover OS practices like data sharing (5%), or pre-registration (0.8%) were exceedingly rare.

While our results did not allow us to make any inferences about the quality of data analyses at CHI PLAY, the quality of reporting we found could not dispel claims of looming *"threats of a replication crisis"*[44, p.1]. The quality of reporting practices varied widely, with many recommendations made in behavioral sciences to address issues with NHST — such as consistently reporting effect sizes [112], power analyses [112], using non-dichotomous analyses [22, 51], or sharing data [134, 216] — not being widely used. To mitigate this, we introduced a checklist that authors could follow when writing up their study, and reviewers could use to quickly check if all required information was present. Anecdotal evidence in the form of conversations with peers

[1] https://chiplay.acm.org

and observations during peer review indicates that the checklist has found some use in teaching, for write-ups, and to communicate reporting recommendations to authors.

Moreover, aside from the mixed reporting quality, our study also indicated that many of the papers published at CHI PLAY may not benefit from using NHST in the first place. As I noted above, NHST is only suitable for very specific, *confirmatory* research questions [130, 186]. i.e., those research questions that seek to test a precise hypothesis with known error rates. It is, however, very common to use NHST to test hypotheses regardless of the research question or underlying theory supporting it [86, 187, 186]. Our analysis showed, that of the included papers, we could only identify clear confirmatory research goals in little more than a third (37%). This hints at *testing* theories potentially not being the primary goal of these works, which in turn indicates that this "ritualistic" use of NHST [86] might also be happening in empirical PX research.

The idea, that empirical PX research is often not aimed at testing theory is further supported by recent work on HCI games research, which highlights that little HCI games research meaningfully interacts with theory [208, 209]. In fact, PX research [62, 223], as well as the wider field of HCI [112], often does not seek to test theory, but instead evaluate design implementations, for which the use of NHST is questionable [112]

Moreover, the theories that are used in PX may not readily lend themselves to the type of precise hypothesis testing required for NHST to be used well. Recent work in psychology [e.g., 74, 224] has called out psychological theories to be unsuited for NHST, due to them being vulnerable to hidden assumptions [74], having only weak logical relation to the from theoretical tenets to derived hypotheses [145], and sometimes not being (scientific) theories at all, but rather *vague, imprecise and/or practically unfalsifiable"* [85, p.196] surrogates of theory. This critique of weak theories has since been repeated for psychological games research by Ballou [9], who argues that *"a purely confirmatory approach is ill-advised [as] games research simply lacks nuanced enough theory to make specific predictions"* [9, p.2:6]. Indeed, one of our most popular theories [163] — SDT [184] — has been criticized as difficult to derive precise claims from [59, 209].

Given that empirical PX research rarely formulates confirmatory research goals (*Pub. I*), PX rarely deeply engages with theory [209], and theories used in PX may not be ready to be thoroughly tested [59], are there better ways for PX researchers to conduct theory-informed research?

3.3 Exploratory Research

In HCI [45], as in other fields [e.g., 84, 220] exploratory research is a vital component, often overshadowed by confirmatory hypothesis testing. Unlike confirmatory research, exploratory

research does not seek to provide explanations for phenomena, but instead describe them [69]. It aims to obtain empirical regularities and make broad inquiries into complex systems, while usually not necessarily having access to any guiding "local" theory [202, 148].

This is not to say, that exploratory research does not utilize theory [220]. Exploratory research is guided by *background theories*[46], which are theories which are relevant to the phenomenon in question, but have not yet been shown to directly apply to it. While it may be unknown if a background theory applies to the phenomenon in question or not, they constrain the search space of effects and mechanisms researchers look for [145]. This has the benefit, that exploratory experiments can be used to investigate scarcely mapped phenomena, while raising the chances of finding robust effects by having a theoretical framework to scaffold results onto [73]. In this, exploratory studies are theory-informed, not theory-driven [220].

If exploratory research is done in this way — i.e., theory-informed — it can have a valuable impact for the theory it uses, and, in turn, could be highly useful in empirical PX research. As mentioned above, many theories in psychology are are too underdeveloped to derive the kind of hypotheses from which are meaningfully tested with NHST [187].

To develop theories to the point where they can be properly tested, Scheel et al. [187] recommend researchers to engage with the *Derivation Chain* [130] — i.e., focus their research not on testing theory, but on exploratory work forming concepts, developing measurements, defining relationships between concepts, and identifying boundary conditions and auxiliary assumptions [187].

This strengthening of the Derivation Chain can be achieved by several non-confirmatory, exploratory research activities. For example, researchers may engage in descriptive [84], discovery-oriented work [145], identifying robust phenomena to understand what a theory needs to explain. Another activity is a-priori evaluation of theory plausibility, for example by formal modeling [179], computational modeling [94] or theory mapping [91], which are ways to more easily check weak, verbal theories [74] for inconsistencies or uncertainties. Moreover, *exploratory experiments* allow for discerning plausible parameter ranges, identifying boundary conditions and auxiliary assumptions, as well as establishing empirical patterns [e.g., 73, 148, 202, 186]

Despite the importance of exploratory work as part of the derivation chain [e.g., 9], and HCI research practice [e.g., 45], there are little to no guides or accounts on how to perform exploratory empirical PX research. As discussed in the previous sections, guidance on how to use theory in HCI practice is primarily focused on design [e.g., 15, 40, 52, 203], while those works seeking to provide guidance on empirical work are mostly focused on improving statistical methods [e.g., 22, 37, 44, 112], or transparency [e.g., 104, 217], or suggest research activities without specific recommendations [e.g., 9, 209].

In the following chapters, I address the lack of accounts of theory-informed exploratory re-

search practice, by describing and reflecting on four research projects, representing different appraches to PX research. *Pub. II* (Chapter 4) and *Pub. III* (Chapter 5) are descriptive works which seek to improve our understanding of emotional moving, and empowering and disempowering gaming experiences respectively. *Proj. I* (Chapter 6) is an attempt at outlining how emotionally videogame experiences become enjoyed and appreciated, highlighting difficulties when attempting to test theories in empirical PX research. Lastly, *Proj. II* (Chapter 7) is an exploratory experiment, using CET [183] as a background theory to approach the experience of videogame rewards.

Chapter 4

Aesthetic Game Experiences

The project "'My Soul Got a Little Bit Cleaner': Art Experiences in Videogames" (*Pub. II*) employs an empirical aesthetic theory — the VIMAP [155] — to explore what kind of videogame experiences people consider to be art experiences, and what effects they had on them. *Pub. II* consists of an online survey in which participants were asked to recall and describe a recent instance of an art experience with games, explain, why they thought of it as an art experience, how the experience impacted them in the long term, and if they saw intention by the developer behind their experience. In addition, participants indicated the emotional spectrum the experience afforded using the Aesthetic Emotions Scale (AESTHEMOS) [191]. For the analysis, we chose to utilize *Qualitative Content Analysis* [128].

In this chapter, I will discuss this project by first, giving a general overview of emotional player experience and empirical aesthetic. I will then discuss why we chose to employ empirical aesthetics to study PX, highlight issues during the research project, and finally recap the results and reflect on the process.

4.1 Aim of Study

The videogame experience is rich with emotions [27]. While traditionally, PX research is primarily interested in positive experiences — i.e., fun [131] — focus has recently shifted to account more for experiences that aren't purely positive [27]. Past work on emotional videogame experience highlighted a wide range of both positive, and negative emotions [e.g., 29, 158], resulting in both positive [e.g., 28] and negative experiences [e.g., 89]. These findings indicate that the

experiences videogames are able to afford could closely mirror the experiences evoked by other forms of art [27]. Consequently, it might be be fruitful for the conceptualization of emotional videogame experience to take the perspective of *empirical aesthetics*.

4.1.1 Empirical Aesthetics

Empirical aesthetic [68] is a psychological subdiscipline primarily based in cognitive science [20] and neuroscience [42]. At its core, empirical aesthetics asks why and how we experience art. A notable questions as the human behavior of seeking out art is, on the face of it, paradoxical: We generally seek out situations that elicit positive emotions and avoid those that elicit negative ones; and while experiencing art is usually ultimately enjoyable, the art experience itself is fundamentally linked to negative emotions [133]. So while we dread the sight of spoiled food, we can appreciate it if the images are framed as art [218, 81]. While we try to avoid sadness in real life, we enjoy watching movies that make us sad [146]. And even though we experience fear in videogames, we experience pleasure from playing them [29].

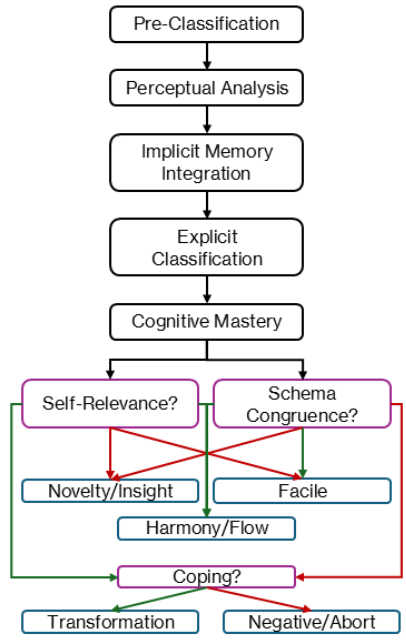

Fig. 4.1: The VIMAP (adapted from [155]). Once perceivers reach the Cognitive Mastery stage, they make two or three appraisal checks (purple), leading to different outcomes (blue), depending on the self-relevance, schema congruence and coping potential being high (green) or low (red). Note, the coping check is proposed to be only done when the artwork has high self-relevance but low schema congruence.

One recent model that seeks to explain how people react to art is the Vienna Integrated Model of top-down and bottom-up processes in Art Perception (VIMAP) [155]. The model is an amalgam of the authors previous models [117, 152], and rooted in the theory of sequential appraisal checks: A theory that broadly posits that living things, when faced with a situation, will go through a series of appraisal checks in order to come to the appropriate emotional and behavioral reaction [190]. For example, when encountering a new stimulus it is first appraised as novel. Based on this appraisal, the stimulus is then either appraised to be hard to comprehend, which leads to the negatively valenced confusion, or as easy to comprehend, which leads to the more positively valenced interest [198].

The VIMAP (Figure 4.1) proposes five distinct experiential outcomes to an art experiences based on a series of such appraisal checks: Facile, Novelty, Harmony, Abort, and Transformation [155]. The outcome the perceiver arrives at depends on three specific appraisal checks: an artwork being (1) congruent with the perceivers self-image (e.g., it fits with what they consider an artwork to be), on the artwork being (2) relevant (e.g., it is important to the perceiver to have an opinion of the artwork), and on the perceiver needing to (3) cope with it (e.g., the artwork challenges a perceivers beliefs in a form the perceiver cannot easily dismiss). These discrepancy checks and subsequent outcomes tend to be characterized by specific emotional outcomes, which are suggested to be stronger the more self-relevant the artwork is considered to be [153]: For example perceivers experience awe when experiencing a transformative event [155], or are being moved to tears when gaining self-relevant insights [156].

This central mechanism of art experiences depending on the experience and resolution of discrepancies, may be useful for studying PX, as it could apply to and explain the more emotionally complex and reflective responses to videogames we have increasingly become interested in [28, 89, 221].

4.2 Role of Theory

We chose to use the VIMAP to explore emotionally moving PX for two reasons: First, as outlined previously, discrepancies — and how perceivers deal with them — are a potential mechanism to account for particularly mixed emotional experiences, like those observed in prior empirical PX works. For example, the transformative experiences reported by Whitby et al. [221] are reminiscent of the transformative outcome of the VIMAP: Recountings by their participants often featured a form of reevaluation of the game as new information was presented, causing a discrepancy by going against their current interpretation of the artwork. Similarly the discomforting outcomes described by Gowler et al. [89], are similar to abort outcomes, as their participants reported discrepancies between their interpretation of the games, and their personal values and expectations.

Second, the VIMAP accounts for emotions that have not been covered much in prior PX work. Namely epistemic emotions such as interest and insight — which are particularly important for higher level art outcomes, such as harmony and transformation —- as well as negative emotions such as confusion and uneasiness [155, 191]. While some work on emotional gaming experiences [e.g., 29, 158] have covered a wider range of emotional responses, they have not incorporated theories that could explain how these emotions came to be or what effect they had on the player experience.

As such, the VIMAP served as the *theoretical background* [73, 46] of *Pub. II*. With it's foundation in emotional appraisal theory [190], it provided the study with a lens though which to conceptualize emotional game experience and consequently give them some theoretical grounding, by relating them to specific explanatory mechanisms — i.e., discrepancies. Consequently, *Pub. II* aids the concept formation of emotionally moving gaming experiences [187].

4.3 Study

The VIMAP feeds into our study design by guiding our research questions, giving implications for survey questions and measurements, aiding our analysis, and relating our results to prior research.

For our study design the VIMAP serves as a lens, focusing our survey on emotional reactions and post experience impact [95]. The VIMAP posits different art outcomes depending, accompanied by emotional reactions and dependent on three discrepancy checks [155]. As most of the process of the VIMAP is subconscious [155], we decided not to directly ask about the experience of discrepancies. Instead we opted to just ask participants to recount an art experiences with videogames, elaborate why they thought it was an art experience, and to describe the impact these experiences had on them. We expected these questions to give us insights into what aspects of games cause experiences to be considered art experiences, and provide accounts of when these experiences are particularly impactful. In addition, we asked for the aesthetic emotions [27, 133] felt by the participants during the experience. We used the AESTHEMOS [191], a questionnaire designed specifically to cover a wide range of aesthetic emotions (e.g., Prototypical aesthetic emotions, nostalgia/relaxation, sadness, and amusement) in regards to a wide range of aesthetic artifacts (e.g., paintings, music, and architecture). Notably, the AESTHEMOS, like the VIMAP, was not designed to cover games.

Based on the VIMAP, we expected our results to resemble the five distinct outcomes of art experiences. Consequently, we planned on conducting a *Latent Class Analysis* [144] to explore the emotional profiles of each outcome. The clustering of the Latent Class analysis was in turn to be guided by Thematic Analysis (TA) [32]. However, while expected the clusters to be similar to the distinct outcomes proposed by the VIMAP, this did not turn out to be the case, with clusters being too manifold and non-exclusive to be meaningfully analyzed with our sample size [see 26, for the analysis].

Consequently we adjusted the approach by opting for Qualitative Content Analysis [128], a codebook based approach which allowed for the explicit incorporation of prior knowledge and expectations — i.e., the VIMAP — into the analysis. We derived our codebook from two prior

works in empirical aesthetics [77, 154] as well prior work on emotionally moving videogame experiences [29].

4.4 Results and Discussion

In recent years, PX research has become increasingly interested in emotionally complex videogame experiences [27] and scholars have called for an *"Empirical Game Aesthetics"* [14]. Yet, no empirical work has attempted to ground the phenomenon of emotionally moving game experiences in theory. By using the VIMAP as a background theory [46] throughout the study process, we could constrain our search area on aesthetic emotions and the discrepancies causing them.

In particular, we were able to highlight emotions that are rarely acknowledged in PX research — such as epistemic (knowledge) emotions, and prototypical aesthetic emotions — as well as illustrated the long term impact these emotional experiences could have. Moreover, some of our participants accounts closely matched both the transformative outcomes of the VIMAP, and the endo- and exo-transformative gaming experiences observed by Whitby et al. [221]. While this in no way says that these experiences have the same origin (or that the VIMAP constitutes the causal explanation for these experiences), it highlights that this theory-informed descriptive research can help synthesizing the existing corpus on the phenomenon of emotionally moving player experiences.

This being said, utilizing the VIMAP poised several challenges throughout the research process. In this study, participants reliably reported (mostly) positive experiences that impacted their lives in some way. Given that the discussions about videogames being art has been long going in the public [see 27], it was not surprising that people who may want to see the medium worthy of the label "art" could readily recount such an experience. In this sense, theory did not only focus our own analysis onto certain aspects of the game experience, but it may have both pre-selected participants who are interested in games as an art form, as well as focused the recountings of our participants, who primarily focused on videogames' aesthetic qualities (e.g., visuals, sounds, and narratives) and not on gameplay.

At the same time, the experiences in our dataset did not directly map onto the art experience proposed by VIMAP. As the VIMAP is a cognitive theory and therefore primarily focused on (mostly subconscious) cognitive processes [155], its core tenets do not necessarily show in the written accounts remembered experiences. Consequently, our qualitative analysis may relate player accounts to the theory, helping to shape the concept of an art experience in videogames, but we cannot make statements about the explanatory ability of the VIMAP.

Lastly, it would have been beneficial to use some form of measure to capture the aesthetic

experience. However, no quantitative, behavioral measures of the VIMAP outcomes were available. Instead, we used the AESTHEMOS which covers a wide range of emotions— including enjoyment, a common indicator of good PX [1, 131] — but was not validated for games and had worse than expected internal consistency (see the supplementary material).

In conclusion, using the VIMAP as a background theory did shift our attention to hitherto sidelined emotions and allowed us to synthesize past results, indicating some descriptive and rhetorical power [95] of the VIMAP and offering a new analytical lens [15] for emotionally moving game experiences. However, issues arose during analysis and interpretation of the data as the VIMAP's proposed cognitive mechanisms are hard to measure quantitatively, and hard to map to qualitative responses. Moreover, art experience being a "loaded" term may have heavily influenced both the kind of experiences reported and the types of people who answered the questionnaire.

Chapter 5

Empowering and Disempowering PX

Pub. III is an exploratory survey of what kind of experiences players consider to be empowering and disempowering. While the study approach is similar to *Pub. II* — particular in terms of study design — this paper engaged with theory quite differently: While emotionally moving videogame experiences are a rather well described phenomenon in lack of a theoretical foundation, empowerment in videogames is rarely defined itself, but may have conceptual overlap to several other PX theories, such as such as SDT [184], and Attribution Theory [58]. Therefore, in *Pub. III*, we set out to explore and conceptualize the phenomenon of empowering and disempowering videogame experiences by examining the type of experiences players themselves find empowering or disempowering, and how these experiences relate to other PX concepts.

In this chapter, I briefly recapitulate conceptualizations and uses of empowerment in other fields and discuss how it is used in HCI and PX research. I then discuss the study process, how we decided to approach the phenomenon and the process of relating it to existing PX concepts.

5.1 Aim of Study

Empowerment as a scientific term has a long history. It was first used in a sociological context [e.g., 107, 111, 180, 201] where it was positioned as a bottom-up process to alleviate systemic pressures on marginalized communities so that they would have the rights and means to live a good life [167].

Later, Zimmerman [237, 235, 236] shifted the level of analysis from the sociological focus on communities to the individual. He termed the concept of *Psychological Empowerment*, which is an individual, context-dependent phenomenon, grounded in intrapersonal, interactional and behavioral components: I.e., a person is empowered if they can (and believe they can) change, understand, and control their current situation [236].

In HCI, empowerment is a widely, but inconsistently used term [193]. Schneider [193] identified eight different lines of empowerment research in HCI, differing on their (1) concept of power (power-to, or power-over), (2) psychological component (feeling, knowing, or doing), (3) persistence of empowerment (persistent or transient), and (4) design mindset (participatory, or expert). In PX research alone, empowerment has been used to describe giving differently abled players the ability to play and compete with one-another [82, 90], to exaggerate players' movements in a mixed reality game giving the in-game appearance of "super-powers" [119], to offering coding-free development tools [12]. Another different use focuses on AI and player experience, where empowerment is considered a *"context-free"* [p.1 114] universal function; a concept closer to what Zimmerman coined individual empowerment [236]. Here empowerment is the quantification of *"an agent's potential perceivable influence"*[93, p.3].

These different understandings of empowerment lead to what is termed "conceptual stretching" [83, 187], where the concept of empowerment is stretched over a range of phenomena that are not necessarily aligned. While it is generally understood that empowerment is good [193], there is no agreement on what it means for players to be empowered. Moreover, empowerment is primarily seen as an *outcome*. How empowerment is *experienced* by players — what causes it and how it makes them feel — has seen little empirical attention. This project aimed at providing descriptions of when they felt empowered and what these experiences meant to them in order to aid the conceptualization [207] of empowerment as PX.

5.2 Role of Theory

Empowerment has a long history in the behavioral sciences [167, 235, 193]. However, it is unclear what it means for videogame players to be empowered, and if these experiences are always desirable. Moreover, while disempowerment in real life is generally undesirable [167], in games where negative emotions can be highly appreciated (see Chapter 4), there could be value in disempowering players. In order to understand what it means for players to be empowered and disempowered and how they experience either we use several established concepts of what makes good PX and relate them to the concept of Psychological Empowerment [236].

The goal of this theoretical triangulation was to interlink the commonly used term empow-

erment with existing PX constructs in order to map what it might mean for a player to feel either empowered or disempowered, and therefore how the term can be used in future research and game design.

5.2.1 Empowerment

In order to further our understanding of empowerment and disempowerment as PX we started with the concept of Psychological Empowerment [236], which we then related to established PX constructs. To do this, we took the descriptions of Psychological Empowerment and compared them to the descriptions of common PX concepts of SDT [184], attribution theory [58], and the Player Experience Inventory (PXI) framework [1]. This revealed several points of overlap.

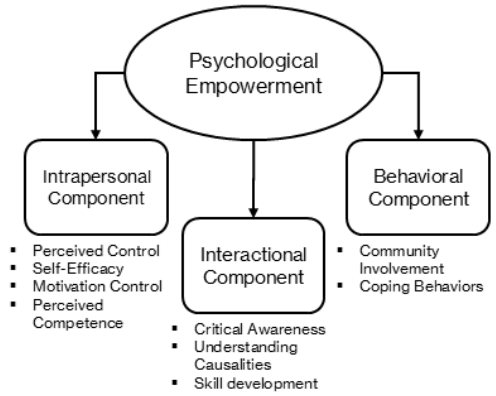

Fig. 5.1: The nomological network of Psychological Empowerment according to Zimmerman [236]

Zimmerman [236] proposes Psychological Empowerment as an open-ended construct with three components: Intrapersonal, Interactional, and Behavioral (Figure 5.1). Intrapersonal components of Psychological Empowerment are primarily related to a sense of self-efficacy, perceived control, and perceived competence. The Interactional component relates to people being aware of their given context, and understanding their options and chances for impact. The Behavioral component relates to the actual behaviors people may take.

These three components of Psychological Empowerment provide some overlap with established PX concepts. First, the theory of Basic Psychological Needs — a subtheory of SDT [184] — highlights the importance of the satisfaction of three needs for an activity to support intrinsic motivation: (1) autonomy — feeling ownership and self-endorsement of ones actions — (2) competence — feeling effective when interaction with the environment — and (3) relatedness — feeling connected with others. These three needs are reminiscent of particularly the intrapersonal and interactional components of Psychological Empowerment: Experiencing domain-specific control, self-efficacy, and perceived competence can mean satisfaction of competence and autonomy, while successfully navigating ones community may satisfy relatedness.

Second, Attribution Theory in PX research aims to explain to which factors players attribute success and failure [58]. Success and failure can be (1) internally or externally attributed —

i.e., this outcome was because of me or because of others —- can be (2) stable or transient — i.e., future outcomes are determined by the same or other factors — can be (3) controllable — i.e., the outcome was dependent on something the player controlled like effort, or something they cannot change like talent — and (4) can be global or specific — i.e., success or failure translate to other tasks or are specific to this singular type of event. Players attribution of success and failure as either their fault or out of their hands directly corresponds to Zimmerman's [236] account of people feeling being in control of their actions.

Lastly, *Pub. III* employs the PXI. While the PXI has overlap [1] with SDT [184] (and particularly the PENS [185]), and therefore the same overlap with Psychological Empowerment, it further highlights players understanding of their environment: The "Goals and Rules" and "Progress Feedback" dimensions cover the players understanding of how their environment works and what they have to do to successfully act in it. This is reminiscent again of the Interactional component of Psychological Empowerment. Moreover, the PXI is positioned to measure positive PX based on low level aspects of the game and higher level psychosocial evaluations by they player [1]

5.2.2 Disempowerment

In addition to empowerment, *Pub. III* further investigated the concept of disempowerment. While disempowerment is usually not a concept in itself but just the absence of empowerment [e.g., 168, 193, 236], in videogames, disempowerment could potentially be an interesting concept in and of itself: As discussed in Chapter 4, videogames (like art) often feature negative experiences to a positive effect. Likewise, videogames often benefit from [75, 160], or even require [109] the at least credible threat of failure. Moreover, disempowerment, in form of surrendering control, has been shown to be a powerful subversive act [18]. Consequently, in addition to focusing on the conceptualization of empowerment as PX, *Pub. III* looked at the experience of disempowerment.

5.3 Study

There is considerable overlap between the verbal accounts of Psychological Empowerment [236] and SDT [184], Attribution Theory [58], and the PXI [1]. Yet, there are no accounts on how either empowerment or disempowerment are experienced by players, and when or whether these experiences make for good or bad PX. Therefore, we designed an exploratory mixed-methods approach to help understand what kind of videogame experiences players find empowering and disempowering, and how they come to enjoy or dislike them.

One issue when trying to explore the phenomenon of Psychological Empowerment is its individual nature: Psychological Empowerment takes different forms for different people, in different contexts, and fluctuates over time [236]. In turn, there is no questionnaire to measure Psychological Empowerment. Instead, as empowerment may not be meaningfully capturable by a standardised questionnaire at all [236], we chose to instead collect qualitative reports of players experiencing either empowering or disempowering experiences. We then used the PXI [1], an attribution questionnaire [58], and a compound questionnaire for need satisfaction and frustration [1, 7, 43, 197] to explore the relation between empowerment and these concepts, as well as the valence of the PX.

To analyze the qualitative answers, we employed Reflexive Thematic Analysis (TA) [32]. TA is an approach to qualitative data analysis and interpretation which implicitly entangles the researchers subjectivity in the analysis process. In context of the given study, this approach allowed us to incorporate prior theoretical understandings and expectation of both empowerment and disempowerment, as well as the related concepts to aid in the creation of the seven themes.

5.4 Results & Discussion

Our approach of identifying themes of empowerment and disempowerment and subsequently investigating their relation to other PX concepts provided us with a detailed view of empowerment and disempowerment as PX. Most notably, while it did support the general understanding of empowerment being good, it further highlighted that in videogames, pure empowerment is less appreciated than the empowerment that comes after considerable struggle. Likewise, while in game disempowerment can be bad — particularly if it relates to real-world discrimination vectors such as gender — it can likewise pose an opportunity for highly appreciated, deeply moving experiences.

Yet, exploring the player experience of empowerment was not without challenges. First, both Psychological Empowerment [236] and the theories and concepts we connected it to — Basic Psychological Needs [184], Attribution Theory [58], and the PXI [1] — are primarily verbal [74], which makes relating them to each other not straight forward. For example, while both Psychological Empowerment and SDT highlight a feeling of competence, it is not necessarily clear if these concepts are the same in both theories. Similarly, the concept of attributing event outcomes to oneself in Attribution Theory may or may not be the same as experiencing self-efficacy and understanding the causal connections of our surrounding in Psychological Empowerment.

Notably, particularly the attribution questionnaire [58] showed considerable overlap between empowering and disempowering experiences. While there was often some form of at-

tribution in the qualitative accounts of players — i.e., the themes "Out of Their Hands" and "Being the Worst" primarily differed on failure being described as the fault of others or the player themselves — this seemingly did not directly translate into the quantitative scales. It should be noted, however, that our approach of separating participants via TA, before comparing at their emotional experiences has very little validity outside of the study: As the act of separation entails an active interpretive effort by the researchers, these groups should not be treated as distinct classes that could represent different populations.

Moreover, our focus on player accounts did not result in a clearer conceptualization of empowerment and disempowerment, but instead highlighted the range of experiences players consider to be either. While we did find interesting overlaps to the other concepts, our results indicated both empowerment and disempowerment meaning a lot of different things to different people, in different situations and even in different games [236]. For example, being disempowered in a multi-player game was generally bad, but being disempowered in a horror game was the intended, enjoyed experience.

This being said, *Pub. III* highlighted elements important to players during empowering and disempowering experiences, and what made these experiences positive and negative. Even with the limited generalizability of the themes and associated differences in quantative scales, this unearthed several interesting research directions — such as the potential importance of first disempowering players to ultimately create more empowering experiences. In this, while this study complicated what empowerment as a player experience may mean, it offered a starting point to further develop empowerment and disempowerment as valuable player experiences

Chapter 6

Explaining Positive Discomfort

In Chapter 4, I discussed previous work exploring emotional videogame experiences from an empirical aesthetic perspective. One take away from *Pub. II* was that empirical aesthetics may be utilized to study emotional videogame experiences. In particular, it highlighted the role of epistemic emotions — e.g., insight and understanding — which could position appraisal theories as a tool to understand how some emotional videogame experiences come to be experienced as negative [e.g., 89], while others come to be experienced as positive [e.g., 28]. However, the descriptive design of *Pub. II* did not allow to ascribe any explanatory power to the VIMAP when it comes to emotional videogame experiences.

Consequently, while planning a follow-up on *Pub. II*, we aimed to test the core mechanics of the VIMAP, resulting in *Proj. I*. In the following, I will reflect on *Proj. I*by first briefly discussing emotional challenge. I will then recaputlate how theory fed into the study design — focusing on the hypothesis derivation and theory-driven game design — the analyses and results, and finally reflect on the outcome of this study. The data, analyses and all research artifacts relating to *Proj. I* are available at https://osf.io/rnbse/?view_only=d59e36415fe3418c8d6ebaacd264de6f.

6.1 Aim of Study

In games research, a growing body of works has emerged, examining the appeal of games that are uncomfortable [89, 110] or frustrating to play [109, 226], that afford strong negative player responses [25, 28, 126, 135], or feature emotional challenge [29, 47, 158]. However, these works remained primarily descriptive and non-generalisable. For example the work by Bopp et al. [29]

on emotional challenge notes that *"[...] no causal inferences may be drawn due to the exploratory nature of [their] study"*[29, p.41]. Similarly, Jørgensen [110] discusses the positive discomfort of *Spec Ops: The Line*'s[232], but explicitly notes the findings may not apply to even other military shooters.

Overall, while regularly described, no theory has been brought forward to explain emotionally moving videogame experiences. This is particularity pertinent as these experiences can be highly appreciated [e.g., 29, 221], but also be source of considerable discomfort [89]. Consequently, being able to predict when (or for whom) a certain experience will be positive or negative, would be highly useful for game designers seeking to evoke these experiences. At the moment, while there are game design textbooks which provide techniques for eliciting flow [76] or intrinsic motivation [189], there are few pointers for how to create games that perturb without repelling players [36, 56]. Likewise, it would be beneficial for games researchers who try to understand how different player experience (PX) concepts (e.g., discomfort, psychological needs, enjoyment) are related, as it could subsequently improve our field's capacity to study games phenomena beyond enjoyment.

Proj. I aimed to understand the explanatory mechanisms behind emotional challenge — that is, mechanisms that allows us to predict when and how an emotionally challenging experience will become appreciated — by approaching the phenomenon with two theories: SDT [184], and the VIMAP [155].

6.2 Role of Theory

Proj. I aims to compare two different perspectives on emotional challenge: SDT [184] — a common PX framework [163, 223] — and the previously mentioned Empirical Aesthetic perspective (Chapter 4). I will discuss them here in turn.

6.2.1 The Motivational Perspective

A common way to frame PX is by way of motivation: SDT [184] is widely used in PX research [163, 223] and centers good PX around the satisfaction of psychological needs: Autonomy, Competence, and Relatedness. Should these needs be satisfied, players will enjoy the experience. This motivational approach to PX, however, does not really account for mixed affect experiences that are not need fulfilling experiences, yet, still highly valued [e.g., 29, 110]. While enjoyment is usually somewhat lower in emotionally challenging experiences compared to conventionally challenging experiences [29, 158], work on emotionally challenging [e.g., 28, 29] and need frustrating

experiences [10] found players to tolerate and even expect some need frustration in games. Likewise, both game design practice [e.g., 36] and the games research literature [126, 102] has argued "breaking" needs to be a good way to create memorable, impactful experiences. The theory itself, however, does not provide an explanation how this may be.

6.2.2 The Empirical Aesthetic Perspective

At the same time, the VIMAP [155] (as discussed in Chapter 4) may provide some explanatory power in regards to negative affect experiences. According to the VIMAP, during the art experience perceivers will, among other factors, incorporate context information [153] to reach an opinion about an artwork. In particular when engaging with challenging and discomforting art, successful resolution of discrepancies during the *cognitive mastery* stage (see Figure 4.1) can result in highly positive experiences, while failure to resolve discrepancies — i.e., not "getting" the art — may in turn results in feelings of confusion [199] and subsequently distaste for the artwork [155].

Several studies have shown that the experience of artworks can substantially change when the perceiver view art — for example — in different contexts [33, 34, 35] or with extra contextual information such as the title [118] or creator [127] provided. Moreover, it has been shown that discomforting images (i.e., depictions of bodily fluids, rotten food, etc.) are being experienced more positively [81], or considered as more beautiful [218] when being framed as art as opposed to documentary photographs.

Moreover, in *Pub. II*, player accounts evoked experiences, hinting at the value of resolving discrepancies: For example, two participants in *Pub. II* talked about *"NieR: Automata"* [162]: While one participant understood the games artistic ambitions, they found the game "pretentious and difficult to understand" [30, 237:12]. Yet the other participant experienced the same game as a great source of comfort during a difficult life experience, highly lauding the game and their time with it. These accounts are in line with the VIMAP: Two players are confronted with the same stimuli, but came to a different conclusion, potentially based on different cognitive mastery.

6.2.3 Bringing the Perspectives together

In regards to emotional challenge, these two theories provide different perspectives. The SDT perspective would expect players experiencing need frustration to enjoy a game less, while the VIMAP perspective would instead expect a player to positively react to a frustrating game if they can solve the discrepancy that caused the frustration.

Notably, however, these two theories do not *directly* contradict each other — primarily because the two theories built on different starting points of human behavior. At its core, the VIMAP does not acknowledge "motivation" — aside from a nod towards motivation being a positive aspect of early art experience process that fosters continued engagement with the artwork [155]. Likewise, SDT [185] does not acknowledge emotions. More broken down and directly relevant to the explanation of phenomena, the VIMAP does not include the concept of basic psychological needs, while SDT does not include cognitive appraisal checks that incorporate contextual information. Moreover, SDT as PX uses enjoyment as the primary outcome dimension, while the VIMAP has many different outcome dimensions (e.g., emotions, arousal, valence, liking, appreciation).

Still, both theories allow us to make predictions about the experience of emotionally challenging videogame experiences. In the next section, I will discuss how these theories fed into the study design and hypothesis test of *Proj. I*.

6.3 Study

6.3.1 Deliminating the Phenomenon

To study emotional challenge, we first further broke down the phenomenon of interest. Given that the phenomenon of emotional challenge is still quite broad, we further focus our inquiry on *discomfort*.

Discomfort, like the wider concept of emotional challenge, is marked by negative emotional reactions [18]. Discomforting experiences can be fostered by subverting established genre conventions, and can create unexpected or shocking situations, elicit extreme affective changes, and prompt further consideration of a game's subject matter [110]. Moreover, Burnell [36] claimed that discomfort can be elicited by "breaking" psychological needs. Furthermore, discomfort may form a dialogue between player and designer, and in doing so *"give players something idiosyncratic, weird, and confrontational"* [226, p. 7]. It has been argued, that at least one benefit of uncomfortable interactions is their potential to *"stimulate powerful emotions"*, and heighten *"the subjective intensity and memorability of the experience"* [18, p. 2006], which can consequently foster striking emotional experiences [e.g., 29].

However, evoking these experiences is not without risk: Gowler et al. [89] examined uncomfortable game experiences, observing that discomfort can arise when players are presented with disturbing themes, or when control over outcomes are limited. They observed, that players who were unexpectedly exposed to discomforting subject matter, or found it overwhelming,

regarded their experience as more negative, while players' experience of loss could lead to them questioning their competence and confidence. Hence while negative emotions can be a valuable part of the gaming experience, it is not a foregone conclusion.

This concept of discomfort, which can be caused by frustrating psychological needs [36, 89], may cause averse outcomes [89], but may be elated by meaning-making [18, 110] lends itself to be adressed with both the SDT and the VIMAP.

6.3.2 Hypothesis Generation

From our literature review, we had selected two potential core mechanisms: (1) Discomfort in games may be caused by the frustration of needs [36, 89]; and (2) Discomfort may become experienced positively if the reason for the discomfort is understood [110, 18]. This links the experience of discomfort to the two theories we employed in this study.

First, according to SDT [184] need frustration thwarts intrinsic motivation and should thereby decrease the enjoyment of an activity. In fact, SDT does not offer an explanation why need frustration could or should be desirable. Consequently, we posited that frustrating the needs of players decreases the enjoyment of the game. Second, the VIMAP [155] posits, that discrepancies in art experiences may be resolved by providing contextual information to the perceiver [e.g., 35, 81, 218]. Consequently, we posited that providing a thematic context for the game will increase appreciation and overall liking of the game.

Based on these mechanisms, we derived the following, statistical hypotheses for a 2x2 design (high/low need frustration & thematic context/no context):

H1: Participants in the higher frustration conditions will enjoy the game less than players playing the low frustration condition, regardless of context.

H2: Participants will appreciate and enjoy the game more in conditions where a thematic context is provided.

H3: Participants playing the frustrating and thematically contextualized condition will enjoy and appreciate the game most.

We operationalized these variables using several validated scales. The main outcome variables were appreciation and enjoyment, as they capture distinct facets of players' liking of a game [28, 147]. Enjoyment pertains to the experience of fun, while appreciation is about experiencing media as meaningful [146]. Both scales have been previously used in studies exploring discomforting game experiences [e.g., 28, 29, 158]. In addition, to measure need satisfaction and frustration we used a compound questionnaire for need satisfaction and frustration [1, 7, 43, 197]. Lastly, to cover the emotional experiences we selected the Insight and Confusion subscales from

Fig. 6.1: Screenshot from the final version of the game. The screenshot on the left depicts the low need frustration condition with the warning symbols and checkpoints, the screenshot on the right depicts the high need frustration condition without warning symbols and checkpoints. Not visible are the difference in controls between the conditions. The game can be played at https://experimentgames.itch.io/

the AESTHEMOS [191], as prior empirical aesthetic works indicate their importance for liking artworks [155, 199]. Furthermore, to cover negative emotions deemed important for discomfort [e.g., 29, 89], we included Sadness, Anger, Boredom and Uneasiness in the questionnaire as well.

6.3.3 Stimulus Design

Lastly, before running the study, we required a game to test our hypotheses with. We briefly scoped commercial videogames [210], however, after considering the specific manipulation we required, as well as potential licensing issues when running the study online, we chose to create a bespoke experimental game.

Our experimental game was inspired by *Getting Over It with Bennett Foddy* [140] — a climbing game specifically designed to frustrate players by featuring unintuitive controls and punishing game design, while, at the same time featuring a voice over contextualizing (and reacting to) the players' failure[1].

Our own resulting vertical platformer (Figure 6.1) featured similar gameplay in that the player tries to reach the top of the play field. We emphasized verticality in the level design, producing an environment where a single error could cause the player to lose all their progress — while having to watch their character fall down. Moreover, we adopted an unpredictable movement system [see 165], which afforded players few opportunities for skill development. In the high frustration condition, control parameters were randomized, (e.g., making the controls more or less "floaty") at each button input. This impeded players' ability to adjust to the game controls, and hence worked to actively thwart players' competence needs. Lastly, while players

[1] see https://store.steampowered.com/app/240720/Getting_Over_It_with_Bennett_Foddy/

in the low frustration condition were made aware of upcoming danger via warning symbols (Figure 6.1, left), the high frustration condition could only be overcome by trial and error, as spikes appeared without telegraphing and platforms shifted mid-jump. The latter is reminiscent of the unfair and often unexpected design of "abusive" games [226], and was expected to further frustrate players' needs for autonomy and competence.

To aid and hinder cognitive mastery we further manipulated the games thematic context. Analogous to studies in empirical aesthetics [e.g., 81, 118, 127] we provided players with either no meaningful context, in which the game is simply labeled as a prototype, or we gave them context from which they could derive reason for the bad controls: We labeled the game as an art project that seeks to make a statement about the unpredictability of life. In line with prior work on experience of discomforting art [81, 218], where the art context improved peoples appreciation, we expected that the art framing similarly reduced the negative impact of the abrasive game design.

6.4 Results & Discussion

After running the study, our pre-planned analysis showed no significant main effect for neither enjoyment nor appreciation. The only significant effect was the main effect of context on appreciation ($F(1, 370) = 5.7, p = .02, f = 0.12$). As discussed in Chapter 3, null results posit a problem as they do not allow to make a decision between the absence of an effect or the inability of the study to detect it. An additional exploratory Bayesian analysis [61] did provide some further insight — for example again showing an effect of context on appreciation — but the models did not perform well enough for us to deem them reliable[2]. This left the study at an impasse, as it was not possible to make a decision between the study not performing well enough, or either theory not being supported.

The central issue during analysis was uncertainty whether the experimental manipulation properly elicited need frustration and the context effect. While we based our design on both theoretical tenets [e.g., 165] and successful commercial games [140], there is no way to make sure this translation into design worked. Consequently, should we reject our alternative hypotheses and therefore theoretical mechanisms, we might commit a *Type III* error — *"concluding a finding does not exist when [...] the study was not designed properly"*[96, p.3314].

Moreover, if there was an effect, it could have been to minuscule to register on the used measures. As discussed in Chapter 4, the AESTHEMOS [191] is not validated for games, nor did

[2]Please see the repository for further details:
https://osf.io/rnbse/?view_only=d59e36415fe3418c8d6ebaacd264de6f

not show satisfactory scale validity, potentially obfuscating any effects. Moreover, in Empirical Aesthetic research, studies that manipulate framing of artworks [e.g., 81, 169] usually invoke effects over several artworks. While this is feasible with artworks which can be fully perceived in seconds [155], this is not feasible for a game that participants have to spend some time playing. Consequently, while the context manipulation may work in empirical aesthetics, it may not be translatable into PX studies.

Another important aspect of the VIMAP [155] is that especially higher order art experiences are dependent on the perceiver being sufficiently interested in the artwork to begin with. If the art does not strike them in any way, they continue on their way without any particular reaction [155]. A potential alternative explanation for our results may therefore be that the participants hired through *Prolific* were not engage enough with the game. Consequently, they may have not been invested enough to feel particularly frustrated by lackluster gameplay or influenced by the context, or may even saw the game as a welcome change of pace from the other tasks offered on the platform.

In conclusion, given that we did not fully trust our manipulation to have worked, we decided to no longer seek publication. It was not possible for us, to make a call whether our study failed in way that we could generate knowledge from — in regards to theory or methodology — which would warrant publication, or if the null results were simple researcher error from which little knowledge could be gained.

Chapter 7

Rewards in Videogames

Proj. II is an on-going study exploring the effects of rewards on the player experience from the perspective of Cognitive Evaluation Theory (CET) [183]. Following recent calls for deeper engagement with SDT [59, 209], this study aims to understand, if and how the predictions CET makes about the experience of rewards applies to videogames as proposed by some SDT scholars [170]. To this end, this study attempts to derive specific claims from CET [183] and CET-based game design recommendations [e.g., 170], implement them in an experimental game, and relate findings back to how CET may predict player experiences.

In this chapter, I will first discuss how both CET and videogame rewards are discussed among game scholars and practitioners. Then I will outline the challenges of deriving clear CET-based experimental manipulations, of turning them into game design elements, and of talking back to theory in an exploratory experiment. All relevant artifacts in regards to *Proj. II* can be found at https://osf.io/zgb9d/?view_only=8e8099e913174e699dce9c827172d8fc

7.1 Aim of Study

Rewards are considered a central element of game design [e.g., 4]. Game industry professionals promote rewards as a way to improve player experience and retention [e.g., 171, 103, 54, 122] and as tools to shape player behaviors [e.g., 66, 87]. Games researchers have studied both form [e.g., 219, 125, 166, 161] and effect of rewards [e.g., 57, 64]. Moreover, rewards have been utilized to increase the efficacy of gamified systems [e.g., 120]. In games, rewards can take many forms: They can be new abilities, in-game valuables, praise, and bragging rights [125, 106, 170, 219,

166]. They can be collectibles and social markers, proof of skill and time invested, can motivate behavior and ease disappointment [219]. While there is substantial prior work on reward types and their *impact* on the player experience, there is little research about why rewards may shape players' experience and behavior. In HCI games research, rewards are primarily studied in terms of the form they take in games [e.g., 125, 166, 161]. These studies, however, rarely engage with why rewards of certain types may impact the player experience differently.

One theory that may explain difference in the impact of rewards is Cognitive Evaluation Theory (CET) [183]. CET has been placed as an explanans for the experience of rewards in recreational play [e.g., 170], however, these claims have so far seen little empirical attention, with the only works looking at CET in the context of games focusing on gamification [e.g., 132, 151, 213]. Addressing this lack of empirical research on the experience of rewards, *Proj. II* seeks to explore how CET may apply to videogame rewards.

7.2 Role of Theory

Proj. II focuses on CET [183], a sub theory of SDT [184] explaining the effect of *"externally administered rewards"*[184, p.125] on intrinsic motivation. It particularly emphasizes that rewarding intrinsically motivated behavior can harm the intrinsic motivation of the receiver, causing them to cease the behavior. The theory was primarily developed in work and school contexts [55] where it was found that *tangible* rewards — e.g., money — tend to thwart intrinsic motivation, while *verbal* rewards — e.g., encouragement — do not.

According to CET, the experience of a specific reward is based on its expectancy and contingency. According to a meta-analysis by Deci et al. [55], tangible rewards thwart free choice behavior (i.e., the continuation of a behavior once rewards have ceased) if they are expected, and if the reward is contingent on the receiver engaging, completing or showing a certain performance in the task. Notably, interest — an indicator of intrinsic motivation [184] — was similarly thwarted by expected rewards, but was not thwarted by performance contingent rewards. Potentially because performance contingency may support perceived competence [183].

Depending on its expectancy and contingency rewards will have a different *functional significance* [183]: The perception of a reward as either informational or controlling (or amotivating). If a reward is primarily experienced as informational — i.e., the receiver perceives the reward as feedback about their actions, creating an *internal perceived locus of causality* — the reward will increase the perceived competence and foster *intrinsic motivation*. Intrinsically motivated behavior is experienced as fun and interesting and will likely be engaged with more. On the flip side, if a reward's controlling aspect is most salient — i.e., the receiver understands the reward to be

an attempt by others to control their behavior, creating an *external perceived locus of causality* — the reward will decrease the receivers sense of autonomy and consequently thwart intrinsic motivation. For intrinsic motivation to be fostered, both autonomy and competence needs must be satisfied [183]. This process is mapped in Figure 7.1

7.2.1 CET and Videogame Rewards

SDT scholars have used CET to give specific recommendations targeted at game designers, in the form of Game Developers Conference (GDC) talks [e.g., 170, 171] and popular science books [e.g., 172]. However, there is scant evidence that CET tenets apply to videogames. In particular, the nature of some videogame rewards being verbal and some tangible, as well as videogame rewards differing in regards to their functional significance has so far not been investigated. *Proj. II* attempts to mitigate this, by investigating if and how CET tenets correspond to player experiences. Particularly if common videogame reward types behave as verbal or tangible rewards according to CET, if they have differing functional significances, and how they impact the player experience.

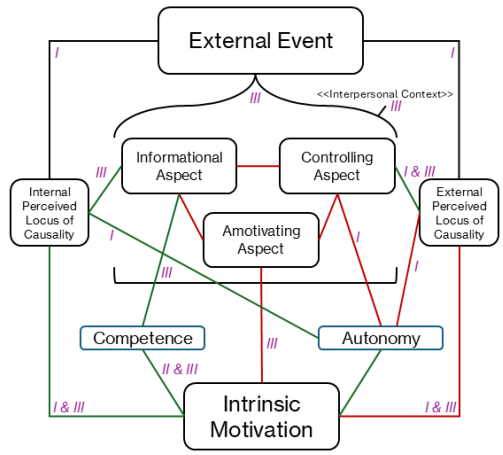

Fig. 7.1: A map of CET [183] based on the three core tenets, using the theory mapping approach by Gray [91]. Green and red lines indicate positive and negative associations respectively. Black lines indicate an undefined association. Elements in brackets constitute fundamental elements. Text in arrows indicate a modulator. Purple numerals refer to the associated CET tenet. Note, that Autonomy is not mentioned to affect intrinsic motivation in the core tenets, however, other places within the chapter [183] argue for both competence and autonomy needing to be satisfied for intrinsic motivation to not be thwarted.

To this end, *Proj. II* uses CET as a background theory [46]. i.e., CET provides a theoretical scaffold to constrain our study design, form our research question, and guide our interpretation. It constrains our search space, in that it proposes a specific mechanism — functional significance — to explain differing experiences of rewards, as well as provides methods to measure it in the form of a questionnaires — the Intrinsic Motivation Invetory (IMI) [41] — and the free-choice paradigm [183].

Table 7.1: Table depicting the reward types used in *Proj. II* (rows) and our expectations regarding the experience of these rewards on several factors (columns). The cells contain the expected effect of the given reward type based on claims made in regards to CET and videogames.

	Reward Contingency	Verbal or Tangible	Functional Significance	Need Satisfaction	Intrinsic Motivation
Praise	Completion	Verbal	Informational	Competence ↑	↑
Facility	Completion	Tangible	Controlling	Autonomy ↓	↓
Glory	Performance	?	?	?	?
All	Both	Both	?	?	?
No Reward	N/A	N/A	N/A	—	—

7.3 Study

To explore the phenomenon of videogame rewards using CET, we first derived specific expectations in regards to how reward types will impact the player experiences, before designing conditions that should be able to either support or subvert our expectations.

7.3.1 Designing Theory-Informed Game Mechanics

After laying out the CET tenets and formulating how rewards should impact player experiences, we needed to create videogame rewards that would match the reward experiences postulated by CET. For this study, we chose three different reward types from the taxonomy of Phillips et al. [161], which should evoke different kinds of experience according to CET [170, 183]. First, a Reward of Praise [161], should constitute a verbal reward in the CET sense [170], which should instill a feeling of competence and therefore foster intrinsic motivation. Second, a Reward of Facility [161] — an increase in the players fire rate — should constitute a tangible reward according to Rigby [170], which in turn should have a functional significance of controlling, therefore thwarting autonomy and subsequently intrinsic motivation. Lastly, we chose a Reward of Glory, which are rewards that are quantifiable, but of not direct use to the gameplay (e.g., a highscore) [161]. In the experimental game, this reward was implemented via coins, dropped after killing an enemy. We chose this reward in addition to rewards of praise and facility (which have been argued to be verbal and tangible rewards respectively) as it was argued to be a tangible reward by Rigby [170], but has no direct value to the player as tangible rewards should have. While prior CET studies have shown a negative impact of rewards with no real value (i.e., badges) on intrinsic motivation [121], a score can also give fine-grained feedback on the players performance, making it hard to predict how players may experience the reward. A brief overview of the rewards and CET based expectations can be found in Table 7.1

While choosing and implementing the reward types was mostly straight forward, as the GDC talk by Rigby [170] provides a design focused breakdown of the CET predictions and the taxonomy by Phillips et al. [161] likewise provides a good guide to videogame rewards, other issues emerged during the design process.

7.4 Results & Discussion

At this time, *Proj. II* is not completed. However, a pre-study with a small sample was conducted to estimate the feasibility of the study design and get outside feedback on the design decisions. The full dataset and analysis are available at https://osf.io/zgb9d/?view_only=8e8099e913174e699dce9c827172d8fc.

First, the impact of rewards is argued to be dependent on rewards being expected [170, 183]. While Rigby [170] argued that rewards in games should come unexpected (e.g., random drops), this was hardly feasible to implement into our wave-based game as (1) after the first wave players would be aware that rewards were given, and (2) — given the aforementioned ubiquity of rewards — players may very well expect to be rewarded in some form when playing a videogame. Moreover, in order to impact motivation, rewards have to be salient. Therefore it is necessary for the rewards to be highlighted in some way. Notably, while we particularly expected the verbal reward to potentially not parse as such (as it is simply the phrase "Good Work"), several participants of the pre-study critiqued not the praise, but both the coins and the fire rate as not being rewards. In regards to the coins, one participant noted that they could not be used for anything and therefore did not feel like a reward, while another participant noted that the fire rate increase was not a reward as they felt it was necessary for the next stage.

In addition to the qualitative feedback, we did use a questionnaire to measure the experience of the rewards as informational and controlling. For this we adopted a scale developed by Thibault Landry et al. [206]. However, the informational and controlling subscales are highly correlated, both in the original validations [206] and our dataset ($r = .67$). This may indicate that while the questionnaire may measure if rewards are perceived, it does not necessarily allow to infer which functional significance was most salient to the receiver. Consequently, we cannot be sure if the reward types in our study evoke the effects we expect them to.

Next, another important aspect of the impact of rewards is their contingency [183]. If an expected reward is contingent on engagement, completion or a certain performance in a task, they may thwart intrinsic motivation, while an expected reward given without requirement does not [55]. In games, however, rewards are rarely not contingent on anything. At the very least, they require players to interact with the game (e.g., kill enemies, solve quests, or login) and should

therefore be engagement contingent. In the experimental game, rewards should be completion contingent, as they require the player to complete a wave. An exception is the reward of glory which is linked to the player's performance as they have to pick up coins for their score to increase. While they cannot get fail to pickup enough coins, they can receive less coins in a wave than they did the one before, resulting in a lower score. This was, in fact, noted by some participants as a point of frustration. This may introduce unwanted effects: The coins add an additional game element compared to the other conditions and the score itself may provide more fine grained performance feedback to the player. This may increase their enjoyment of the game as (1) there is more for them to do, and (2) the detailed feedback may instill more competence satisfaction. This being said, counter to this expectation, in our pre-study the coin reward had the highest reported competence frustration and several participants noted that the level ending before they could pickup the rest of the coins left them feel cheated. This may indicate, that at least for the reward of glory, our specific design implementation may be lackluster.

Lastly, when varying the rewards a player receives, it is impossible to not inadvertently change the game in another way. Per definition, the reward of facility has to have some use in (and therefore impact on) the game [161]: In the specific case of *Proj. II*, the increased fire rate allows players to more quickly dispatch the enemies and therefore beat the given wave. This increase in capabilities may impact the players sense of competence, potentially counteracting the detrimental effect of the reward being experienced as tangible.

Overall, even in the arguably best case scenario where scholars of the theory provide some translations of the theory to game design themselves, many more open questions about any given implementation remain, and the complex nature of games as stimuli make precise manipulations hard, and vulnerable to design errors. Moreover, games speak differently to different people creating a natural range in responses, which complicate analyses — particularly with small sample sizes.

Chapter 8

Discussion

In the previous chapters, I first discussed the role of NHST in empirical PX research, arguing that its confirmatory nature may be suboptimal for the kind of research questions we are interested in: Much empirical PX research is not asking confirmatory research questions, and the theories we work with do not lend themselves to precise hypothesis tests. Instead, I argue, we should perform more exploratory, theory-informed research. I then presented four such research projects — two completed (*Pub. II*, *Pub. III*), one discontinued (*Proj. I*), and one in progress (*Proj. II*) — reflecting on the role and use of theory in each, and the challenges faced during the process. In this last chapter, I will further collate and discuss the challenges of theory-use in PX, but also the opportunities it may provide.

8.1 Opportunities for Theory-Informed Empirical PX Research

As discussed in chapter 2 and 3, theory-use in empirical PX research has been criticized as shallow [209], and wrongly focused on hypothesis testing [9]. But why should we bother to involve theory in our work in the first place?

There is promise in better theory-use: For PX research, working with theory may unify our research efforts, allowing us to answer open questions [9], while the unique position of PX research as an applied science [48], puts us in a valuable position to both inform design practice and talk-back to the theory [16, 209]. Moreover, as illustrated by *Pub. II*and *Pub. III*, working with theory to explore and define phenomena and concepts may allow us to collate and synthesize

our considerable body of descriptive research. While singular descriptive works may provide valuable data, it is only when we combine research via theory, that we can properly utilize them [85], and consequently increasing our field's ability to solve problems [149]. For example, by providing theoretical backing to emotionally moving gaming experiences, we can use the different descriptions of the nature and impact of emotionally moving experiences (i.e., Chapter 4), and shift our questions towards how these experiences come to be (i.e., Chapter 6). In turn, this may help designers to create these experiences, and feed back into existing theories, increasing our knowledge of how we experience media.

Moreover, theory-use may help us during the research process itself. HCI scholars have increasingly called for more transparency throughout the research process [e.g., 11, 104, 216]. While transparency is an important factor to assure research being reliable [136], it does not come without drawbacks. For one, transparency forces the myriad of decisions researchers make during the research process [105, 212, 222] into the open. This requires each decision to be understandable to those evaluating our research [138], for example when filling out a pre-registration [142, 187], submitting a registered report [21, 188], or simply during peer review. While theory-use does not guarantee to make correct decisions, it can help arguing why a given choice of measures, analyses, or stimulus design was made.

8.2 Challenges for Theory-Informed Empirical PX Research

Doing theory-informed research is difficult. While better theory-use is called for in psychology [e.g., 31, 67], HCI [e.g., 149], and PX research [e.g., 9, 209], many challenges can arise during the research process that are seldom acknowledged.

Similarly to design research [176], theory-use in PX research is hindered by the distance between cognitive theories and their implementation in our studies. In the projects discussed here, this manifested in several ways.

First, the claims that theories make about PX phenomena are often vague and verbal [9, 59], making it difficult to determine what effects to expect and how to measure the phenomena we are interest in. As illustrated previously, many aspects of PX, such as the emotional spectrum (Chapter 4 & 6), feelings of dis- and empowerment (Chapter 5), or the functional significance of rewards (Chapter 7) are not measurable with existing questionnaires. Consequently, we either have to rely on qualitative data (e.g., chapter 4 & 5), or adopt questionnaires from other fields which may not be reliable (e.g., chapter 6 & 7); something we can usually only determine after already committing resources to the study. While a solution to this would be the development of more measurement tools, this would in turn require understanding of the underlying

phenomena we seek to construct [72, 187].

Another way this distance manifests is in the challenge of translating theoretical tenets into specific design decisions. Even in scenario when theoretical tenets are seemingly already translated — as with rewards, where SDT scholars themselves have made predictions regarding specific game design elements [e.g., 170] — there can still be a considerable gap between a theory and the actual implementation. This is further exacerbated by the intricate nature of our object of study: As discussed in Chapter 7, videogames are complex and interconnected. Subsequently, manipulating a single, theory-derived game element might impact the player experience in more than the intended ways. When, however, manipulations are reduced to a state where they no longer interfere with any other game elements, such as the context manipulation in *Proj. I*, effects may be reduced, or the relevant manipulation not even triggered as the experimental environment is now too far removed from the actual phenomenon [53, 224].

This distance between theory and design has been an often noted issue in HCI [e.g., 39, 176]. However, the recommendations to alleviate this are primarily aimed at practitioners or design researchers, not at researchers doing theory-informed work. The issues outlined in chapter 6 and 7 however, seem reminiscent of the issues practitioners have been argued to face [16, 176]. Consequently, it may be fruitful for empirical PX research to incorporate design recommendations into our study design — see our practice not just as research, but also as a design practice. For example, Sutcliffe's [203] "generic models" could be used to create a tool-set of game design elements we know can spark certain effects. These could consequently be easier ported into different studies, facilitating more robust experiments. This could potentially alleviate the uncertainties we face when having to decide if it was the theory, or the study design failing.

The uncertainty in theory-informed design for and the measuring certain effects has direct detrimental effects on our ability to say something about the theory we used: If we do not know if we failed to elicit an effect, or if the theory failed to predict the data, we cannot "talk-back" to the theory [16, 209]. Even when designing studies trying to say something about theory (i.e., *Proj. I* & *Proj. II*), the myriad of decisions during operationalization and design can introduce alternative explanations to any given study outcome.

Part of this issue may be our reliance on NHST. As discussed in Chapter 3, NHST is the default approach of empirical PX researchers seeking to perform a quantitative study. However, and as illustrated in Chapter 6, NHST is hard to do right unless the derivation chain from theory to hypothesis test is strong [9, 130, 187], i.e., if we know exactly when the theory applies, how it maps to measurable constructs, how constructs relate, and with which tools to reliably measure and evoke effects. In PX research, this is rarely a given [9].

Therefore, it may be prudent to ask if we in empirical PX research are able to test theory at all (at this point in time). Many of our theories are under-specified — e.g., SDT [59] — and verbal in

nature, which can make them virtually unfalsifiable, as any failure to predict an outcome may be due to any given auxiliary assumption, measurement choice, or design implementation [9]. For example, and as discussed in Chapter 6, *Proj. I* may have non-significant results because the theories were wrong, or due to the participants being not invested enough, the questionnaires not being sensitive enough, the game not evoking the desired effect, or the chosen analysis method not being suitable. As Ballou put it: *"Any failure to find an expected result can be explained away by decisions in the study design or analysis"* [9, p.2:4].

Furthermore, it should be noted that this discussion arises particularly with *Proj. I* and *Proj. II*, as they produced unexpected results; therefore sparking doubt in their ability to talk-back to theory. However, what if these studies had "succeeded", i.e., produced expected results? Would they then be able to speak to the theories, or would the results then be just as likely due to any of the auxiliary assumptions or design decisions [187]?

Consequently — and as pointed out in *Pub. I* and argued for by Ballou [9] — empirical PX research may be better served by working further "up" the Derivation Chain. For example, embracing more descriptive and exploratory research could help tie down auxiliary assumptions, form strong concepts [130, 187].

However, there is rather little guidance on how to do exploratory and descriptive research. Both *Pub. II* and *Pub. III* are such descriptive works, and both struggled with linking the qualitative and quantitative data to specific theories. While several works have recommended more theory-informed exploratory PX work [9, 209] (including *Pub. I*), these recommendations did not give specific pointers on how to do them. There are some recommendations in psychology for works up the Derivation Chain — such as discussing quality markers of concept formation [83] or different types of descriptive work [84] — though these as well remain high-level. At the current point in time, empirical PX research may benefit from a more open discussion on how we practice theory-informed exploratory work, what challenges we face during it, and where we might fail.

8.2.1 Opportunities and Challenge of Failure

In the prior section, I reflected on challenges during empirical PX research projects. Challenges, that in some cases lead to failure. On paper, failure and being wrong is an accepted part of science. For example, at the core of the hypothetico-deductive model lays falsification [164], i.e., proving a theory wrong. For HCI, Hornbæk argued that *we must be more wrong*[98, p.2], as being wrong is the natural occasional outcome of asking interesting and important questions, i.e., questions we do not know the answer to beforehand.

Yet, despite the nominal acceptance of failure, our current work and publishing landscape is

ill-suited to accommodate it. Scholars have long noted the issue of publication biases. Editors and reviewers favor successful (read: significant) studies [181] with little room for imperfections [100]. In fact, acknowledging short-comings and limitations can open researchers up to increased scrutiny [97]. In turn, having research projects fail, or caught up in peer review can jeopardize the acquisition of degrees, or worsen ones academic profile, therefore hurting job and grant prospects [19, 38].

Failure, or even imperfections, having a hard time to find formal publishing may in part be due to it being hard to tell if the failure in questions is the knowledge producing "good" kind of failure or the simply a mistake [e.g., 98]. As both discussed in Chapter 6 and noted by Deterding & Cutting [59], it is generally difficult in empirical PX research to properly attribute experimental failure. The theories we use can be too under-specified, so that any number of auxiliary assumptions or experimental implementations are as likely to be the fault of an unsuccessful experiment as the theory. Consequently, while there is the theoretical possibility of benefiting from failure — revising our beliefs, updating theories, gaining knowledge of how to iterate our practice going forward [99, 196] — given the complexity of the games that we study, the player experiences they evoke, and the theories we use it seems difficult to discern the value of being wrong in a given specific case.

Consequently, it may be understandable that clearly successful papers have a much easier time to be accepted over messy papers of maybe more limited knowledge gain — especially as venues as Conference on Human Factors in Computing Systems (CHI) see ever increasing submission numbers [234], straining editors, and reviewers who have to tend to more and more papers. Still, it is only when being wrong, when failing, that we learn what does not work [98] and which effects do not exist [6].

8.3 Open Questions and Future Work

How should we deal with challenges and failure in research? As discussed in the previous section, there can be value in acknowledging the challenges we face in research, yet, the research culture incentivizes us to not acknowledge weaknesses [97, 100, 233]. In part, this may be due to not all failure being equally productive [98], and deciding how productive any given failed study is can be hard to discern to the nature of our stimuli games (Chapter 6 & 7), as well as the theories [59, 209] and the methodologies we use (Chapter 3).

Yet, if we as a field do want to be more productive [9], if we want to become better at solving problems [149], and if we value openness and transparency [104], we need to appreciate challenges and failure more. Being wrong, being challenged in our understandings and beliefs is

indispensable to progress [98, 164], and if not accepted as such, researchers encountering them will either be forced to abandon projects that could still provide valuable knowledge, or forced to mask issues to present a perfect, polished, unassailable work [138]. Only by highlighting problems we usually do not talk about, but could all encounter, we may find valuable new questions to ask regarding our research process. For example, as many of the issues of theory-use in empirical PX research discussed in this thesis are linked to the design of our stimuli game, should we maybe start consider our practice as a design practices as well, seeing experiments not as one-offs, but iterate on them [e.g., 79] — for example, with perturbation studies [224].

This being said, this would not be without other challenges. For example, NHST would be poor choice of analysis method for this kind of repeated experiments as it, by design, will occasionally produce false results (see Chapter 3). This would be fatal if a falsely significant result would stop the iteration. Moreover, iterative experimental practices would likely require more time and participants, which directly translates into higher costs which many researchers may not be able to stem.

Even when retaining are current approach, empirical PX research may also benefit from a wider array of tools [see also 112]. For example, Bayesian analysis [60, 61] can be used in cases where precise hypothesis testing is not possible. Likewise, exploratory experiments [70, 73, 220] can be useful to better deliminate phenomena, define constructs, or explore interrelations [187].

In conclusion, PX research stands to benefit from more theory-informed research. Tyack & Mekler [209] list a host of fruitful research avenues just for SDT alone, which could substantially improve our understanding of PX. While it is challenging to incorporate theory, it stands to argue that it may become easier the more we do it: The more descriptive research explicitly links to theoretical constructs, the easier it is so synthesize results and built measurements. The more theory-informed experimental games we make (or commercial games we adopt [210]), the easier it will be to identify manipulations that do and do not work. The more we translate and operationalize theory, the easier it will be to work with it in future studies.

Chapter 9

Conclusion

Theory is a fundamental aspect of empirical PX research. It is only by theory that we can conceptualize PX and synthesize our research. Yet, theory-use in empirical PX research is often lacking. This thesis uses findings and experiences from three publications and two unfinished research projects to discuss the challenges when conducting theory-informed empirical player experience research.

This thesis showcases the use and discusses the role of NHST in PX research, pointing out that empirical PX research may be better served by adopting exploratory methods. As there are no practical accounts of how to do theory-informed empirical PX research, I reflect on four of my own research projects. Using two published papers and two unfinished projects, I highlighted how these projects engaged with theory and identified challenges I faced when attempting to work with theory.

During these four reflections, I identify several challenges PX researchers may face when attempting to work with theory: Theory is often too general to readily map onto PX phenomena, PX research lacks robust concepts and measurement tools, and theoretical tenets are difficult to translate into experimental manipulations. Based on these challenges I posit that PX research should adopt more exploratory and design methods, and work closer with theory to built a better understanding of PX phenomena. To facilitate this, I further point out the value of failure in PX research.

Taken together, my thesis gives insight into the practical aspect of succeeding and failing in performing theory-informed PX research and offers impulses for more productive theory-use when studying videogames.

References

[1] Vero Vanden Abeele, Katta Spiel, Lennart Nacke, Daniel Johnson, and Kathrin Gerling. 2020. Development and validation of the player experience inventory: A scale to measure player experiences at the level of functional and psychosocial consequences. *International Journal of Human Computer Studies* 135, October 2019 (2020), 102370–102370. https://doi.org/10.1016/j.ijhcs.2019.102370 Number: October 2019 Publisher: Elsevier Ltd.

[2] Gabriel Abend. 2008. The Meaning of 'Theory'. *Sociological Theory* 26, 2 (June 2008), 173–199. https://doi.org/10.1111/j.1467-9558.2008.00324.x Publisher: SAGE Publications Inc.

[3] Balazs Aczel, Barnabas Szaszi, Alexandra Sarafoglou, Zoltan Kekecs, Šimon Kucharský, Daniel Benjamin, Christopher D Chambers, Agneta Fisher, Andrew Gelman, Morton A Gernsbacher, John P Ioannidis, Eric Johnson, Kai Jonas, Stavroula Kousta, Scott O Lilienfeld, D Stephen Lindsay, Candice C Morey, Marcus Monafò, Benjamin R Newell, Harold Pashler, David R Shanks, Daniel J Simons, Jelte M Wicherts, Dolores Albarracin, Nicole D Anderson, John Antonakis, Hal R Arkes, Mitja D Back, George C Banks, Christopher Beevers, Andrew A Bennett, Wiebke Bleidorn, Ty W Boyer, Cristina Cacciari, Alice S Carter, Joseph Cesario, Charles Clifton, Ronán M Conroy, Mike Cortese, Fiammetta Cosci, Nelson Cowan, Jarret Crawford, Eveline A Crone, John Curtin, Randall Engle, Simon Farrell, Pasco Fearon, Mark Fichman, Willem Frankenhuis, Alexandra M Freund, M Gareth Gaskell, Roger Giner-Sorolla, Don P Green, Robert L Greene, Lisa L Harlow, Fernando Hoces de la Guardia, Derek Isaacowitz, Janet Kolodner, Debra Lieberman, Gordon D Logan, Wendy B Mendes, Lea Moersdorf, Brendan Nyhan, Jeffrey Pollack, Christopher Sullivan, Simine Vazire, and Eric Jan Wagenmakers. 2019. A consensus-based transparency checklist. *Nature Human Behaviour* (Dec. 2019). https://doi.org/10.1038/s41562-019-0772-6

[4] Ernest (Ernest W.) Adams. 2014. *Fundamentals of game design*. New Riders. https://learning.oreilly.com/library/view/fundamentals-of-game/9780133435726/

[5] Lena Fanya Aeschbach, Sebastian A.C. Perrig, Lorena Weder, Klaus Opwis, and Florian Brühlmann. 2021. Transparency in Measurement Reporting: A Systematic Literature Review of CHI PLAY. *Proceedings of the ACM on Human-Computer Interaction* 5, CHI PLAY (Oct. 2021), 233:1–233:21. https://doi.org/10.1145/3474660 Number: CHI PLAY.

[6] Marcel A. L. M. van Assen, Robbie C. M. van Aert, Michèle B. Nuijten, and Jelte M. Wicherts. 2014. Why Publishing Everything Is More Effective than Selective Publishing of Statistically Significant Results. *PLOS ONE* 9, 1 (Jan. 2014), e84896. https://doi.org/10.1371/journal.pone.0084896 Publisher: Public Library of Science.

[7] Ahmad Azadvar and Alessandro Canossa. 2018. UPEQ: ubisoft perceived experience questionnaire: a self-determination evaluation tool for video games. In *Proceedings of the 13th International Conference on the Foundations of Digital Games (FDG '18)*. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3235765.3235780

[8] Marjan Bakker, Annette Van Dijk, and Jelte M. Wicherts. 2012. The Rules of the Game Called Psychological Science. *Perspectives on Psychological Science* 7, 6 (Nov. 2012), 543–554. https://doi.org/10.1177/1745691612459060

[9] Nick Ballou. 2023. A Manifesto for More Productive Psychological Games Research. *Games: Research and Practice* 1, 1 (March 2023), 2:1–2:26. https://doi.org/10.1145/3582929

[10] Nick Ballou and Sebastian Deterding. 2023. 'I Just Wanted to Get It Over and Done With': A Grounded Theory of Psychological Need Frustration in Video Games. *Proceedings of the ACM on Human-Computer Interaction* 7, CHI PLAY (Oct. 2023), 382:217–382:236. https://doi.org/10.1145/3611028

[11] Nicholas Ballou, Vivek R Warriar, and Christoph Sebastian Deterding. 2021. Are You Open? : A Content Analysis of Transparency and Openness Guidelines in HCI Journals. (2021), 11. https://doi.org/10.1145/3411764.3445584

[12] Rahul Banerjee, Jason Yip, Kung Jin Lee, and Zoran Popović. 2016. Empowering Children To Rapidly Author Games and Animations Without Writing Code. In *Proceedings of the The 15th International Conference on Interaction Design and Children*. ACM, Manchester United Kingdom, 230–237. https://doi.org/10.1145/2930674.2930688

[13] Phillip J. Barnard. 1991. Connecting psychological theory to HCl: science, craft or just plain craftiness?. In *IEE Colloquium on Theory in Human-Computer Interaction (HCI)*. IET, 3–1.

[14] Chris Bateman. 2014. Empirical Game Aesthetics. In *Handbook of Digital Games* (1 ed.), Marios C. Angelides and Harry Agius (Eds.). Wiley, 411–443. https://doi.org/10.1002/9781118796443.ch15

[15] Michel Beaudouin-Lafon, Susanne Bødker, and Wendy E. Mackay. 2021. Generative Theories of Interaction. *ACM Transactions on Computer-Human Interaction* 28, 6 (Dec. 2021), 1–54. https://doi.org/10.1145/3468505 Number: 6.

[16] Jordan Beck and Hamid R. Ekbia. 2018. The Theory-Practice Gap as Generative Metaphor. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3173574.3174194

[17] Benjamin B Bederson and Ben Shneiderman. 2003. Theories for Understanding Information Visualization. In *The Craft of Information Visualization*. Elsevier, 349–351. https://doi.org/10.1016/B978-1-55860-915-0.X5000-8

[18] Steve Benford, Chris Greenhalgh, Gabriella Giannachi, Brendan Walker, Joe Marshall, and Tom Rodden. 2012. Uncomfortable interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. Association for Computing Machinery, New York, NY, USA, 2005–2014. https://doi.org/10.1145/2207676.2208347

[19] Lawrence D. Berg, Edward H. Huijbens, and Henrik Gutzon Larsen. 2016. Producing anxiety in the neoliberal university. *Canadian Geographies / Géographies canadiennes* 60, 2 (2016), 168–180. https://doi.org/10.1111/cag.12261 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cag.12261.

[20] D. E. Berlyne. 1974. *Studies in the new experimental aesthetics : steps toward an objective psychology of aesthetic appreciation*. Hemisphere Publishing Corp. https://aalto.finna.fi/Record/alli.754268 Pages: 340.

[21] Lonni Besançon, Anastasia Bezerianos, Pierre Dragicevic, Petra Isenberg, and Yvonne Jansen. 2021. *Publishing Visualization Studies as Registered Reports: Expected Benefits and Researchers' Attitudes*. Technical Report. OSF Preprints. https://doi.org/10.31219/osf.io/3z7kx type: article.

[22] Lonni Besançon and Pierre Dragicevic. 2019. The Continued Prevalence of Dichotomous Inferences at CHI. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–11. https://doi.org/10.1145/3290607.3310432

[23] Lonni Besançon, Florian Echtler, Matthew Kay, and Chat Wacharamanotham. 2024. Experimenting with new review methods, open practices, and interactive publications in HCI. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, 1–4. https://doi.org/10.1145/3613905.3643987

[24] Max V. Birk, Maximilian A. Friehs, and Regan L. Mandryk. 2017. Age-Based Preferences and Player Experience: A Crowdsourced Cross-sectional Study. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM, Amsterdam The Netherlands, 157–170. https://doi.org/10.1145/3116595.3116608

[25] Max V. Birk, Ioanna Iacovides, Daniel Johnson, and Regan L. Mandryk. 2015. The False Dichotomy between Positive and Negative Affect in Game Play. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*. ACM Press, London, United Kingdom, 799–804. https://doi.org/10.1145/2793107.2810258

[26] Julia Bopp, Jan Vornhagen, Roosa Piitulainen, Barbara Keller, and Elisa Mekler. 2022. GamesAsArt. https://doi.org/10.17605/OSF.IO/RYVT6 Publisher: Open Science Framework.

[27] Julia Ayumi Bopp. 2020. *Aesthetic Emotions in Digital Games: The Appeal of Moving, Challenging, and Thought-Provoking Player Experiences*. Aalto University. https://aaltodoc.aalto.fi:443/handle/123456789/47415 Accepted: 2020-11-04T10:00:06Z ISSN: 1799-4942 (electronic).

[28] Julia Ayumi Bopp, Elisa D. Mekler, and Klaus Opwis. 2016. Negative Emotion, Positive Experience? Emotionally Moving Moments in Digital Games. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2996–3006. https://doi.org/10.1145/2858036.2858227

[29] Julia Ayumi Bopp, Klaus Opwis, and Elisa D. Mekler. 2018. "An Odd Kind of Pleasure": Differentiating Emotional Challenge in Digital Games. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173615

[30] Julia A. Bopp, Jan B. Vornhagen, and Elisa D. Mekler. 2021. "My Soul Got a Little Bit Cleaner": Art Experience in Videogames. *Proceedings of the ACM on Human-Computer Interaction* 5, CHI PLAY (Oct. 2021), 237:1–237:19. https://doi.org/10.1145/3474664

[31] Denny Borsboom, Han L. J. van der Maas, Jonas Dalege, Rogier A. Kievit, and Brian D. Haig. 2021. Theory Construction Methodology: A Practical Framework for Building Theories in Psychology. *Perspectives on Psychological Science* 16, 4 (July 2021), 756–766. https://doi.org/10.1177/1745691620969647

[32] Virginia Braun and Victoria Clarke. 2022. *Thematic analysis: a practical guide*. SAGE, London ; Thousand Oaks, California. OCLC: on1247204005.

[33] David Brieber, Helmut Leder, and Marcos Nadal. 2015. The experience of art in museums: An attempt to dissociate the role of physical context and genuineness. *Empirical Studies of the Arts* 33, 1 (2015), 95–105. https://doi.org/10.1177/0276237415570000 Number: 1.

[34] David Brieber, Marcos Nadal, and Helmut Leder. 2015. In the white cube: Museum context enhances the valuation and memory of art. *Acta Psychologica* 154 (Jan. 2015), 36–42. https://doi.org/10.1016/j.actpsy.2014.11.004 Publisher: Elsevier B.V..

[35] David Brieber, Marcos Nadal, Helmut Leder, and Raphael Rosenberg. 2014. Art in time and space: Context modulates the relation between art experience and viewing time. *PLoS ONE* 9, 6 (2014), 1–8. https://doi.org/10.1371/journal.pone.0099019 Number: 6.

[36] Kaitlyn Burnell. 2012. Breaking the Rules of Game Design: When to Go Against Competence, Autonomy and Relatedness. https://www.gdcvault.com/play/1015398/Breaking-the-Rules-of-Game Presenters: _:n11580.

[37] Paul Cairns. 2007. HCI.. not as it should be: Inferential statistics in HCI research. In *People and Computers XXI HCI.But Not as We Know It - Proceedings of HCI 2007: The 21st British HCI Group Annual Conference*, Vol. 1. 195–201. https://doi.org/10.14236/ewic/hci2007.20

[38] Martina Angela Caretta, Danielle Drozdzewski, Johanna Carolina Jokinen, and Emily Falconer. 2018. " *Who* can play *this* game?" The lived experiences of doctoral candidates and early career women in the neoliberal university. *Journal of Geography in Higher Education* 42, 2 (April 2018), 261–275. https://doi.org/10.1080/03098265.2018.1434762

[39] John M. Carroll. 2003. Introduction: Toward a Multidisciplinary Science of Human-Computer Interaction. In *HCI Models, Theories, and Frameworks*. Elsevier, 1–9. https://doi.org/10.1016/B978-155860808-5/50001-0

[40] John M. Carroll and Mary Beth Rosson. 2003. Design Rationale as Theory. In *HCI Models, Theories, and Frameworks*. Elsevier, 431–461. https://doi.org/10.1016/B978-155860808-5/50015-0

[41] Center For Self-Determination Theory. [n. d.]. Intrinsic Motivation Inventory (IMI). https://selfdeterminationtheory.org/intrinsic-motivation-inventory/

[42] Anjan Chatterjee and Oshin Vartanian. 2016. Neuroscience of aesthetics. *Annals of the New York Academy of Sciences* 1369, 1 (2016). https://doi.org/10.1111/nyas.13035 Number: 1.

[43] Beiwen Chen, Maarten Vansteenkiste, Wim Beyers, Liesbet Boone, Edward L. Deci, Jolene Van der Kaap-Deeder, Bart Duriez, Willy Lens, Lennia Matos, Athanasios Mouratidis, Richard M. Ryan, Kennon M. Sheldon, Bart Soenens, Stijn Van Petegem, and Joke Verstuyf. 2015. Basic psychological need satisfaction, need frustration, and need strength across four cultures. *Motivation and Emotion* 39, 2 (April 2015), 216–236. https://doi.org/10.1007/s11031-014-9450-1 Number: 2.

[44] Andy Cockburn, Pierre Dragicevic, Lonni Besançon, and Carl Gutwin. 2020. Threats of a replication crisis in empirical computer science. *Commun. ACM* 63, 8 (2020), 70–79. https://doi.org/10.1145/3360311 Number: 8.

[45] Andy Cockburn, Carl Gutwin, and Alan Dix. 2018. HARK No More: On the Preregistration of CHI Experiments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, Montreal QC, Canada, 1–12. https://doi.org/10.1145/3173574.3173715

[46] David Colaço. 2018. Rethinking the role of theory in exploratory experimentation. *Biology & Philosophy* 33, 5-6 (Dec. 2018), 38. https://doi.org/10.1007/s10539-018-9648-9

[47] Tom Cole, Paul Cairns, and Marco Gillies. 2015. Emotional and Functional Challenge in Core and Avant-garde Games. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15 (CHI PLAY '15)*. ACM Press, New York, New York, USA, 121–126. https://doi.org/10.1145/2793107.2793147

[48] Lucas Colusso, Ridley Jones, Sean A. Munson, and Gary Hsieh. 2019. A Translational Science Model for HCI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300231

[49] Ben Cowley, Darryl Charles, Michaela Black, and Ray Hickey. 2008. Toward an understanding of flow in video games. *Computers in Entertainment* 6, 2 (2008), 1–1. https://doi.org/10.1145/1371216.1371223 Number: 2.

[50] Mihaly Csikszentmihalyi. 2014. *Flow and the Foundations of Positive Psychology*. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-94-017-9088-8

[51] Geoff Cumming and Sue Finch. 2005. Inference by Eye: Confidence Intervals and How to Read Pictures of Data. *American Psychologist* 60, 2 (2005), 170–180. https://doi.org/10.1037/0003-066X.60.2.170 Number: 2.

[52] Peter Dalsgaard and Christian Dindler. 2014. Between theory and practice: bridging concepts in HCI research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Toronto Ontario Canada, 1635–1644. https://doi.org/10.1145/2556288.2557342

[53] Kurt Danziger. 1998. *Constructing the subject: historical origins of psychological research* (reprinted ed.). Cambridge Univ. Press, Cambridge.

[54] Travis Day. 2019. Rewards in Video Games. https://www.youtube.com/watch?v=urijgWXLYck

[55] Edward L. Deci, Richard Koestner, and Richard M. Ryan. 1999. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin* 125 (1999), 627–668. https://doi.org/10.1037/0033-2909.125.6.627 Place: US Publisher: American Psychological Association.

[56] Alena Denisova, Julia Ayumi Bopp, Thuy Duong Nguyen, and Elisa D Mekler. 2021. "Whatever the Emotional Experience, It's Up to Them": Insights from Designers of Emotionally Impactful Games. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–9. https://doi.org/10.1145/3411764.3445286

[57] Alena Denisova and Eliott Cook. 2019. Power-Ups in Digital Games: The Rewarding Effect of Phantom Game Elements on Player Experience. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '19)*. Association for Computing Machinery, New York, NY, USA, 161–168. https://doi.org/10.1145/3311350.3347173

[58] Ansgar E. Depping and Regan L. Mandryk. 2017. Why is This Happening to Me?: How Player Attribution can Broaden our Understanding of Player Experience. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, Denver Colorado USA, 1040–1052. https://doi.org/10.1145/3025453.3025648

[59] Sebastian Deterding and Joe Cutting. 2023. Objective Difficulty-Skill Balance Impacts Perceived Balance but Not Behaviour: A Test of Flow and Self-Determination Theory Predictions. *Proceedings of the ACM on Human-Computer Interaction* 7, CHI PLAY (Sept. 2023), 1179–1205. https://doi.org/10.1145/3611065

[60] Zoltan Dienes. 2008. *Understanding psychology as a science: An introduction to scientific and statistical inference*. Macmillan International Higher Education.

[61] Zoltan Dienes. 2014. Using Bayes to get the most out of non-significant results. *Frontiers in Psychology* 5 (July 2014), 781–781. https://doi.org/10.3389/fpsyg.2014.00781 Publisher: Frontiers Media SA.

[62] Anders Drachen, Pejman Mirza-Babaei, and Lennart E. Nacke. 2018. *Games User Research*. Oxford University Press. Google-Books-ID: W3FGDwAAQBAJ.

[63] Pierre Dragicevic. 2016. Fair Statistical Communication in HCI. In *Modern Statistical Methods for HCI*. 291–330. https://doi.org/10.1007/978-3-319-26633-6_13

[64] Luís Duarte and Luís Carriço. 2012. Power me Up! an interactive and physiological perspective on videogames' temporary bonus rewards. In *Proceedings of the 4th International Conference on Fun and Games (FnG '12)*. Association for Computing Machinery, New York, NY, USA, 55–63. https://doi.org/10.1145/2367616.2367623

[65] Alexander Eiselmayer, Chat Wacharamanotham, Michel Beaudouin-Lafon, and Wendy E. Mackay. 2019. Touchstone2: An Interactive Environment for Exploring Trade-offs in HCI Experiment Design. In *Conference on Human Factors in Computing Systems - Proceedings*. 1–11. https://doi.org/10.1145/3290605.3300447

[66] Jack Emmert. 2007. To Reward or Not to Reward? How to Create a Positive Community in an MMO World. https://www.gdcvault.com/play/578/To-Reward-or-Not-to

[67] Markus I. Eronen and Laura F. Bringmann. 2021. The Theory Crisis in Psychology: How to Move Forward. *Perspectives on Psychological Science* 16, 4 (July 2021), 779–788. https://doi.org/10.1177/1745691620970586

[68] Gustav Theodor Fechner. 2013. *Vorschule der Aesthetik, Volume 1*. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781139854580

[69] Uljana Feest. 2017. Phenomena and Objects of Research in the Cognitive and Behavioral Sciences. *Philosophy of Science* 84, 5 (Dec. 2017), 1165–1176. https://doi.org/10.1086/694155

[70] Uljana Feest and Friedrich Steinle (Eds.). 2012. *Scientific concepts and investigative practice*. Number volume 3 in Berlin studies in knowledge research. De Gruyter, Berlin.

[71] Andy P. Field. 2009. *Discovering statistics using SPSS: and sex, drugs and rock 'n' roll* (3rd ed ed.). SAGE Publications, Los Angeles.

[72] Jessica K. Flake, Jolynn Pek, and Eric Hehman. 2017. Construct Validation in Social and Personality Research: Current Practice and Recommendations. *Social Psychological and Personality Science* 8, 4 (May 2017), 370–378. https://doi.org/10.1177/1948550617693063 Number: 4 Publisher: SAGE Publications Inc.

[73] L. R. Franklin. 2005. Exploratory Experiments. *Philosophy of Science* 72, 5 (Dec. 2005), 888–899. https://doi.org/10.1086/508117

[74] Eiko I. Fried. 2020. Lack of Theory Building and Testing Impedes Progress in The Factor and Network Literature. *Psychological Inquiry* 31, 4 (Oct. 2020), 271–288. https://doi.org/10.1080/1047840X.2020.1853461 Number: 4 Publisher: Routledge _eprint: https://doi.org/10.1080/1047840X.2020.1853461.

[75] Julian Frommel, Madison Klarkowski, and Regan L. Mandryk. 2021. The Struggle is Spiel: On Failure and Success in Games. In *The 16th International Conference on the Foundations of Digital Games (FDG) 2021 (FDG'21)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3472538.3472565

[76] Tracy Fullerton. 2019. *Game design workshop: a playcentric approach to creating innovative games* (fourth edition ed.). Taylor & Francis, CRC Press, Boca Raton.

[77] Alf Gabrielsson and Erik Lindström. 2010. The role of structure in the musical expression of emotions. In *Handbook of music and emotion: Theory, research, applications*. Oxford University Press, New York, NY, US, 367–400.

[78] Andrew Gambino and Bingjie Liu. 2022. Considering the Context to Build Theory in HCI, HRI, and HMC: Explicating Differences in Processes of Communication and Socialization with Social Technologies. *Human-Machine Communication* 4, 1 (April 2022). https://doi.org/10.30658/hmc.4.6 Number: 1.

[79] William Gaver, John Bowers, Tobie Kerridge, Andy Boucher, and Nadine Jarvis. 2009. Anatomy of a failure: how we knew when our design went wrong, and what we learned from it. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Boston MA USA, 2213–2222. https://doi.org/10.1145/1518701.1519040

[80] Bertram Gawronski and Galen V. Bodenhausen (Eds.). 2015. *Theory and explanation in social psychology*. The Guilford Press, New York, NY.

[81] Gernot Gerger, Helmut Leder, and Alexandra Kremer. 2014. Context effects on emotional and aesthetic evaluations of artworks and IAPS pictures. *Acta Psychologica* 151 (Sept. 2014), 174–183. https://doi.org/10.1016/j.actpsy.2014.06.008

[82] Kathrin Maria Gerling, Matthew Miller, Regan L. Mandryk, Max Valentin Birk, and Jan David Smeddinck. 2014. Effects of balancing for physical abilities on player performance, experience and self-esteem in exergames. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Toronto Ontario Canada, 2201–2210. https://doi.org/10.1145/2556288.2556963

[83] John Gerring. 1999. What Makes a Concept Good? A Criterial Framework for Understanding Concept Formation in the Social Sciences. *Polity* 31, 3 (March 1999), 357–393. https://doi.org/10.2307/3235246

[84] John Gerring. 2012. Mere Description. *British Journal of Political Science* 42, 4 (Oct. 2012), 721–746. https://doi.org/10.1017/S0007123412000130

[85] Gerd Gigerenzer. 1998. Surrogates for Theories. *Theory & Psychology* 8, 2 (April 1998), 195–204. https://doi.org/10.1177/0959354398082006

[86] Gerd Gigerenzer. 2004. Mindless statistics. *Journal of Socio-Economics* 33, 5 (2004), 587–606. https://doi.org/10.1016/j.socec.2004.09.033 Number: 5.

[87] Christopher Gile. 2016. Achievements are Permission. https://www.gamedeveloper.com/design/achievements-are-permission

[88] Lahari Goswami, Amelia McNamara, Viktorija Paneva, Jan Benjamin Vornhagen, and Erich Weichselgartner. 2022. A Cheat Sheet for a Transparent CHI paper. (Sept. 2022). https://doi.org/10.17605/OSF.IO/YHWUQ Publisher: OSF.

[89] Chad Phoenix Rose Gowler and Ioanna Iacovides. 2019. "Horror, guilt and shame" – Uncomfortable Experiences in Digital Games. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '19*. ACM Press, New York, New York, USA, 325–337. https://doi.org/10.1145/3311350.3347179

[90] Roland Graf, Pallavi Benawri, Amy E. Whitesall, Dashiell Carichner, Zixuan Li, Michael Nebeling, and Hun Seok Kim. 2019. iGYM: An Interactive Floor Projection System for Inclusive Exergame Environments. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM, Barcelona Spain, 31–43. https://doi.org/10.1145/3311350.3347161

[91] Kurt Gray. 2017. How to Map Theory: Reliable Methods Are Fruitless Without Rigorous Theory. *Perspectives on Psychological Science* 12, 5 (Sept. 2017), 731–741. https://doi.org/10.1177/1745691617691949

[92] Shirley Gregor. 2006. The Nature of Theory in Information Systems. *MIS Quarterly* 30, 3 (2006), 611–642. https://doi.org/10.2307/25148742 Publisher: Management Information Systems Research Center, University of Minnesota.

[93] Christian Guckelsberger, Christoph Salge, Jeremy Gow, and Paul Cairns. 2017. Predicting Player Experience without the Player.: An Exploratory Study. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM, Amsterdam The Netherlands, 305–315. https://doi.org/10.1145/3116595.3116631

[94] Olivia Guest and Andrea E. Martin. 2021. How Computational Modeling Can Force Theory Building in Psychological Science. *Perspectives on Psychological Science* 16, 4 (July 2021), 789–802. https://doi.org/10.1177/1745691620970585

[95] Christine A. Halverson. 2002. Activity Theory and Distributed Cognition: Or What Does CSCW Need to DO with Theories? *Computer Supported Cooperative Work (CSCW)* 11, 1-2 (March 2002), 243–267. https://doi.org/10.1023/A:1015298005381

[96] Eric B. Hekler, Predrag Klasnja, Jon E. Froehlich, and Matthew P. Buman. 2013. Mind the theoretical gap: interpreting, using, and developing behavioral theory in HCI research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Paris France, 3307–3316. https://doi.org/10.1145/2470654.2466452

[97] Rink Hoekstra and Simine Vazire. 2021. Aspiring to greater intellectual humility in science. *Nature Human Behaviour* (Oct. 2021). https://doi.org/10.1038/s41562-021-01203-8

[98] Kasper Hornbæk. 2015. We must be more wrong in HCI research. *Interactions* 22, 6 (Oct. 2015), 20–21. https://doi.org/10.1145/2833093 Number: 6.

[99] Kasper Hornbæk, Søren S. Sander, Javier Andrés Bargas-Avila, and Jakob Grue Simonsen. 2014. Is once enough?: on the extent and content of replications in human-computer interaction. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*. ACM Press, Toronto, Ontario, Canada, 3523–3532. https://doi.org/10.1145/2556288.2557004

[100] Noura Howell, Audrey Desjardins, and Sarah Fox. 2021. Cracks in the Success Narrative: Rethinking Failure in Design Research through a Retrospective Trioethnography. *ACM Transactions on Computer-Human Interaction* 28, 6 (Dec. 2021), 1–31. https://doi.org/10.1145/3462447

[101] Kristina Höök and Jonas Löwgren. 2012. Strong concepts: Intermediate-level knowledge in interaction design research. *ACM Trans. Comput.-Hum. Interact.* 19, 3 (Oct. 2012), 23:1–23:18. https://doi.org/10.1145/2362364.2362371

[102] Ioanna Iacovides and Anna L. Cox. 2015. Moving Beyond Fun: Evaluating Serious Experience in Digital Games. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. ACM Press, Seoul, Republic of Korea, 2245–2254. https://doi.org/10.1145/2702123.2702204

[103] Matthew Jackson. 2016. Intrinsic vs Extrinsic Rewards: Why you Need Both. https://www.gamedeveloper.com/design/intrinsic-vs-extrinsic-rewards-why-you-need-both

[104] Yvonne Jansen, Jan Benjamin Vornhagen, Olga Iarygina, Kavous Salehzadeh Niksirat, Lonni Besançon, Pierre Dragicevic, Julien Gori, and Chat Wacharamanotham. 2024. The Many Ways of Being Transparent in Human-Computer Interaction Research. https://doi.org/10.31219/osf.io/2wze6

[105] Leslie K. John, George Loewenstein, and Drazen Prelec. 2012. Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychological Science* 23, 5 (2012), 524–532. https://doi.org/10.1177/0956797611430953 Number: 5.

[106] Daniel Johnson, Madison Klarkowski, Kellie Vella, Cody Phillips, Mitchell McEwan, and Christopher N. Watling. 2018. Greater rewards in videogames lead to more presence, enjoyment and effort. *Computers in Human Behavior* 87 (Oct. 2018), 66–74. https://doi.org/10.1016/j.chb.2018.05.025

[107] Mack H. Jones. 1978. Black Political Empowerment in Atlanta: Myth and Reality. *The ANNALS of the American Academy of Political and Social Science* 439, 1 (Sept. 1978), 90–117. https://doi.org/10.1177/000271627843900108 Number: 1.

[108] Eunice Jun, Melissa Birchfield, Nicole De Moura, Jeffrey Heer, and René Just. 2022. Hypothesis Formalization: Empirical Findings, Software Limitations, and Design Implications. *ACM Transactions on Computer-Human Interaction* 29, 1 (Jan. 2022), 6:1–6:28. https://doi.org/10.1145/3476980

[109] Jesper Juul. 2013. *The Art of Failure: An Essay on the Pain of PLaying Video Games*. The MIT Press. Pages: 173.

[110] Kristine Jørgensen. 2016. The Positive Discomfort of Spec Ops: The Line. *Game Studies* 16, 2 (Dec. 2016). http://gamestudies.org/1602/articles/jorgensenkristine Number: 2.

[111] Rosabeth Moss Kanter. 1977. *Men and women of the corporation*. Basic Books, New York.

[112] Maurits Kaptein and Judy Robertson. 2012. Rethinking statistical analysis methods for CHI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1105–1114. https://doi.org/10.1145/2207676.2208557

[113] Matthew Kay, Steve Haroz, Shion Guha, and Pierre Dragicevic. 2016. Special Interest Group on Transparent Statistics in HCI. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. Association for Computing Machinery, New York, NY, USA, 1081–1084. https://doi.org/10.1145/2851581.2886442

[114] Alexander S. Klyubin, Daniel Polani, and Chrystopher L. Nehaniv. 2008. Keep Your Options Open: An Information-Based Driving Principle for Sensorimotor Systems. *PLoS ONE* 3, 12 (Dec. 2008), e4018. https://doi.org/10.1371/journal.pone.0004018 Number: 12.

[115] Vassilis Kostakos. 2015. The big hole in HCI research. *interactions* 22, 2 (Feb. 2015), 48–51. https://doi.org/10.1145/2729103

[116] Thomas S. Kuhn. 1996. *The structure of scientific revolutions* (3rd ed ed.). University of Chicago Press, Chicago, IL.

[117] Helmut Leder, Benno Belke, Andries Oeberst, and Dorothee Augustin. 2004. A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology* 95, 4 (Nov. 2004), 489–508. https://doi.org/10.1348/0007126042369811 Number: 4.

[118] Helmut Leder, Claus-Christian Carbon, and Ai-Leen Ripsas. 2006. Entitling art: Influence of title information on understanding and appreciation of paintings. *Acta Psychologica* 121, 2 (Feb. 2006), 176–198. https://doi.org/10.1016/j.actpsy.2005.08.005 Number: 2.

[119] Lauri Lehtonen, Maximus D. Kaos, Raine Kajastila, Leo Holsti, Janne Karsisto, Sami Pekkola, Joni Vähämäki, Lassi Vapaakallio, and Perttu Hämäläinen. 2019. Movement Empowerment in a Multiplayer Mixed-Reality Trampoline Game. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM, Barcelona Spain, 19–29. https://doi.org/10.1145/3311350.3347181

[120] Rui Leitão, Martin Maguire, Sarah Turner, and Laura Guimarães. 2022. A systematic evaluation of game elements effects on students' motivation. *Education and Information Technologies* 27, 1 (Jan. 2022), 1081–1103. https://doi.org/10.1007/s10639-021-10651-8

[121] Mark R. Lepper, David Greene, and Richard E. Nisbett. 1973. Undermining children's intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis. *Journal of Personality and Social Psychology* 28, 1 (Oct. 1973), 129–137. https://doi.org/10.1037/h0035519

[122] Yongcheng Liu. 2021. Player pleasure: exciting rewards. https://www.gamedeveloper.com/design/player-pleasure-exciting-rewards

[123] Yong Liu, Jorge Goncalves, Denzil Ferreira, Bei Xiao, Simo Hosio, and Vassilis Kostakos. 2014. CHI 1994-2013: mapping two decades of intellectual progress through co-word analysis. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*. ACM Press, Toronto, Ontario, Canada, 3553–3562. https://doi.org/10.1145/2556288.2556969

[124] J. Long. 1991. Theory in human-computer interaction?. In *IEE Colloquium on Theory in Human-Computer Interaction (HCI)*. 2/1–2/6.

[125] Elizabeth J. Lyons. 2015. Cultivating Engagement and Enjoyment in Exergames Using Feedback, Challenge, and Rewards. *Games for Health Journal* 4, 1 (Feb. 2015), 12–18. https://doi.org/10.1089/g4h.2014.0072 Number: 1.

[126] Tim Marsh and Brigid Costello. 2012. Experience in Serious Games: Between Positive and Serious Experience. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 7528 LNCS. 255–267. https://doi.org/10.1007/978-3-642-33687-4_22

[127] Stefano Mastandrea and William D. Crano. 2019. Peripheral Factors Affecting the Evaluation of Artworks. *Empirical Studies of the Arts* 37, 1 (Jan. 2019), 82–91. https://doi.org/10.1177/0276237418790916 Number: 1 Publisher: SAGE Publications Inc.

[128] Philipp Mayring. 2004. Qualitative Content Analysis. In *A Companion to Qualitative Research*, Uwe Flick, Ernst von Kardoff, and Ines Steinke (Eds.). SAGE. Google-Books-ID: lRSL1KJjEPoC.

[129] Paul E. Meehl. 1967. Theory-Testing in Psychology and Physics: A Methodological Paradox. *Philosophy of Science* 34, 2 (June 1967), 103–115. https://doi.org/10.1086/288135 Publisher: Cambridge University Press.

[130] Paul E. Meehl. 1990. Why Summaries of Research on Psychological Theories are Often Uninterpretable. *Psychological Reports* 66, 1 (Feb. 1990), 195–244. https://doi.org/10.2466/pr0.1990.66.1.195 Publisher: SAGE Publications Inc.

[131] Elisa D. Mekler, Julia Ayumi Bopp, Alexandre N. Tuch, and Klaus Opwis. 2014. A systematic review of quantitative studies on the enjoyment of digital entertainment games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 927–936. https://doi.org/10.1145/2556288.2557078

[132] Elisa D. Mekler, Florian Brühlmann, Alexandre N. Tuch, and Klaus Opwis. 2017. Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior* 71 (June 2017), 525–534. https://doi.org/10.1016/j.chb.2015.08.048

[133] Winfried Menninghaus, Valentin Wagner, Julian Hanich, Eugen Wassiliwizky, Thomas Jacobsen, and Stefan Koelsch. 2017. The Distancing-Embracing model of the enjoyment of negative emotions in art reception. *Behavioral and Brain Sciences* 40 (2017). https://doi.org/10.1017/S0140525X17000309

[134] Tsuyoshi Miyakawa. 2020. No raw data, no science: another possible source of the reproducibility crisis. *Molecular brain* 13, 1 (2020), 24–24. https://doi.org/10.1186/s13041-020-0552-2 Number: 1 Publisher: Molecular Brain.

[135] Torill Elvira Mortensen and Kristine Jørgensen. 2020. *The Paradox of Transgression in Games*. Routledge, London. https://doi.org/10.4324/9780367816476

[136] Marcus R. Munafò, Brian A. Nosek, Dorothy V.M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie Du Sert, Uri Simonsohn, Eric Jan Wagenmakers, Jennifer J. Ware, and John P.A. Ioannidis. 2017. A manifesto for reproducible science. *Nature Human Behaviour* 1, 1 (2017), 1–9. https://doi.org/10.1038/s41562-016-0021 Number: 1 Publisher: Macmillan Publishers Limited.

[137] Brad A. Myers. 1998. A brief history of human-computer interaction technology. *Interactions* 5, 2 (March 1998), 44–54. https://doi.org/10.1145/274430.274436

[138] C. Thi Nguyen. 2021. Transparency is Surveillance. *Philosophy and Phenomenological Research* (2021). https://doi.org/10.1111/phpr.12823 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/phpr.12823.

[139] Raymond S. Nickerson. 2000. Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods* 5, 2 (2000), 241–301. https://doi.org/10.1037/1082-989X.5.2.241 Number: 2.

[140] Noodlecake Studios. 2017. *Getting Over It with Bennett Foddy*.

[141] B. A. Nosek, G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, C. D. Chambers, G. Chin, G. Christensen, M. Contestabile, A. Dafoe, E. Eich, J. Freese, R. Glennerster, D. Goroff, D. P. Green, B. Hesse, M. Humphreys, J. Ishiyama, D. Karlan, A. Kraut, A. Lupia, P. Mabry, T. Madon, N. Malhotra, E. Mayo-Wilson, M. McNutt, E. Miguel, E. Levy Paluck, U. Simonsohn, C. Soderberg, B. A. Spellman, J. Turitto, G. VandenBos, S. Vazire, E. J. Wagenmakers, R. Wilson, and T. Yarkoni. 2015. Promoting an open research culture. *Science* 348, 6242 (June 2015), 1422–1425. https://doi.org/10.1126/science.aab2374 Publisher: American Association for the Advancement of Science.

[142] Brian A. Nosek, Charles R. Ebersole, Alexander C. DeHaven, and David T. Mellor. 2018. The preregistration revolution. *Proceedings of the National Academy of Sciences of the United States of America* 115, 11 (2018), 2600–2606. https://doi.org/10.1073/pnas.1708274114 Number: 11.

[143] Brian A Nosek, Tom E Hardwicke, Hannah Moshontz, Aurélien Allard, Katherine S Corker, Anna Dreber, Fiona Fidler, Joe Hilgard, Melissa Kline Struhl, Michèle B Nuijten, Julia M Rohrer, Felipe Romero, Anne M Scheel, Laura D Scherer, Felix D Schönbrodt, and Simine Vazire. 2022. Replicability, Robustness, and Reproducibility in Psychological Science. (2022). https://doi.org/10.1146/annurev-psych-020821114157

[144] Karen Nylund-Gibson and Andrew Young Choi. 2018. Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science* 4, 4 (Dec. 2018), 440–461. https://doi.org/10.1037/tps0000176

[145] Klaus Oberauer and Stephan Lewandowsky. 2019. Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review* 26, 5 (Oct. 2019), 1596–1618. https://doi.org/10.3758/s13423-019-01645-2

[146] Mary Beth Oliver and Anne Bartsch. 2010. Appreciation as audience response: Exploring entertainment gratifications beyond hedonism. *Human Communication Research* 36, 1 (2010), 53–81. https://doi.org/10.1111/j.1468-2958.2009.01368.x Number: 1.

[147] Mary Beth Oliver, Nicholas David Bowman, Julia K. Woolley, Ryan Rogers, Brett I. Sherrick, and Mun-Young Chung. 2016. Video games as meaningful entertainment experiences. *Psychology of Popular Media Culture* 5, 4 (Oct. 2016), 390–405. https://doi.org/10.1037/ppm0000066 Number: 4.

[148] Maureen A. O'Malley. 2007. Exploratory Experimentation and Scientific Practice: Metagenomics and the Proteorhodopsin Case. *History and Philosophy of the Life Sciences* 29, 3 (2007), 337–360. https://www.jstor.org/stable/23334265 Publisher: Stazione Zoologica Anton Dohrn - Napoli.

[149] Antti Oulasvirta and Kasper Hornbæk. 2016. HCI Research as Problem-Solving. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 4956–4967. https://doi.org/10.1145/2858036.2858283

[150] Antti Oulasvirta and Kasper Hornbæk. 2022. Counterfactual Thinking: What Theories *Do* in Design. *International Journal of Human–Computer Interaction* 38, 1 (Jan. 2022), 78–92. https://doi.org/10.1080/10447318.2021.1925436

[151] Juneyoung Park, Seunghyun Kim, Auk Kim, and Mun Y. Yi. 2019. Learning to be better at the game: Performance vs. completion contingent reward for game-based learning. *Computers & Education* 139, C (Oct. 2019), 1–15. https://doi.org/10.1016/j.compedu.2019.04.016

[152] Matthew Pelowski and Fuminori Akiba. 2011. A model of art perception, evaluation and emotion in transformative aesthetic experience. *New Ideas in Psychology* 29, 2 (2011), 80–97. https://doi.org/10.1016/j.newideapsych.2010.04.001 Number: 2.

[153] Matthew Pelowski, Michael Forster, Pablo P. L. Tinio, Maria Scholl, and Helmut Leder. 2017. Beyond the lab: An examination of key factors influencing interaction with 'real' and museum-based art. *Psychology of Aesthetics, Creativity, and the Arts* 11, 3 (Aug. 2017), 245–264. https://doi.org/10.1037/aca0000141 Number: 3.

[154] Matthew Pelowski, Young-Jin Hur, Katherine N. Cotter, Tomohiro Ishizu, Alexander P. Christensen, Helmut Leder, and I. C. McManus. 2021. Quantifying the if, the when, and the what of the sublime: A survey and latent class analysis of incidence, emotions, and distinct varieties of personal sublime experiences. *Psychology of Aesthetics, Creativity, and the Arts* 15, 2 (May 2021), 216–240. https://doi.org/10.1037/aca0000273

[155] Matthew Pelowski, Patrick S. Markey, Michael Forster, Gernot Gerger, and Helmut Leder. 2017. Move me, astonish me… delight my eyes and brain: The Vienna Integrated Model of top-down and bottom-up processes in Art Perception (VIMAP) and corresponding affective, evaluative, and neurophysiological correlates. *Physics of Life Reviews* 21 (July 2017), 80–125. https://doi.org/10.1016/j.plrev.2017.02.003 Publisher: Elsevier B.V..

[156] Matthew John Pelowski. 2015. Tears and transformation: feeling like crying as an indicator of insightful or "aesthetic" experience with art. *Frontiers in Psychology* 0 (2015). https://doi.org/10.3389/fpsyg.2015.01006 Publisher: Frontiers.

[157] Ella Peltonen, Nitinder Mohan, Peter Zdankin, Tanya Shreedhar, Tri Nguyen, Suzan Bayhan, Jon Crowcroft, Jussi Kangasharju, and Daniela Nicklas. 2023. Perspectives on Negative Research Results in Pervasive Computing. *IEEE Pervasive Computing* 22, 3 (July 2023), 63–72. https://doi.org/10.1109/MPRV.2023.3273718

[158] Xiaolan Peng, Jin Huang, Alena Denisova, Hui Chen, Feng Tian, and Hongan Wang. 2020. A Palette of Deepened Emotions: Exploring Emotional Challenge in Virtual Reality Games. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3313831.3376221

[159] Charlotte R. Pennington. 2023. *A student's guide to open science: using the replication crisis to reform psychology*. Open University Press, Maidenhead.

[160] Serge Petralito, Florian Brühlmann, Glena Iten, Elisa D. Mekler, and Klaus Opwis. 2017. A Good Reason to Die: How Avatar Death and High Challenges Enable Positive Experiences. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, Denver Colorado USA, 5087–5097. https://doi.org/10.1145/3025453.3026047

[161] Cody Phillips, Daniel Johnson, Madison Klarkowski, Melanie Jade White, and Leanne Hides. 2018. The Impact of Rewards and Trait Reward Responsiveness on Player Motivation. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. ACM, Melbourne VIC Australia, 393–404. https://doi.org/10.1145/3242671.3242713

[162] PlatinumGames. 2017. *NieR: Automata*. Place: Osaka, Japan Published: Game [PlayStation 4, Windows, Xbox One, Nintendo Switch].

[163] Susanne Poeller and Cody J. Phillips. 2022. Self-Determination Theory — I Choose You!: The Limitations of Viewing Motivation in HCI Research Through the Lens of a Single Theory. In *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play*. ACM, Bremen Germany, 261–262. https://doi.org/10.1145/3505270.3558361

[164] Karl R. Popper. 2005. *Logik der Forschung* (11. aufl. durchges. u. erg ed.). Number 3 in Gesammelte Werke in deutscher Sprache / Karl R. Popper. Hrsg.: W. W. Bartley III. Mohr Siebeck, Tübingen.

[165] Andrew K. Przybylski, C. Scott Rigby, Edward L. Deci, and Richard M. Ryan. 2014. Competence-impeding electronic games and players' aggressive feelings, thoughts, and behaviors. *Journal of Personality and Social Psychology* 106, 3 (2014), 441–457. https://doi.org/10.1037/a0034820 Number: 3.

[166] Amon Rapp. 2017. From Games to Gamification: A Classification of Rewards in World of Warcraft for the Design of Gamified Systems. *Simulation & Gaming* 48, 3 (June 2017), 381–401. https://doi.org/10.1177/1046878117697147 Publisher: SAGE Publications Inc.

[167] Julian Rappaport. 1981. In praise of paradox: A social policy of empowerment over prevention. *American Journal of Community Psychology* 9, 1 (1981), 1–25. https://doi.org/10.1007/BF00896357 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1007/BF00896357.

[168] Julian Rappaport. 1984. Studies in Empowerment. *Prevention in Human Services* (1984). https://doi.org/10.1300/J293v03n02_02 Publisher: Taylor & Francis Group.

[169] Claire Reymond, Jan B. Vornhagen, Matthew Pelowski, Klaus Opwis, and Elisa D. Mekler. 2023. Images influencing images: How pictorial context affects the emotional interpretation of art photographs. *Psychology of Aesthetics, Creativity, and the Arts* (2023), No Pagination Specified–No Pagination Specified. https://doi.org/10.1037/aca0000523 Place: US Publisher: Educational Publishing Foundation.

[170] C. Scott Rigby. 2009. Finding The Right Rewards To Sustain Player Engagement: A Close-up Look At The RPG. https://gdcvault.com/play/1011921/Finding-The-Right-Rewards-To

[171] Scott Rigby. 2010. Tools of Engagement: Five Tested Techniques to Improve Player Retention. https://www.gdcvault.com/play/1013770/Tools-of-Engagement-Five-Tested

[172] Scott Rigby. 2011. *Glued to games : how video games draw us in and hold us spellbound*. Santa Barbara, Calif. : ABC-CLIO. http://archive.org/details/gluedtogameshowv0000rigb

[173] Scott Rigby and Troy Skinner. 2014. The Importance of Player Autonomy: Motivating Sustained Engagement through Volition and Choice. https://gdcvault.com/play/1020455/The-Importance-of-Player-Autonomy

[174] Raquel B. Robinson, Jan B. Vornhagen, Guo Freeman, Brendan Keogh, Regan L. Mandryk, and Brendan Sinclair. 2024. The Impact of Mass Game Industry Layoffs on the CHI PLAY Community. In *Companion Proceedings of the 2024 Annual Symposium on Computer-Human*

Interaction in Play (CHI PLAY Companion '24). Association for Computing Machinery, New York, NY, USA, 475–477. https://doi.org/10.1145/3665463.3678861

[175] Yvonne Rogers. 2004. New Theoretical Approaches for Human-Computer Interaction. *Annual Review of Information Science and Technology (ARIST)* 38 (2004), 87–143. https://www.learntechlib.org/p/97535/

[176] Yvonne Rogers. 2012. *HCI Theory: Classical, Modern, and Contemporary*. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-031-02197-8

[177] Felipe Romero. 2018. Who Should Do Replication Labor? *Advances in Methods and Practices in Psychological Science* 1, 4 (Dec. 2018), 516–537. https://doi.org/10.1177/2515245918803619 Publisher: SAGE Publications Inc.

[178] Felipe Romero. 2020. The Division of Replication Labor. *Philosophy of Science* 87, 5 (Dec. 2020), 1014–1025. https://doi.org/10.1086/710625 Publisher: Cambridge University Press.

[179] Iris van Rooij and Mark Blokpoel. 2020. Formalizing Verbal Theories. *Social Psychology* (Oct. 2020). https://econtent.hogrefe.com/doi/10.1027/1864-9335/a000428 Publisher: Hogrefe Publishing.

[180] Mark Rosenman. 1980. Empowerment as a purpose of education. *Alternative Higher Education* 4, 4 (June 1980), 248–259. https://doi.org/10.1007/BF01079732 Number: 4.

[181] Robert Rosenthal. 1979. The file drawer problem and tolerance for null results. *Psychological Bulletin* 86, 3 (1979), 638–641. https://doi.org/10.1037/0033-2909.86.3.638 Number: 3.

[182] Tony Ross-Hellauer. 2017. What is open peer review? A systematic review. *F1000Research* 6 (Aug. 2017). https://doi.org/10.12688/f1000research.11369.2

[183] Richard M Ryan and Edward L Deci. 2017. Cognitive Evaluation Theory, Part I The Effects of rewards, feedback, and other external events on intrinsic motivation. In *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness.* Guilford Press, New York, USA, 123–157.

[184] Richard M. Ryan and Edward L. Deci. 2017. *Self-determination theory: basic psychological needs in motivation, development, and wellness*. Guilford Press, New York.

[185] Richard M. Ryan, C. Scott Rigby, and Andrew Przybylski. 2006. The Motivational Pull of Video Games: A Self-Determination Theory Approach. *Motivation and Emotion* 30, 4 (Dec. 2006), 344–360. https://doi.org/10.1007/s11031-006-9051-8 Number: 4.

[186] Anne M. Scheel. 2022. Why most psychological research findings are not even wrong. *Infant and Child Development* n/a, n/a (Jan. 2022), e2295. https://doi.org/10.1002/icd.2295 Number: n/a _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/icd.2295.

[187] Anne M. Scheel, Leonid Tiokhin, Peder M. Isager, and Daniël Lakens. 2021. Why Hypothesis Testers Should Spend Less Time Testing Hypotheses. *Perspectives on Psychological Science* 16, 4 (July 2021), 744–755. https://doi.org/10.1177/1745691620966795 Number: 4.

[188] Anne M. Scheel, Tim van der Zee, James A. Grange, Chris Chambers, Aaron R. Caldwell, Hannah Fraser, Timo B. Roettger, Xenia Schmalz, Alexander C. DeHaven, and Claire Riss. 2017. Registered Reports Now! (May 2017). https://osf.io/3wct2/ Publisher: OSF.

[189] Jesse Schell. 2019. *The Art of Game Design*. A K Peters/CRC Press, Third edition. | Boca Raton : Taylor & Francis, a CRC title, part of the Taylor & Francis imprint, a member of the Taylor & Francis Group, the academic division of T&F Informa, plc, 2019. https://doi.org/10.1201/b22101

[190] Klaus R. Scherer. 2001. Appraisal considered as a process of multilevel sequential checking. In *Appraisal processes in emotion: Theory, methods, research*. Oxford University Press, New York, NY, US, 92–120.

[191] Ines Schindler, Georg Hosoya, Winfried Menninghaus, Ursula Beermann, Valentin Wagner, Michael Eid, and Klaus R. Scherer. 2017. Measuring aesthetic emotions: A review of the literature and a new assessment tool. *PLOS ONE* 12, 6 (June 2017), e0178899–e0178899. https://doi.org/10.1371/journal.pone.0178899 Number: 6.

[192] Stefan Schmidt. 2009. Shall we Really do it Again? The Powerful Concept of Replication is Neglected in the Social Sciences. *Review of General Psychology* 13, 2 (June 2009), 90–100. https://doi.org/10.1037/a0015108 Publisher: SAGE Publications Inc.

[193] Hanna Schneider, Malin Eiband, Daniel Ullrich, and Andreas Butz. 2018. Empowerment in HCI - A Survey and Framework. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–14. https://doi.org/10.1145/3173574.3173818

[194] Felix D. Schönbrodt. 2016. p-Hacker: Train your p-hacking skills! http://shinyapps.org/apps/p-hacker/

[195] Larissa Shamseer, David Moher, Mike Clarke, Davina Ghersi, Alessandro Liberati, Mark Petticrew, Paul Shekelle, Lesley A. Stewart, Douglas G. Altman, Alison Booth, An Wen Chan, Stephanie Chang, Tammy Clifford, Kay Dickersin, Matthias Egger, Peter C. Gøtzsche, Jeremy M. Grimshaw, Trish Groves, Mark Helfand, Julian Higgins, Toby Lasserson, Joseph Lau, Kathleen Lohr, Jessie McGowan, Cynthia Mulrow, Melissa Norton, Matthew Page, Margaret Sampson, Holger Schünemann, Iveta Simera, William Summerskill, Jennifer Tetzlaff, Thomas A. Trikalinos, David Tovey, Lucy Turner, and Evelyn Whitlock. 2015. Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015: Elaboration and explanation. *BMJ (Online)* 349, December 2014 (2015), 1–25. https://doi.org/10.1136/bmj.g7647 Number: December 2014.

[196] John Sharp, Colleen Macklin, Steven Davis, Yu Jen Chen, Tuba Ozkan, and Carla Molins Pitarch. 2019. Failure. In *Iterate: Ten Lessons in Design and Failure*. MIT Press, 23–42. https://ieeexplore.ieee.org/document/8709377 Conference Name: Iterate: Ten Lessons in Design and Failure.

[197] Kennon M. Sheldon and Jonathan C. Hilpert. 2012. The balanced measure of psychological needs (BMPN) scale: An alternative domain general measure of need satisfaction. *Motivation and Emotion* 36, 4 (Dec. 2012), 439–451. https://doi.org/10.1007/s11031-012-9279-4 Number: 4.

[198] Paul J. Silvia. 2009. Looking Past Pleasure: Anger, Confusion, Disgust, Pride, Surprise, and Other Unusual Aesthetic Emotions. *Psychology of Aesthetics, Creativity, and the Arts* 3, 1 (2009), 48–51. https://doi.org/10.1037/a0014632 Number: 1.

[199] Paul J. Silvia. 2010. Confusion and interest: The role of knowledge emotions in aesthetic experience. *Psychology of Aesthetics, Creativity, and the Arts* 4, 2 (2010), 75–80. https://doi.org/10.1037/a0017081 Number: 2.

[200] Joseph P. Simmons, Leif D. Nelson, and Uri Simonsohn. 2011. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science* 22, 11 (2011), 1359–1366. https://doi.org/10.1177/0956797611417632 Number: 11.

[201] Barbara Bryant Solomon. 1976. *Black empowerment : social work in oppressed communities*. New York : Columbia University Press. http://archive.org/details/blackempowerment00barb_0

[202] Friedrich Steinle. 1997. Entering New Fields: Exploratory Uses of Experimentation. *Philosophy of Science* 64, S4 (Jan. 1997), S65–S74. https://doi.org/10.1086/392587

[203] Alistair Sutcliffe. 2000. On the effective use and reuse of HCI knowledge. *ACM Trans. Comput.-Hum. Interact.* 7, 2 (June 2000), 197–221. https://doi.org/10.1145/353485.353488

[204] Penelope Sweetser and Peta Wyeth. 2005. GameFlow. *Computers in Entertainment* 3, 3 (July 2005), 3–3. https://doi.org/10.1145/1077246.1077253 Number: 3.

[205] Denes Szucs and John P.A. Ioannidis. 2017. When null hypothesis significance testing is unsuitable for research: A reassessment. *Frontiers in Human Neuroscience* 11 (Aug. 2017). https://doi.org/10.3389/fnhum.2017.00390 Publisher: Frontiers Media S. A.

[206] Anaïs Thibault Landry, Konstantinos Papachristopoulos, Marc-Antoine Gradito Dubord, and Jacques Forest. 2022. "Here's Some Money, Your Work's So Worthy?" A Brief Report on the Validation of the Functional Meaning of Cash Rewards Scale. *Frontiers in Psychology* 13 (2022). https://www.frontiersin.org/articles/10.3389/fpsyg.2022.821501

[207] Noam Tractinsky. 2018. The Usability Construct: A Dead End? *Human–Computer Interaction* 33, 2 (March 2018), 131–177. https://doi.org/10.1080/07370024.2017.1298038

[208] April Tyack and Elisa D. Mekler. 2020. Self-Determination Theory in HCI Games Research: Current Uses and Open Questions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–22. https://doi.org/10.1145/3313831.3376723

[209] April Tyack and Elisa D. Mekler. 2024. Self-Determination Theory and HCI Games Research: Unfulfilled Promises and Unquestioned Paradigms. *ACM Transactions on Computer-Human Interaction* (June 2024). https://doi.org/10.1145/3673230 Just Accepted.

[210] April Tyack, Peta Wyeth, and Madison Klarkowski. 2018. Video Game Selection Procedures For Experimental Research. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, New York, New York, USA, 1–9. https://doi.org/10.1145/3173574.3173760

[211] Niels van Berkel and Kasper Hornbæk. 2023. Implications of Human-Computer Interaction Research. *interactions* 30, 4 (June 2023), 50–55. https://doi.org/10.1145/3600103

[212] Tom van Drimmelen, M. Nienke Slagboom, Ria Reis, Lex M. Bouter, and Jenny T. van der Steen. 2024. Decisions, Decisions, Decisions: An Ethnographic Study of Researcher Discretion in Practice. *Science and Engineering Ethics* 30, 6 (Nov. 2024), 59. https://doi.org/10.1007/s11948-024-00481-5

[213] Rob Van Roy and Bieke Zaman. 2019. Unravelling the ambivalent motivational power of gamification: A basic psychological needs perspective. *International Journal of Human-Computer Studies* 127 (July 2019), 38–50. https://doi.org/10.1016/j.ijhcs.2018.04.009

[214] Jan B. Vornhagen, Dan Bennett, Dooley Murphy, and Elisa D. Mekler. 2023. "I'm the leader and I'm going to save the world": Characterizing Empowering and Disempowering Game Experiences. *Proceedings of the ACM on Human-Computer Interaction* 7, CHI PLAY (Sept. 2023), 1330–1360. https://doi.org/10.1145/3611071

[215] Jan B. Vornhagen, April Tyack, and Elisa D. Mekler. 2020. Statistical Significance Testing at CHI PLAY: Challenges and Opportunities for More Transparency. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '20)*. Association for Computing Machinery, New York, NY, USA, 4–18. https://doi.org/10.1145/3410404.3414229

[216] Chat Wacharamanotham, Lukas Eisenring, Steve Haroz, and Florian Echtler. 2020. Transparency of CHI Research Artifacts : Results of a Self-Reported Survey. (2020). https://doi.org/10.31219/osf.io/3bu6t

[217] Chat Wacharamanotham, Fumeng Yang, Xiaoying Pu, and Abhraneel Sarma. 2023. Transparent Practices for Quantitative Empirical Research. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, 1–5. https://doi.org/10.1145/3544549.3574168

[218] Valentin Wagner, Winfried Menninghaus, Julian Hanich, and Thomas Jacobsen. 2014. Art schema effects on affective experience: The case of disgusting images. *Psychology of Aesthetics, Creativity, and the Arts* 8, 2 (2014), 120–129. https://doi.org/10.1037/a0036126 Number: 2.

[219] Hao Wang and Chuen-Tsai Sun. 2011. Game Reward Systems: Gaming Experiences and Social Meanings. *Proceedings of DiGRA 2011 Conference: Think Design Play* (2011).

[220] C. Kenneth Waters. 2007. The Nature and Context of Exploratory Experimentation: An Introduction to Three Case Studies of Exploratory Research. *History and Philosophy of the Life Sciences* 29, 3 (2007), 275–284. https://www.jstor.org/stable/23334262 Publisher: Stazione Zoologica Anton Dohrn - Napoli.

[221] Matthew Alexander Whitby, Sebastian Deterding, and Ioanna Iacovides. 2019. "One of the baddies all along": Moments that Challenge a Player's Perspective. In *Proceedings of the*

Annual Symposium on Computer-Human Interaction in Play. ACM, Barcelona Spain, 339–350. https://doi.org/10.1145/3311350.3347192

[222] Jelte M. Wicherts, Coosje L.S. Veldkamp, Hilde E.M. Augusteijn, Marjan Bakker, Robbie C.M. van Aert, and Marcel A.L.M. van Assen. 2016. Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid P-hacking. *Frontiers in Psychology* 7, NOV (2016), 1–12. https://doi.org/10.3389/fpsyg.2016.01832 Number: NOV.

[223] Josef Wiemeyer, Lennart Nacke, Christiane Moser, and Florian 'Floyd' Mueller. 2016. Player Experience. In *Serious Games*, Ralf Dörner, Stefan Göbel, Wolfgang Effelsberg, and Josef Wiemeyer (Eds.). Springer International Publishing, Cham, 243–271. https://doi.org/10.1007/978-3-319-40612-1_9

[224] Robyn Wilford, Juan Ardila-Cifuentes, Edward Baggs, and Michael L. Anderson. 2022. The stimulus-response crisis. *The Behavioral and Brain Sciences* 45 (Feb. 2022), e39. https://doi.org/10.1017/S0140525X21000285

[225] Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J. G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A. C. 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3, 1 (March 2016), 160018. https://doi.org/10.1038/sdata.2016.18 Bandiera_abtest: a Cg_type: Nature Research Journals Number: 1 Primary_atype: Comments & Opinion Publisher: Nature Publishing Group Subject_term: Publication characteristics;Research data Subject_term_id: publication-characteristics;research-data.

[226] Douglas Wilson and Miguel Sicart. 2010. Now it's personal: on abusive game design. In *Proceedings of the International Academic Conference on the Future of Game Design and Technology*. ACM, Vancouver British Columbia Canada, 40–47. https://doi.org/10.1145/1920778.1920785

[227] Max Wilson, Wendy Mackay, Ed Chi, Michael Bernstein, and Jeffrey Nichols. 2012. RepliCHI SIG: from a panel to a new submission venue for replication. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems (CHI EA '12)*. ACM, New York, NY, USA, 1185–1188. https://doi.org/10.1145/2212776.2212419

[228] Max L. Wilson, Ed H. Chi, Stuart Reeves, and David Coyle. 2014. RepliCHI: the workshop II. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems (CHI EA '14)*. ACM, New York, NY, USA, 33–36. https://doi.org/10.1145/2559206.2559233

[229] Max L. Wilson, Wendy Mackay, Ed Chi, Michael Bernstein, Dan Russell, and Harold Thimbleby. 2011. RepliCHI - CHI should be replicating and validating results more: discuss. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems (CHI EA '11)*. ACM, New York, NY, USA, 463–466. https://doi.org/10.1145/1979742.1979491

[230] Max L. L. Wilson, Paul Resnick, David Coyle, and Ed H. Chi. 2013. RepliCHI: the workshop. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13)*. ACM, New York, NY, USA, 3159–3162. https://doi.org/10.1145/2468356.2479636

[231] Jacob O. Wobbrock and Julie A. Kientz. 2016. Research contributions in human-computer interaction. *Interactions* 23, 3 (April 2016), 38–44. https://doi.org/10.1145/2907069 Number: 3.

[232] Yager Development. 2012. *Spec Ops: The Line*. Place: Berlin, Germany Published: Game [Windows, PlayStation 3, Xbox 360, macOS, Linux].

[233] Tal Yarkoni. 2019. The Generalizability Crisis. (2019). https://doi.org/10.31234/osf.io/jqw35

[234] Johannes Zagermann. 2024. CHI 2025 — Papers Track, post-review report (Round 1). https://chi2025.acm.org/

[235] Marc A. Zimmerman. 1990. Taking aim on empowerment research: On the distinction between individual and psychological conceptions. *American Journal of Community Psychology* 18, 1 (1990), 169–177. https://doi.org/10.1007/BF00922695 Number: 1.

[236] Marc A. Zimmerman. 1995. Psychological empowerment: Issues and illustrations. *American Journal of Community Psychology* 23, 5 (Oct. 1995), 581–599. https://doi.org/10.1007/BF02506983

[237] Marc A. Zimmerman and Julian Rappaport. 1988. Citizen participation, perceived control, and psychological empowerment. *American Journal of Community*

Psychology 16, 5 (1988), 725–750. https://doi.org/10.1007/BF00930023 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1007/BF00930023.

[238] Rolf A. Zwaan, Alexander Etz, Richard E. Lucas, and M. Brent Donnellan. 2018. Making replication mainstream. *Behavioral and Brain Sciences* 41 (Jan. 2018), e120. https://doi.org/10.1017/S0140525X17001972

Part II

Publications

Paper I

Statistical Significance Testing at CHI PLAY: Challenges and Opportunities for More Transparency

Jan B. Vornhagen, April Tyack, & Elisa D. Mekler

The paper has been published in the
Proceedings of the Annual Symposium on Computer-Human Interaction in Play Vol. 6, pp. 4–18, 2020.
https://doi.org/10.1145/3410404.3414229

The layout has been revised.

Abstract

Statistical Significance Testing – or Null Hypothesis Significance Testing (NHST) – is common to quantitative CHI PLAY research. Drawing from recent work in HCI and psychology promoting transparent statistics and the reduction of questionable research practices, we systematically review the reporting quality of 119 CHI PLAY papers using NHST (data and analysis plan at OSF.io). We find that over half of these papers employ NHST without specific statistical hypotheses or research questions, which may risk the proliferation of false positive findings. Moreover, we observe inconsistencies in the reporting of sample sizes and statistical tests. These issues reflect fundamental incompatibilities between NHST and the frequently exploratory work common to CHI PLAY. We discuss the complementary roles of exploratory and confirmatory research, and provide a template for more transparent research and reporting practices.

I.1 Introduction

A primary goal of CHI PLAY is to provide a space for "*high quality research in games and HCI*" while "*embracing a wide variety of research contributions*" [3]. Many of these contributions emerge from empirical user studies of videogames and other game-like artefacts, whereby statistical analysis is applied to quantitative (or quantified) data to produce new insights regarding player-computer interaction [179]. Often, data analysis proceeds by way of p values (e.g., as computed via t-test or ANOVA), which are used to understand whether trends in data represent real effects, or merely noise. This is commonly called *Null Hypothesis Significance Testing* (NHST).

However, NHST methods have become increasingly subject to critique. False positive results, whereby noise is misidentified as a real effect, can easily occur as a result of common practices performed during analysis [79, 154]. These *Questionable Research Practices* [177, QRPs] threaten the legitimacy of statistical significance and therefore complicate interpretation of published research findings [79, 154]. QRPs are facilitated by a publishing climate biased towards statistically significant results[1], leaving non-significant research findings in the file drawer [33, 49, 131, 170].

A growing number of HCI scholars have consequently called for greater consideration of the quality of NHST analyses, and statistical reporting more broadly [26, 75, 48, 27, 88]. However, the extent to which these issues affect HCI research on games and play – and CHI PLAY in particular – is yet to be determined.

Yet CHI PLAY arguably has much to gain from other fields where similar problems have begun to be addressed. Quantitative games research in HCI often draws from psychological

[1] A phenomenon we have also occasionally observed during peer review at CHI PLAY.

theory and methodology [e.g., 165] – as such, we argue that recent psychological work on Open Science has much to offer to the CHI PLAY community.

In this paper, we examine the quality of statistical reporting at CHI PLAY, reviewing 119 publications employing NHST in their analysis. We observe wide variation in reporting quality: about two thirds (67.2%) of these papers consistently report full test statistics, 28.6% contain inconsistent p values, and only seven papers (5.9%) justify their sample size. Moreover, NHST is often (mis)applied to exploratory research questions, risking the propagation of false positive findings. Our results demonstrate that quantitative research at CHI PLAY exhibits similar issues as other HCI research domains [26], suggesting a need to improve research practice, peer review, and publication guidelines. To help address these issues, we offer a comprehensive, easy to use template for authors and reviewers to assess the reporting quality of papers that employ NHST in their analysis. Moreover, we argue for the widespread adoption of Open Science practices, such as data sharing and pre-registration, by the CHI PLAY community.

In the following, we first describe the NHST method, and summarize pertinent issues with the approach identified by scholars in HCI [e.g., 26, 48, 33] and other fields [e.g., 79, 154, 158, 177]. We then present a systematic literature review of 119 papers employing NHST to evaluate the quality of statistical reporting at the venue. Lastly, we present a template for authors and reviewers, and propose recommendations for future quantitative games and play research at CHI PLAY.

I.2 Related Work

We begin with a brief summary of NHST methods, from which we situate our critique. A comprehensive overview is beyond the scope of this paper; however, for a more in-depth review of the underlying mathematical principles, we refer readers to [45, 48, 102].

Quantitative research usually takes one of two forms: *Exploratory* work, through which hypotheses are generated, or *confirmatory* studies, where hypotheses can be tested [48, 75]. Exploratory research is typically conducted in the early stages of forming theories and conceptual frameworks, where researchers are unsure what effects to expect. Data are collected and explored for potential effects, informing theory and concept development, which facilitate the generation of concrete hypotheses. The goal of confirmatory research, then, is to test these hypotheses through their use in predicting one or more study outcomes. In this way, correct predictions can represent support for a theory.

Researchers often test hypotheses using Null Hypothesis Significance Testing (NHST). NHST is an inferential approach to statistics that allows researchers to decide between two competing

hypotheses about data: The *null hypothesis* (H_0) and the alternative hypothesis (H_1).

Researchers are usually interested in an entire population (e.g., online game players); however, it is usually impossible to test every person in a population of interest. Methods such as NHST can produce valid inferences regarding the complete population from a smaller sample.

However, extrapolating from incomplete data inherently comes with the risk of making one of two errors: falsely rejecting the null hypothesis when it is actually true (Type I), and failing to reject the null hypothesis when it is actually false (Type II) [88].

It is not possible to find support for a specific hypothesis with NHST, as p values represent the probability of observing at least equally extreme data under the null hypothesis. NHST instead allows for precise control over how often a hypothesis is falsely accepted or rejected *in the long run*. The significance (or α) threshold (usually .05) controls how often we will make a Type I error and incorrectly reject the null hypothesis (false positive). Statistical power ($1 - \beta$, usually set to .8) controls how often we will make a Type II error and incorrectly fail to reject the null hypothesis (false negative).

Using NHST, it is not possible to know whether a null hypothesis is truly false following a single significant result. Instead, NHST demonstrates the frequency of incorrect judgements in the long run when rejecting the null hypothesis. Because error rates are controlled over a series of hypothetically infinite studies, rather than just one, NHST is also referred to as a frequentist approach to statistics [45].

I.2.1 Criticism of NHST

While widespread, the use of NHST in HCI is controversial [48, 88, 33]. Two main issues emerge: First, a single p value constitutes only weak evidence; second, long-term error control can occur only when researchers abide by a number of specific rules. Error rates become inflated when these rules are not followed, which limits the validity of results. We discuss each issue in turn.

It is a common misunderstanding that p values represent the probability that a hypothesis is true [158, 65, 48, 88, 45]. More worryingly, p values are not a measure of strength of evidence like a standardized effect size, as they randomly fluctuate [88, 158] in what was coined the "dance of the p values" [49, 34]. For example, when performing t-tests on two groups with the same mean and standard deviation – in other words, when no true effect exists – all p values are equally likely; the proportion of significant p values is equal to the α threshold (i.e., 5%). Even when a true between-group difference exists, repeated tests will not produce the same p value: The p values "dance" around the significance threshold. This is illustrated in Figure I.1.

This issue is compounded for studies with low *statistical* power. In Figure I.1 left, power $\approx$ 80%; i.e., approximately 80% of p values correctly fall below the significance threshold, and

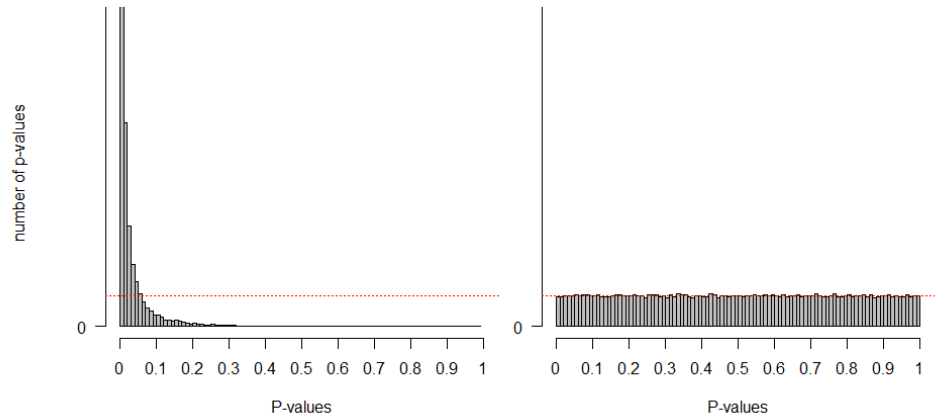

Fig. I.1: Histograms of p values, each produced from 100,000 simulated t-tests. For each t-test (n=50), samples were randomly drawn from two normal distributions. On the left, the two distributions had $m_1 = 100$ and $m_2 = 105$ ($sd_{1,2} = 10$); on the right, the two distributions were completely equal ($m_{1,2} = 100$; $sd_{1,2} = 10$). In both histograms, the first five bars include all p values < 0.05, constituting significant results. Note that $\sim$5% of results in the right distribution were significant. The red bar represents the uniform distribution of p values for infinite simulations where the true effect is zero. Plots generated with code by [102] (CC BY-NC-SA 4.0).

"only" 20% of studies would commit a Type II error. The statistical power of a test is a function of both sample and effect size: power is maximized when a large effect is expected from a large sample. For under-powered studies, Type II errors become more likely, and null results consequently provide limited information[2] [88].

This behavior of the p value when the null is true – the uniform distribution in which every p value is equally likely – is problematic for exploratory study designs. More specifically, when more values are drawn from the uniform p distribution (i.e., as more statistical tests are run), the odds of observing spurious significant results are increased.

Exploratory research takes an essential role in HCI [33], as exploratory works where novel artefacts are designed and evaluated – often via a variety of measures and statistical tests – constitute key contributions to the field [179]. Yet applying NHST in exploratory research presents one of two drawbacks: if the critical α threshold is lowered to prevent an excess of false positive findings, potentially interesting (and non-significant) findings may be overlooked; conversely, not adjusting α substantially increases the Type I error rate such that confidence in test results is attenuated [35, 56]. Accordingly, exploratory work should be explicitly labelled as such, and indeed may benefit from using alternative analytic methods, such as estimation [48].

I.2.2 Questionable Research Practices and Researcher Degrees of Freedom

Sometimes researchers engage in exploratory analyses, yet report the resulting significant findings as if they were confirmatory – effects are claimed to support a hypothesis formulated after data collection has occurred. This practice is called hypothesizing after results are known, or HARKing [91, 33] and is a so-called *Questionable Research Practice* (QRP) [83]

QRPs are decisions made during data collection or analysis whose strategic application can improve the odds of achieving statistically significant results (and thereby inflate the Type I error rate). Typical QRPs include removing data points to change a group mean, collecting data until desired results occur, selectively reporting (in)dependent variables, rounding down p values close to the significance threshold, and HARKing [142, 83, 177].

These practices may, at first glance, seem reasonable: for example, continuing data collection when analysis yields non-significant results [83] does increase the sample size, which is generally desirable. However, because p values behave randomly (Figure Figure I.1), these further tests always increase the chance of false positive findings. In fact, this practice of intermediate testing

[2]It should be noted that we discuss *statistical* power, which is calculated with an expected effect size. So-called *post hoc* power, derived after analysis from the observed effect size and achieved sample size is merely a conversion of the p value and therefore of limited use and misleading [101].

until a significant result is found will eventually yield significant results regardless of what is tested [158].

It has been shown that QRPs are pervasive in the psychological literature [83]. Their impact can be immense: Only reporting one of two dependent variables, collecting ten more observations per cell, controlling for gender, and selectively choosing between three conditions collectively increases the rate of false positive findings from 5% to 60% [154][3]. The influence of QRPs has been deemed so massive that in a widely discussed paper, Ioannidis [79] proposed that most published research findings are false, with some scientific fields potentially only reporting their own biases instead of any true effects.

A number of approaches to prevent QRPs have been identified [33]. One prominent option is pre-registration, whereby scholars record all salient features of their research plan in a time-stamped, immutable form prior to data collection [33, 122]. While the exact contents of pre-registrations tend to differ [see 122], they typically include the research goals, hypotheses to be tested, and the statistical analysis plan [33]. Pre-registration plans represent compelling evidence that confirmatory statistical analyses are independent of the observed data. As such, pre-registration practices represent a positive step towards more open and transparent science; however, they do not constitute an absolute safeguard against QRPs. Pre-registered studies lose value when research plans omit key information, or differ from final analyses without a valid rationale [177].

QRPs are made possible by the substantial amount of leeway afforded during research, otherwise known as Researcher Degrees of Freedom (R-DFs) [154, 177]. These freedoms are not bad in and of themselves: many decisions, such as which participants to exclude from a sample, are essential aspects of research (for example, to ensure data integrity [23]). However, because researchers are incentivized to produce "novel" (i.e., significant) results for publication [33], R-DFs can be readily misused to "improve" results [83]. Conversely, many non-significant results are never published for the same reason, resulting in the *file drawer problem* [33, 137].

We note that QRPs do not constitute fraud. Practices such as intermediate testing have historically been considered unproblematic and defensible [83]. In contrast, fraud requires intent, and is therefore rare. QRPs are widespread, easily performed unintentionally, and difficult to detect, making them a primary threat to the reproducibility of results [154, 142, 177].

R-DFs serve a valid purpose in the scientific process – for example, excluding careless survey respondents can improve data quality [23]. However, their application should be intentional, consistent, and transparently reported. To these ends, Wicherts et al. [177] compiled a list of 34 R-DFs to help researchers identify unintentional p-hacking in their own practices; for example, testing broadly stated hypotheses for which a number of dependent variables *could* apply – or, if

[3] An interactive example can be found in the P-Hacker app, which illustrates this process [144]

non-significant, be removed from the analysis.

In the present work, we apply this list of 34 R-DFs to evaluate the quality of statistical reporting at CHI PLAY. In doing so, we aimed to understand how R-DFs at CHI PLAY are reported, and hence to what extent the rate of false positive results has been inflated. In this way, the present research follows from a long tradition of meta-scientific work on research methods at CHI PLAY, CHI, and the wider HCI literature.

Calls for other changes to research methods and reporting practice have also been made in recent games scholarship. The use of non-violent "control" videogames in aggression research [8], for example, provoked questions as to whether games that vary across genre, pacing, and content can induce comparative experiences [51, 73]. More detailed reporting of game selection procedures, with theoretical or empirical bases, has been suggested as a means by which researchers could more convincingly justify their choice of stimulus games [166]. Psychometric measures have also come under increasing scrutiny in HCI games research. In particular, it was shown that the Game Experience Questionnaire (GEQ) saw wide use as a validated measure, despite the absence of a published validation study [109] – and indeed, the stated factor structure could not be independently validated [24, 85, 109]. Moreover, substantial variation in reporting basic qualities of the GEQ, such as the number of scale items, was observed across the literature [109], suggesting the existence of broader issues in reporting practice in HCI games research.

I.3 Systematic Literature Review

We conducted a systematic literature review to take stock of the quality of statistical reporting at CHI PLAY. Following the PRISMA-P protocol [151], the literature review was pre-registered on January 29 2020, before beginning data collection. Materials and the PRISMA Flow diagram detailing all steps of the review are available at https://osf.io/4mcbn/. The literature review and analysis were performed by the first author, in regular consultation with the third author.

Identification

Using the ACM Digital Library[4], we collected all CHI PLAY papers published since the inaugural conference in 2014 that were classified as "Research-Article" (i.e., full papers). As such, we did not include publications labeled as "Abstract" or "Short-Paper", as the reduced page limit and potential preliminary status of the work (e.g., Works-in-Progress) may have limited what authors could report. This first step resulted in a sample of n=246 papers.

[4] dl.acm.org/conference/chi-play/proceedings

Screening

Next, we screened the sample of 246 papers for Notes. We decided to exclude Notes from our sample, for the same reason that we excluded Abstracts and Short-Papers. We excluded n=8 notes, resulting in n=238 considered for further analysis.

Eligibility

We screened the remaining papers for the presence of inferential statistics; for example, reporting p values or describing results as significant. In this way, a further n=108 papers without inferential statistics were excluded, leaving a sample of n=130 papers (marked with * in the References).

Codebook

Initial coding proceeded by adapting the checklist developed by Wicherts et al. [177] into a preliminary codebook. These early categories included R-DFs organized around Hypothesizing (e.g., "Conducting explorative [sic] research without any hypothesis"), Study Design (e.g., "Measuring additional constructs that could potentially act as primary outcomes"), Data Collection (e.g., "Determining the data collection stopping rule on the basis of desired results or intermediate significance testing"), Statistical Analysis (e.g., "Choosing to include different measured variables as covariates, independent variables, mediators, or moderators"), and Reporting of Results (e.g., "Failing to assure reproducibility (verifying the data collection and data analysis)") [all quotes 177, Table 1, p. 3].

As per our pre-registration report, we randomly selected a subset of papers (n=32; ∼25% of the sample) to assess the preliminary codebook's viability. Based on this analysis, the codebook was revised in several ways:

- As only two studies in our sample were pre-registered (i.e., [81, 175]), R-DFs that could only be identified with knowledge of authors' pre-study intentions were removed (e.g., post-hoc switching of the primary outcome).
- As we will elaborate in our Results section, hypotheses, test statistics, effect sizes, measures, and assumption tests were often reported incorrectly or not at all (e.g., hypotheses were only implicitly linked to tests performed). We therefore added codes referring to complete and clear reporting practices (e.g., "are the (in)dependent variables reported in a way that readers could reproduce them?").
- We added items for statistical reporting (e.g., "are full test statistics reported?") to address concerns previously noted in Related Work regarding the improper use of NHST.

We also examined p value reporting practices with statcheck.io [121], or manual computation where necessary. In particular, we investigated whether reported p values were consistent with their corresponding degrees of freedom and test statistics.

- We added items related to reproducibility (e.g., "does the paper provide raw data or an analysis plan?") to assess to what extent Open Science practices, such as sharing data, have been adopted at CHI PLAY.

The final codebook, which includes code descriptions and relevant citations, can be found in the OSF repository.

Included

We analyzed all 130 papers with the updated codebook. During this final analysis, a further n=11 papers were excluded: Four papers did not feature statistical inferences, and were therefore deemed false positives [15, 31, 47, 74]; two studies used exploratory factor analyses [109, 161], which (unlike NHST) are designed for exploratory analyses; finally, five papers did not employ NHST when reporting results [125, 171, 173], or only reported non-significant results without further information on the statistical analysis [119, 175]. This resulted in a final sample of N=119 papers, which forms the basis of our systematic review results.

I.4 Results

In the following section, we present the results of our literature review. Altogether, our sample spans almost half (48.78%) of all CHI PLAY full papers published between 2014 and 2019, attesting to the popularity of NHST at the conference. We present findings in the general order of our codebook. Importantly, our review concentrates on the quality of the (described) *methods and reporting*; we do not intend to make statements on the research quality of the works.

I.4.1 Hypothesis Reporting

As noted in Related Work, long term error rates are only controlled in confirmatory research designs that test specific statistical hypotheses. We therefore examined whether papers employing NHST reported confirmatory research goals and statistical hypotheses.

Most works contained at least one exploratory research question (n=75, 63.02%), though it was not always labeled as such. For some studies, however, an exploratory focus was explicitly noted; for instance, "in the absence of an existing theoretical framework for the design and

discussion of asymmetric games, we adopted an exploratory approach" [70, p. 350]. Other papers were less clear: in some cases, no research questions were identified [e.g., 13]; in others, research goals were described in ways that could not be interpreted as confirmatory [e.g., 11] – for example, among one study's "primary aims" [14, p. 327] was to "[d]etermine the similarities and differences of the likely impacts of MDDA [...]" (p. 327). Note that while this statement is a useful research question and may be considered a "theoretical" hypothesis, it does not constitute a "statistical" hypothesis or confirmatory research goal, as it does not directly relate to test outcomes.

Of the 119 papers, n=44 (36.98%) stated a confirmatory research goal, outlining at least one hypothesis that was supported or rejected on the basis of their results. For example, two hypotheses are defined in [41], one of which predicts that "[e]xperiences of interdependence (H1a) and cooperation (H1b) are positively associated with in-game social capital" (p. 90). As both H1a and H1b are later defined in terms of self-report measures, this hypothesis can be directly tested.

Notably, over half of the reviewed works (n=64, 53.78%) stated hypotheses (e.g., "There will be no differences in any game experience measures due to graphical fidelity", [21, p. 270]), even where confirmatory research goals had not been formulated. Only n=22 (18.49%) outlined statistical hypotheses that could be directly translated into statistical tests (e.g., "The perspective switching provides significant benefits to spatial orientation and overview" [32, p. 290]). This is mirrored in the reported analyses, where for most papers (n=90, 75.63%), the distinction between confirmatory and exploratory testing is unclear. In comparison, 12.61% (n=15) of papers clarified this distinction – for instance, by using subheadings (e.g., "Exploratory Analysis: Game Experience and Game Behavior", [124, p. 9]; "Further Statistics", [185, p. 7]). In the remaining 11.77% papers (n=14), only confirmatory analyses were conducted.

We were unable to confidently assess whether tests for all hypotheses were reported in 50.42% (n=60) of papers, either because the hypotheses themselves were not clearly stated [e.g., 37], or the reporting did not allow us to assess whether a hypothesis had been answered: For example, one paper sought to study how a system might "collect en-masse achievement data about gamers" [p. 305 174], "[w]hat insights [might] be drawn solely from the data collected [...]" (p. 305), and identify the limitations of "using only one source of usernames for the system [...]" (p. 305). While the first two questions were addressed in the results section, the open wording of the hypothesis makes it impossible to determine whether the questions were addressed completely. Indeed, the third research question is not discussed in the results section at all, but only addressed in the limitations.

In the remaining 49.58% (n=59) of cases, all hypotheses could be clearly linked to a corresponding test [e.g., 72, 157].

I.4.2 Study Design Reporting

Researchers have the most freedom when planning and preparing a study. Detailed and thorough reporting of the study design is therefore crucial for readers to understand how a study was conducted, the ways its enactment may have influenced results, and how it could be replicated.

Sample Size

An important part of NHST is justifying the sample size, usually via power analysis and a defined significance threshold (e.g., $\alpha < .05$). Of the N=119 papers, n=7 (5.88%) justified their sample size [e.g., with a power analysis, as in 44], with one of the seven only reporting post-hoc power [21] – which is somewhat misleading, as post-hoc power is not equivalent to statistical power, but rather a conversion of the p-Value [101].

All papers reported the sample size, where n=54 (45.38%) either did not remove any participants from their analysis [e.g., 9], or provided a rationale for doing so [e.g., where data collection was compromised for participants who had guessed the true intent of the study, as in 168]. A further 21.01% (n=25) of papers did not justify removing participants, or did not mention having done so, despite inconsistencies between reported degrees of freedom and sample size [i.e., indicating that participants were removed from the test, e.g., 6, 28]. In the remaining 33.61% (n=40) papers, insufficient details were reported to determine whether participants were removed.

Significance Threshold

The significance threshold (i.e., α) was defined more frequently, with 18.49% (n=22) of papers opting for a single value (i.e., 0.05) for all tests. All other papers either used multiple thresholds or an implicit threshold. Further, only 4 papers (3.36%) justified their α threshold. For example, [164] adjusted the standard α threshold from 0.05 to 0.0045 "[...] to control the experiment-wise error rate across the 11 tests [...]" (p. 5).

Study Setup and Variables

A clear description of the study design and setup are needed to understand how the study was conducted, what was measured and how, as well as how the researchers managed their degrees of freedom. Independent variables were thoroughly described in most papers (n=92, 77.31%), resulting in a clear vision of the manipulation, and facilitating conceptual replications.

Dependent variables were defined in less detail. Overall, 56.3% (n=67) papers reported them in ways that clarified their role in the study and made them available for replication. For exam-

ple, one paper used a subheading per questionnaire, stating "To quantify the play experience, we measure interest/enjoyment, invested effort and pressure/tension using the IMI" [42, p. 453], followed by a citation.

In papers where dependent variables were not precisely defined, the methods of their construction, relevance to the research, or potential moderator status was rarely stated [e.g., 54]. A majority of studies reported additional measures: in 51.26% (n=61) of papers, these variables were not described in a way they could be replicated or their role in the analysis was unclear (e.g., whether they were intended as dependent or moderator variables). Of the remaining papers, 14.29% (n=17) fully reported their additional variables, and 34.45% (n=41) did not report measuring additional variables. Most papers (n=88, 73.95%) did not employ moderator variables, but among those that did, few explained their use (n=14, 11.77%).

An exemplar of reporting measurements and manipulation can be found in Johanson et al. [81]. Moderator variables were collected "to get a sense of each participant's interest in the task and ability to complete the task" [81, p.174], and are subsumed under the "Questionnaires" subheading. Dependent variables are separately described in the next subsection "Dependent Measures".

I.4.3 Statistical Reporting

Of the 119 papers, n=80 (67.23%) report their tests in a way that communicates (1) the tests used, (2) degrees of freedom, (3) the test statistic, and (4) the p value. In the remaining papers, at least one of these elements was not clearly reported. An example of a well-reported ANOVA that also incorporates effect size can be found in [114, p. 196]: *"(F(2, 122) = 56.8, p < .001, η_p^2 =.482)"*.

When sufficient statistical details were reported, we used the app statcheck.io [121] to review the computation of p values from the test statistic and degrees of freedom. As statcheck only works for papers that report results formatted according to APA guidelines (and PDFs that directly translate into plain text), results were computed by hand where necessary. While 51.26% (n=61) of papers reported consistent p values, inconsistencies were observed in 28.57% (n=34) of papers. In most cases, inconsistencies reflected rounding errors with no meaningful influence on study outcomes – rarely, however, we observed decision inconsistencies, whereby the reported and computed p values supported different decisions. For 20.17% (n=24) of papers, it was not possible to re-compute p values, as they lacked necessary statistical details.

For example, one paper reported "$t(14) = -2.055, p = 0.049$" [22, p. 211], which would result in a significant $p = 0.0295$ for a one-sided test, or a non-significant $p = 0.059$ if a two-sided test was conducted. The same paper reported "$F(2, 12) = 3.775, p = 0.031$" (p. 211),

which would produce $p = 0.053$ [5]. Note that these were two tests among a total of 25 reported in the paper. As such, the observed inconsistencies do not change the overall conclusions of the paper by much.

Assumption Testing

Most papers (n=108, 90.76%) reported parametric tests, which should generally be accompanied by assumption tests – however, the majority of papers (n=64, 53.78%) did not mention these. Of the remaining papers, 36.98% (n=44) explicitly described assumption tests [e.g., 68, 69], the remaining 9.24% (n=11) used non-parametric tests [e.g., 160] or tests, where we were not aware of applicable assumption tests [e.g., 19].

Effect Sizes and Confidence Intervals

We also examined the prevalence of reporting effect sizes and confidence intervals (CIs): While we found effect sizes in 63.02% (n=75) of papers, only 6.72% (n=9) reported CIs [e.g., 58, 86] for point estimates of interest (usually effect sizes or means), suggesting that many studies rely solely on p values for their inferences. Lastly, few papers adjusted their significance threshold for multiple testing: 81.51% (n=97) of papers did not report adjusting their critical α level despite conducting multiple tests related to one hypothesis. Of the remaining works, 10.92% (n=13) performed adjustments [e.g., 114, 138], and a further 7.56% (n=9) did not require adjustment for multiple tests [e.g., 62, 94].

I.4.4 Transparency

Transparent data, analyses, and research goals allow other researchers to independently reproduce the analysis, or perform replication studies.

Only n=5 papers (4.2%) were accompanied by publicly available data [e.g., 40, 148, 160]; the remaining papers did not provide explanations for their non-disclosure. No papers shared the software script used for data analysis.

With regards to sharing experimental software, tools, or other materials, 20.2% (n=24) of papers sourced all relevant materials [e.g., 163], attaching questionnaires to the work [e.g., 162, 181], describing materials exhaustively [e.g., 78], or making software available in a repository [e.g., 148, 145, 147].

[5] This was determined with statcheck.io, and rechecked via www.socscistatistics.com/pvalues/tdistribution.aspx and www.socscistatistics.com/pvalues/fdistribution.aspx. Both methods produced consistent p values that differed from those stated in the text.

Finally, we investigated study pre-registration. Only one[6] paper [81] was pre-registered; curiously, the methods and tests described in the paper sometimes diverged from the pre-registration without explanation.

I.5 Discussion

The present work has reviewed 119 CHI PLAY papers employing NHST to examine the quality of statistical reporting practice. We have identified a number of issues with the ways that study design and data analysis are reported in these papers. NHST is an extremely popular analytic method at CHI PLAY, with our corpus comprising almost half of all published full papers from the venue.

However, critiques of NHST have emphasized the ease by which false positive results can emerge from seemingly reasonable practices conducted during and after data collection [i.e., QRPs, 177]. Our review raises similar concerns about quantitative research at CHI PLAY.

Two main issues can be identified from our review. First, statistical reporting varies widely: unexplained changes to sample size between tests, uncorrected multiple testing, and vaguely specified dependent variables are common. Inconsistent reporting problematizes evaluation of research quality, potentially obscuring questionable practices (e.g., incomplete test statistics). This ambiguity can drastically increase the rate of false positive results in the literature.

Second, we identified a number of arguably exploratory studies that apply methods intended for use in confirmatory research. These papers often *resemble* confirmatory work, featuring hypotheses that are tested using NHST. However, these papers often reflect an exploratory intent; they may, for example, employ a wide array of measures to evaluate a game (or game-like artefact) over a similarly extensive battery of tests. Without correcting for multiple tests, this approach inflates the rate of false positive findings.

Although some works report their results to a high standard, in general, the perfunctory application of NHST in CHI PLAY research is a cause for concern. The pursuit of meaningful claims about player-computer interaction – "discussion of current high quality research in games and HCI" [3] – is impeded by a quantitative literature for which effect sizes and confidence intervals are commonly elided in favour of isolated p values, and whose analytic methods fundamentally conflict with high-level goals of research.

[6] We note another pre-registered study [175] published at CHI PLAY 2019. However, the study was excluded from our review, as it describes a qualitative, exploratory approach and did not employ NHST.

I.5.1 The Value of Confirmatory and Exploratory Research

The majority of reviewed papers contained no confirmatory research goal, and had not formulated any statistical hypotheses. This may suggest that (1) most of these works actually pursue exploratory research aims, even if not explicitly stated in the paper; and that (2) confirmatory research is seemingly of limited use in the context of player-computer interaction.

Indeed, "*intentionally* exploratory studies are a cornerstone of HCI" [33, p. 5, emphasis added]. As such, the prevalence of exploratory work at CHI PLAY is unsurprising. Many publications describe the testing phase of iterative development, and evaluate novel interfaces [33]; for these applications, the rigor of confirmatory approaches is not always needed [76]. Moreover, in contrast to experimental psychology or medicine, player-computer interaction is yet a nascent field of research, in which the discovery of novel phenomena for theory-building remains a priority. Phillips et al. [129], for instance, formulated an exploratory research question to investigate how their reward taxonomy affects the player experience (i.e., "RQ1: Does type of video game reward in a game influence the player experience?", p. 396). As such, exploratory research is well-suited to investigating topics that are not sufficiently understood, or where firm predictions are impractical.

What value, then, does confirmatory research have to CHI PLAY? Recall that confirmatory studies test hypotheses derived from theoretical and conceptual speculation – in other words, they *build on prior work*. For example, the qualitatively greater autonomy identified in solitary play, relative to social play with friends [169] could be formally tested from a confirmatory perspective. While confirmatory research is less common in HCI, its relative absence has contributed to concerns regarding fragmentation and limited progress [75, 88]. Hence, confirmatory research is necessary to advance player-computer interaction by linking empirical work to theoretical considerations, validating conceptual assumptions (e.g., greater variation in rewards increases intrinsic motivation [cf. 129]), and understanding the processes involved in particular phenomena [75]. Together, these efforts facilitate more informed predictions, as well as contribute to a more unified and integrative understanding of player-computer interaction.

I.5.2 Directions

Our review has identified a slew of shortcomings in current research and reporting practices at CHI PLAY. However, we emphasize that the QRPs described in the present work are likely a product of unfamiliarity with the assumptions underlying NHST, rather than intentional data massaging. As noted, many R-DFs have only recently been identified as problematic [83].

We highlight that some CHI PLAY research is, at times, already conducted in partial align-

ment with Open Science principles. Schwind et al. [148], for example, made their *faceMaker* app freely available, facilitating its further use in research. Similarly, despite concerns regarding its arguably low statistical power, Johanson and colleagues' [81] thorough pre-registration and detailed reporting clarify the aims of the work and facilitate replication.

Finally, we highlight that while essential, reporting quality is not the *sine qua non* of publication value – indeed, there are many other aspects (e.g., subject matter, theory) that make papers interesting and worth reading [75, 179].

However, for CHI PLAY research to have proceeded largely in isolation from these discussions is worth further examination. Guidelines for high-quality study design [177, 154], analysis [158], and reporting practices [4] – many of which were compiled from within HCI [27, 33, 48, 75, 88, 170] – have been largely eschewed. While we can only speculate as to why these recommendations have not yet found their way to the field, we urge CHI PLAY scholars to engage with works such as these, and more completely apply these guidelines in their own practice. Moreover, we recommend that HCI and games educators sensitize students to the different roles of confirmatory and exploratory research.

I.6 A template for more transparent quantitative research at CHI PLAY

To facilitate more rigorous and transparent research and reporting practices at CHI PLAY, we contribute a template for researchers to guide their study designs from start to finish, and for reviewers to quickly assess reporting quality. Recommendations mostly focus on confirmatory, NHST-based research, but also include pointers for exploratory work. Suggestions are presented alongside their corresponding citations for further reference.

Note that this template is neither complete nor infallible – research is diverse, and no single template can perfectly address every work – however, it is intended to specifically address concerns of quantitative studies published at CHI PLAY at present. As with p values and effect sizes, researchers and authors should take a critical perspective and carefully consider the rationale for each point to determine where divergence is pertinent.

I.6.1 Deciding on the Research Goals

First, researchers need to consider their research question(s) of interest, and how these may be studied [75]. Specifically, whether their aim is to test specific hypotheses (confirmatory), collect rich descriptions of a phenomenon or artefact (exploratory), or a combination thereof (i.e.,

confirmatory hypothesis testing, followed by exploratory analyses). This decision then informs all subsequent methodological choices, as well as the selection of suitable statistical methods.

I.6.2 Hypothesizing

Designing a study should proceed with the research goals in mind, as they influence a number of decisions that follow. Submissions should clearly report the overarching research goal, research questions, and precise hypotheses.

- **Confirmatory Research Goal [4, 177]:** The work has a clearly stated confirmatory research goal: Authors specify two competing hypotheses. Deciding on a clear research goal early on can drastically improve the study design, as it informs all subsequent study design choices.
- **Precise & Directional Statistical Hypotheses [88, 177]:** Hypotheses should be explicitly defined, including predictions as to how the independent variable will impact specific measures. Steinemann et al. [157], for example, built on previous experiments in media psychology to formulate concrete statistical hypothesis, e.g., "H1: Interactivity will lead to increased donations" [157, p. 321].

I.6.3 Study Design

Designing the study follows directly from the research question(s).

- ***Independent* variables are precisely defined [4, 154, 177]:** Independent variables should be clearly described and justified to facilitate replication. For example, when selecting video games as stimuli for experimental conditions, clarifications should be provided to justify the choice [166].
- ***Dependent* variables are precisely defined [177]:** Full reporting of dependent variables paints a more complete picture of the research (and results), and facilitates replication. Authors should describe which variables they intended to influence, how these were measured, and why this instrument was chosen. Especially in light of the variety of available player experience questionnaires, psychometrics and a clear rationale should be provided [85, 109].
- **All additional and moderator variables are defined [177]:** *All* additional variables collected are clearly described, including their role in the analysis. Demographic details and other interesting constructs with no clearly specified relation to the research question may provide valuable insights, but unless specified in the hypotheses, should not be used in confirmatory analyses.

- **Data cleaning, exclusions and grouping [177]:** All data cleaning practices, exclusion of participants' data, and grouping participants by demographic variables should be clearly summarized [see also 23], alongside a rationale for their use. Justifying these decisions (e.g., listing predetermined exclusion criteria) emphasizes that these measures were not taken post-hoc to fish for significant results. Where participants are grouped via a third variable (e.g., age brackets), a multiverse analysis – in which all reasonable groupings are calculated [49] – may help researchers demonstrate the robustness of their decisions.
- **Data collection and power [158, 177]:** Controlling for statistical power is crucial to understanding the long-term error rate. This is usually done by collecting a sample of a specific size, as determined by a power analysis. Power analysis also helps reduce over-testing (i.e., potentially wasting resources), and facilitates the development of more specific hypotheses, as it requires a prediction concerning the expected effect's magnitude (i.e., effect size).
- **Deciding on an alpha threshold [45, 104]:** As with statistical power, a single critical alpha value should be determined prior to data collection. The alpha threshold should be considered strictly dichotomous: Marginally significant results (typically $0.05 \leq p \leq 0.1$) should not be interpreted [130]. While $\alpha < 0.05$ is standard, the chosen threshold value may differ depending on the needs of the research [104].
- **Visualize the study design [50]:** Tools exist to walk through the aforementioned steps, visualize the study design, and facilitate pre-registration. Touchstone2 [50], for example, allows researchers to set up and compare study designs, and perform power calculations.

I.6.4 Reporting

The aim of NHST is to decide between two competing hypotheses about data. Transparent reporting is necessary to comprehend these decisions, as well as promote reproducibility.

- **Full test statistics are reported [177]:** Statistical tests, degrees of freedom, and statistical values are reported in detail. A (fictitious) example of a fully reported one-sided, non-significant Student's t-test is depicted in Figure I.2.
- **Assumptions are tested and reported [26, 177]:** All statistical tests come with specific assumptions towards the data. While some tests (e.g., a one-way ANOVA) are considered robust to violations of normality [143], other assumptions such as variance homogeneity can affect common NHST tests, including ANOVA and ANCOVA [135], especially for smaller sample sizes. Papers should also clearly report if assumptions were not tested [e.g., tests of the normality assumption have been criticized for their unreliability 27].

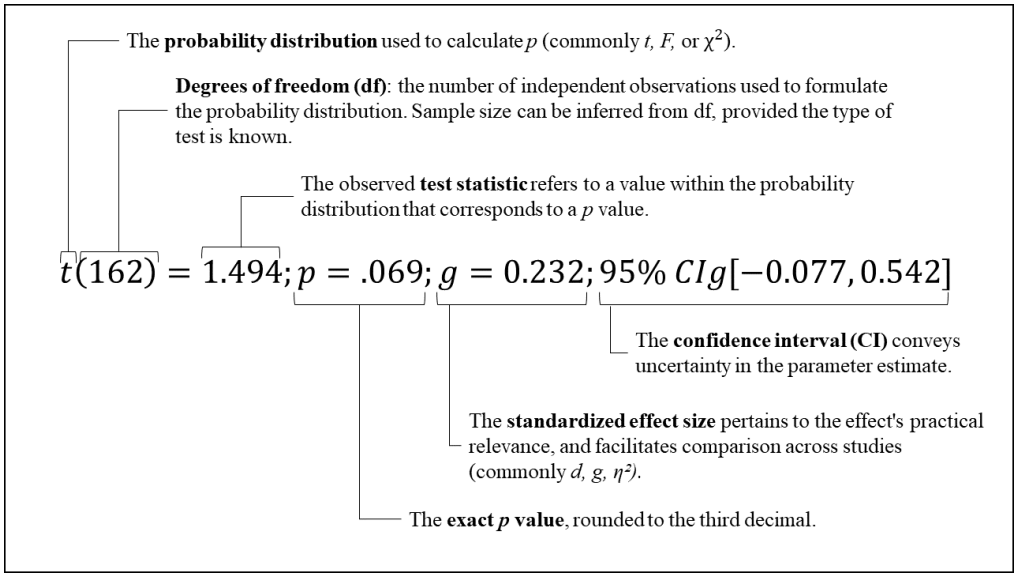

Fig. I.2: An illustration of a full test statistic. It includes all information the reader needs to understand which test was performed, how many data points were included, what the result is, and how the result may be interpreted.

- **Reporting effect sizes and confidence intervals [75, 90, 158]:** In contrast to p values, effect sizes indicate the strength of a statistical effect (e.g., how strongly the independent variable impacted the dependent variable), and provide a more meaningful basis for interpreting results. Confidence intervals should also be reported to indicate the degree of uncertainty around important point estimates (i.e., effect sizes, means, etc.). Reporting effect sizes and CIs also allows for comparisons across studies (e.g., as in meta analyses).
- **Correcting for multiple tests [16, 103]:** Family-wise error control is important to adjust for multiple testing, as every test beyond the first increases the rate of false positive findings. As a general rule, p values should be adjusted per hypothesis, where given k independent tests, the chance of observing a false positive is $1 - (1 - \alpha)^k$ [16]. Harpstead et al. [69], for example, chose a Bonferroni correction to adjust their chosen significance threshold of 0.001 "by the number of statistical tests being performed, (1 + the number of rewards) × (1 + the number of groups) × 2, resulting in a final alpha value of 5.05E^{-6}"[69, p.375].
- **Dealing with non-significant results [46, 130]:** As noted in Related Work, p values do not represent the probability that a hypothesis is true. Equally, non-significant results cannot be interpreted as evidence of "no effect": Any series of tests is likely to produce some non-significant results when statistical power is < 1. "Marginally significant" results are

non-significant, and should not be interpreted otherwise [130].

- **Matching hypotheses to tests [177]:** All hypotheses should be clearly linked to corresponding statistical tests, as well as clearly report any exploratory and post-hoc analyses.

I.6.5 Transparency

Public sharing of study data and materials increases transparency and trust [117, 116], while freeing up space in papers to concentrate on other relevant aspects (e.g., the game design process) and discuss the most interesting findings.

- **The data set is available [4, 117, 116, 170, 177]:** If possible, anonymized raw data should be made openly available in a persistent online repository. The location of the data should be explicitly noted in the text, preferably in the abstract, so that the data are available, even if the paper is not openly accessible. Where anonymized data cannot be made public (e.g., vulnerable populations), researchers could instead generate a distributionally identical 'synthetic' data set [see 132, for a primer].
- **The analysis plan is available [4, 170, 177]:** Sharing R scripts, SPSS syntax, or the data analysis plan allows researchers and reviewers to easily replicate the analysis.
- **Experimental artefacts are available [4, 170, 177]:** All materials necessary to replicate the study should be made openly available. For many CHI PLAY publications, this simply entails listing all questionnaire items, as well as sourcing or uploading all study materials. Where this is not possible (e.g., due to copyrighted hardware, software prototypes), researchers should share alternate resources to facilitate replication. Krekhov et al. [98], for instance, describe a blueprint for their controller prototype.

I.6.6 Pre-registration

As all aforementioned recommendations refer to decisions made prior to data collection, pre-registering a study requires little extra effort [158]. The pre-registration plan provides readers a means to identify confirmatory and exploratory goals of the work, and discourages HARKing [33].

- **Pre-registration [4, 33, 117, 131, 170, 177]:** The study is pre-registered at a permanent third-party archive (e.g., OSF.io or AsPredicted.org) and accompanied by timestamps. Johanson et al. [81], for instance, provide an example of a pre-registration plan in the context of player-computer interaction.

- **Deviations from pre-registration are justified [33, 117, 131, 170, 177]:** Studies do not always go as intended. Deviations from the pre-registration plan are often warranted, but should be clearly highlighted and justified in the paper.

I.6.7 Exploratory Research

The previous sections reflect recommendations primarily tailored to confirmatory work and NHST. However, in light of the importance of exploratory research for player-computer interaction, we provide a few pointers here for more transparent quantitative exploratory research.

- **Stating the exploratory research goal [33, 177]:** Intentional exploratory analyses are perfectly reasonable, when clearly declared as such. Researchers should specify their exploratory focus, and avoid presenting findings as definitive proof (e.g., via p values). In turn, reviewers should not insist on the provision of p values, as they tend to be misconstrued as strong evidence.
- **Adequate statistical approach:** Some research questions may be more appropriately investigated with other statistical methods. When relevant prior information exists, Bayesian approaches [100] may be useful; in other cases, researchers may benefit from estimation-based approaches [48], as in some existing CHI PLAY work [e.g., 47].
- **Interpreting results [36, 48]:** Exploratory analysis is less constrained than a confirmatory approach. Hence, researchers might forgo certain "rules" when interpreting and reporting results. For instance, instead of relying on p values, researchers may gauge graphs "by eye" to compare whether confidence intervals overlap between groups.
- **Transparent reporting:** Exploratory work should also follow the transparency guidelines described above, including pre-registration [33], open data, and complete reporting of all collected measures. Sharing the data openly also allows other researchers to engage in exploratory analyses, and extends the contribution of the original work [170].

I.7 Limitations and Future Work

First, it was not always obvious to determine whether the methodological best practices were followed in the reviewed papers. We could, for example, only investigate outlier removal based on the descriptions and rationales reported in the papers. Had the raw data been made openly available, it would have been possible to reproduce by what means outliers were identified and removed.

Second, screening and coding processes were conducted entirely by the first author. While these procedures were performed with transparency and completeness; however, some readers may disagree with aspects of the analysis.

Third, some points of critique around NHST remain topics of active debate; for example, whether and how to test assumptions [26, 27, 143], and when to adjust for multiple testing [103]. While we have made clear recommendations for many of these topics, we urge CHI PLAY researchers to more deeply engage with these discussions, to make informed decisions regarding their statistical analyses, and provide clear justifications in their papers.

Finally, our review is limited to NHST, researcher degrees of freedom, and their potential inflation of the Type I error rate. Yet our findings showcase further methodological concerns (e.g., suitability of statistical tests or study designs for answering research questions) that warrant consideration. Moreover, the present work mostly focuses on hypothesis-testing. However, recent work on the uses of psychological theory in HCI games research [165] also suggests a need to assess *how* hypotheses are generated and *what* research questions are formulated [75, 115].

I.8 Conclusion

Null Hypothesis Significance Testing (NHST) is a popular analytic tool at CHI PLAY. However, our review of 119 full papers highlights a number of inconsistencies and shortcomings with regards to research and reporting practices. These issues emerged against a backdrop of systematic misuse of confirmatory methods, such as NHST, in seemingly exploratory work. To help counter these issues, we present a template for authors to improve study design and statistical reporting, and for reviewers to evaluate work employing NHST. We are confident that by adopting basic Open Science standards – such as pre-registration, open data, and more uniform statistical reporting – the quality of CHI PLAY research may be further improved, fostering the validity and reliability of research findings.

Data availability Statement

The data, full analysis, as well as supplementary materials are available at https://osf.io/4mcbn/.

Declaration of Conflicting Interests

The authors declare no conflicting interests.

Author Contributions

JBV and EDM conceptualized the paper and designed the study. JBV conducted data collection and analysis. EDM and AT consulted on the analysis. EDM, AT and JBV wrote the paper.

Acknowledgments

We are grateful to Julia Ayumi Bopp for input and feedback during the coding and writing, as well as the reviewers for their encouraging comments.

References

[1] * Vero Vanden Abeele, Jan Wouters, Pol Ghesquière, Ann Goeleven, and Luc Geurts. 2015. Game-based Assessment of Psycho-acoustic Thresholds. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*. ACM Press. https://doi.org/10.1145/2793107.2793132

[2] * Andrea Abney, Brooke White, Jeremy Glick, Andre Bermudez, Paul Breckow, Jason Yow, Rayna Tillinghast-Trickett, and Paul Heath. 2014. Evaluation of recording methods for user test sessions on mobile devices. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play - CHI PLAY '14*. ACM Press. https://doi.org/10.1145/2658537.2658704

[3] ACM SIGCHI. 2020. CHI PLAY 2020 – November 1 – 4, 2020 Ottawa, Canada. Retrieved 2020-04-16 from https://chiplay.acm.org/2020/

[4] Balazs Aczel, Barnabas Szaszi, Alexandra Sarafoglou, Zoltan Kekecs, Šimon Kucharský, Daniel Benjamin, Christopher D Chambers, Agneta Fisher, Andrew Gelman, Morton A Gernsbacher, John P Ioannidis, Eric Johnson, Kai Jonas, Stavroula Kousta, Scott O Lilienfeld, D Stephen Lindsay, Candice C Morey, Marcus Monafò, Benjamin R Newell, Harold Pashler, David R Shanks, Daniel J Simons, Jelte M Wicherts, Dolores Albarracin, Nicole D Anderson, John Antonakis, Hal R Arkes, Mitja D Back, George C Banks, Christopher Beevers, Andrew A Bennett, Wiebke Bleidorn, Ty W Boyer, Cristina Cacciari, Alice S Carter, Joseph Cesario, Charles Clifton, Ronán M Conroy, Mike Cortese, Fiammetta Cosci, Nelson Cowan, Jarret Crawford, Eveline A Crone, John Curtin, Randall Engle, Simon Farrell, Pasco Fearon, Mark Fichman, Willem Frankenhuis, Alexandra M Freund, M Gareth

Gaskell, Roger Giner-Sorolla, Don P Green, Robert L Greene, Lisa L Harlow, Fernando Hoces de la Guardia, Derek Isaacowitz, Janet Kolodner, Debra Lieberman, Gordon D Logan, Wendy B Mendes, Lea Moersdorf, Brendan Nyhan, Jeffrey Pollack, Christopher Sullivan, Simine Vazire, and Eric Jan Wagenmakers. 2019. A consensus-based transparency checklist. *Nature Human Behaviour* (dec 2019). https://doi.org/10.1038/s41562-019-0772-6

[5] * Dmitry Alexandrovsky, Maximilian Achim Friehs, Max V. Birk, Rowan K. Yates, and Regan L. Mandryk. 2019. Game Dynamics that Support Snacking, not Feasting. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3311350.3347151

[6] * Sarah AlSulaiman and Michael S. Horn. 2015. Peter the Fashionista?. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*. ACM Press. https://doi.org/10.1145/2793107.2793127

[7] * Maximilian Altmeyer, Pascal Lessel, Marc Schubhan, Vladislav Hnatovskiy, and Antonio Krüger. 2019. Germ Destroyer - A Gamified System to Increase the Hand Washing Duration in Shared Bathrooms. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3311350.3347157

[8] Craig A. Anderson. 2004. An update on the effects of playing violent video games. *Journal of Adolescence* 27, 1 (feb 2004), 113–122. https://doi.org/10.1016/j.adolescence.2003.10.009

[9] * Dennis Ang and Alex Mitchell. 2017. Comparing Effects of Dynamic Difficulty Adjustment Systems on Video Game Experience. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116623

[10] * Dennis Ang and Alex Mitchell. 2019. Representation and Frequency of Player Choice in Player-Oriented Dynamic Difficulty Adjustment Systems. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3311350.3347165

[11] * Ivon Arroyo, Matthew Micciollo, Jonathan Casano, Erin Ottmar, Taylyn Hulse, and Ma. Mercedes Rodrigo. 2017. Wearable Learning. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116637

[12] * Louise Ashbarry, Benjamin Geelan, Kristy de Salas, and Ian Lewis. 2016. Blood and Violence. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*. ACM Press. https://doi.org/10.1145/2967934.2968111

[13] * Jeremy B. Badler and Alessandro Canossa. 2015. Anticipatory Gaze Shifts during Navigation in a Naturalistic Virtual Environment. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*. ACM Press. https://doi.org/10.1145/2793107.2793136

[14] * Alexander Baldwin, Daniel Johnson, and Peta Wyeth. 2016. Crowd-Pleaser. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*. ACM Press. https://doi.org/10.1145/2967934.2968100

[15] * Gabriel Barata, Sandra Gama, Joaquim A.P. Jorge, and Daniel J.V. Gonçalves. 2014. Relating gaming habits with student performance in a gamified learning experience. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play - CHI PLAY '14*. ACM Press. https://doi.org/10.1145/2658537.2658692

[16] Ralf Bender and Stefan Lange. 2001. Adjusting for multiple testing - When and how? *Journal of Clinical Epidemiology* 54, 4 (2001), 343–349. https://doi.org/10.1016/S0895-4356(00)00314-0

[17] * Max V. Birk, Maximilian A. Friehs, and Regan L. Mandryk. 2017. Age-Based Preferences and Player Experience. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116608

[18] * Max V. Birk, Regan L. Mandryk, and Cheralyn Atkins. 2016. The Motivational Push of Games. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*. ACM Press. https://doi.org/10.1145/2967934.2968091

[19] * Max V. Birk, Regan L. Mandryk, Matthew K. Miller, and Kathrin M. Gerling. 2015. How Self-Esteem Shapes our Interactions with Play Technologies. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*. ACM Press. https://doi.org/10.1145/2793107.2793111

[20] * Marion Boberg, Evangelos Karapanos, Jussi Holopainen, and Andrés Lucero. 2015. PLEXQ. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*. ACM Press. https://doi.org/10.1145/2793107.2793124

[21] * Jason T. Bowey and Regan L. Mandryk. 2017. Those are not the Stories you are Looking For. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116636

[22] * Evren Bozgeyikli, Andrew Raij, Srinivas Katkoori, and Rajiv Dubey. 2016. Point & Teleport Locomotion Technique for Virtual Reality. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*. ACM Press. https://doi.org/10.1145/2967934.2968105

[23] Florian Brühlmann, Serge Petralito, Lena F Aeschbach, and Klaus Opwis. 2020. The Quality of Data Collected Online: An Investigation of Careless Responding in a Crowdsourced Sample. *Methods in Psychology* (2020), 100022. https://doi.org/10.1016/j.metip.2020.100022

[24] Florian Brühlmann and Gian Marco Schmid. 2015. How to measure the game experience? Analysis of the factor structure of two questionnaires. In *Conference on Human Factors in Computing Systems - Proceedings*, Vol. 18. Association for Computing Machinery, New York, New York, USA, 1181–1186. https://doi.org/10.1145/2702613.2732831

[25] * Jie Cai, Donghee Yvette Wohn, and Guo Freeman. 2019. Who Purchases and Why?. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3311350.3347196

[26] Paul Cairns. 2007. HCI.. not as it should be: Inferential statistics in HCI research. *People and Computers XXI HCI.But Not as We Know It - Proceedings of HCI 2007: The 21st British HCI Group Annual Conference* 1 (2007), 195–201. https://doi.org/10.14236/ewic/hci2007.20

[27] Paul Cairns. 2019. *Doing Better Statistics in Human-Computer Interaction*. Cambridge University Press. https://doi.org/10.1017/9781108685139

[28] * Murat Perit Cakir, Nur Akkuş Çakir, Hasan Ayaz, and Frank J. Lee. 2015. An Optical Brain Imaging Study on the Improvements in Mathematical Fluency from Game-based Learning. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*. ACM Press. https://doi.org/10.1145/2793107.2793133

[29] * Jared E. Cechanowicz, Carl Gutwin, Scott Bateman, Regan Mandryk, and Ian Stavness. 2014. Improving player balancing in racing games. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play - CHI PLAY '14*. ACM Press. https://doi.org/10.1145/2658537.2658701

[30] * Anjana Chatta, Tyler Hurst, Gayani Samaraweera, Rongkai Guo, and John Quarles. 2015. Get off the Couch. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*. ACM Press. https://doi.org/10.1145/2793107.2793115

[31] * Jinghui Cheng, Dorian Anderson, Cynthia Putnam, and Jin Guo. 2017. Leveraging Design Patterns to Support Designer-Therapist Collaboration When Ideating Brain Injury Therapy Games. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116600

[32] * Sebastian Cmentowski, Andrey Krekhov, and Jens Krüger. 2019. Outstanding: A Multi-Perspective Travel Approach for Virtual Reality Games. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3311350.3347183

[33] Andy Cockburn, Carl Gutwin, and Alan Dix. 2018. HARK no more: On the preregistration of chi experiments. *Conference on Human Factors in Computing Systems - Proceedings* 2018-April (2018), 1–12. https://doi.org/10.1145/3173574.3173715

[34] Geoff Cumming. 2009. Dance p 3 Mar09. https://www.youtube.com/watch?v=ez4DgdurRPg

[35] Geoff Cumming. 2013. *Understanding The New Statistics*. https://doi.org/10.4324/9780203807002

[36] Geoff Cumming and Sue Finch. 2005. Inference by eye confidence intervals and how to read pictures of data. *American Psychologist* 60, 2 (2005), 170–180. https://doi.org/10.1037/0003-066X.60.2.170

[37] * Martin Dechant, Ian Stavness, Aristides Mairena, and Regan L. Mandryk. 2018. Empirical Evaluation of Hybrid Gaze-Controller Selection Techniques in a Gaming Context. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. ACM Press. https://doi.org/10.1145/3242671.3242699

[38] * Alena Denisova and Paul Cairns. 2015. The Placebo Effect in Digital Games. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*. ACM Press. https://doi.org/10.1145/2793107.2793109

[39] * Alena Denisova and Eliott Cook. 2019. Power-Ups in Digital Games. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3311350.3347173

[40] * Alena Denisova, A. Imran Nordin, and Paul Cairns. 2016. The Convergence of Player Experience Questionnaires. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*. ACM Press. https://doi.org/10.1145/2967934.2968095

[41] * Ansgar E. Depping, Colby Johanson, and Regan L. Mandryk. 2018. Designing for Friendship. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. ACM Press. https://doi.org/10.1145/3242671.3242702

[42] * Ansgar E. Depping and Regan L. Mandryk. 2017. Cooperation and Interdependence. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116639

[43] * Ansgar E. Depping, Regan L. Mandryk, Colby Johanson, Jason T. Bowey, and Shelby C. Thomson. 2016. Trust Me. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*. ACM Press. https://doi.org/10.1145/2967934.2968097

[44] * Arindam Dey, Hao Chen, Mark Billinghurst, and Robert W. Lindeman. 2018. Effects of Manipulating Physiological Feedback in Immersive Virtual Environments. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. ACM Press. https://doi.org/10.1145/3242671.3242676

[45] Zoltan Dienes. 2008. *Understanding psychology as a science: An introduction to scientific and statistical inference*. Macmillan International Higher Education.

[46] Zoltan Dienes. 2014. Using Bayes to get the most out of non-significant results. *Frontiers in Psychology* 5 (jul 2014), 781. https://doi.org/10.3389/fpsyg.2014.00781

[47] * Gabriella Dodero, Rosella Gennari, Alessandra Melonio, and Santina Torello. 2014. Towards tangible gamified co-design at school. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play - CHI PLAY '14*. ACM Press. https://doi.org/10.1145/2658537.2658688

[48] Pierre Dragicevic. 2016. Fair Statistical Communication in HCI. In *Modern Statistical Methods for HCI*. 291–330. https://doi.org/10.1007/978-3-319-26633-6_13 arXiv:arXiv:1011.1669v3

[49] Pierre Dragicevic, Yvonne Jansen, Abhraneel Sarma, Matthew Kay, and Fanny Chevalier. 2019. Increasing the transparency of research papers with explorable multiverse analyses. *Conference on Human Factors in Computing Systems - Proceedings* (2019), 1–15. https://doi.org/10.1145/3290605.3300295

[50] Alexander Eiselmayer, Chat Wacharamanotham, Michel Beaudouin-Lafon, and Wendy E. Mackay. 2019. Touchstone2: An Interactive Environment for Exploring Trade-offs in HCI Experiment Design. *Conference on Human Factors in Computing Systems - Proceedings* (2019), 1–11. https://doi.org/10.1145/3290605.3300447

[51] Malte Elson and Thorsten Quandt. 2016. Digital games in laboratory experiments: Controlling a complex stimulus through modding. *Psychology of Popular Media Culture* 5, 1 (jan 2016), 52–65. https://doi.org/10.1037/ppm0000033

[52] * Katharina Emmerich and Maic Masuch. 2016. The Influence of Virtual Agents on Player Experience and Performance. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*. ACM Press. https://doi.org/10.1145/2967934.2968092

[53] * Katharina Emmerich and Maic Masuch. 2017. The Impact of Game Patterns on Player Experience and Social Interaction in Co-Located Multiplayer Games. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116606

[54] * Katharina Emmerich, Patrizia Ring, and Maic Masuch. 2018. I'm Glad You Are on My Side. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. ACM Press. https://doi.org/10.1145/3242671.3242709

[55] * Zachary Fitz-Walter, Peta Wyeth, Dian Tjondronegoro, and Daniel Johnson. 2014. Exploring the effect of achievements on students attending university orientation. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play - CHI PLAY '14*. ACM Press. https://doi.org/10.1145/2658537.2658700

[56] Wolfgang Forstmeier, Eric Jan Wagenmakers, and Timothy H. Parker. 2017. Detecting and avoiding likely false-positive findings – a practical guide. *Biological Reviews* 92, 4 (2017), 1941–1968. https://doi.org/10.1111/brv.12315

[57] * Julian Frommel, Kim Fahlbusch, Julia Brich, and Michael Weber. 2017. The Effects of Context-Sensitive Tutorials in Virtual Reality Games. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116610

[58] * Julian Frommel, Fabian Fischbach, Katja Rogers, and Michael Weber. 2018. Emotion-based Dynamic Difficulty Adjustment Using Parameterized Difficulty and Self-Reports of

Emotion. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. ACM Press. https://doi.org/10.1145/3242671.3242682

[59] * Julian Frommel, Katja Rogers, Thomas Dreja, Julian Winterfeldt, Christian Hunger, Maximilian Bär, and Michael Weber. 2016. 2084 – Safe New World. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*. ACM Press. https://doi.org/10.1145/2967934.2968087

[60] * Julian Frommel, Michael Weber, Katja Rogers, Julia Brich, Daniel Besserer, Leonard Bradatsch, Isabel Ortinau, Ramona Schabenberger, Valentin Riemer, and Claudia Schrader. 2015. Integrated Questionnaires. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*. ACM Press. https://doi.org/10.1145/2793107.2793130

[61] * Yue Gao, Kathrin M. Gerling, Regan L. Mandryk, and Kevin G. Stanley. 2014. Decreasing sedentary behaviours in pre-adolescents using casual exergames at school. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play - CHI PLAY '14*. ACM Press. https://doi.org/10.1145/2658537.2658693

[62] * Luc Geurts, Vero Vanden Abeele, Kevin Van Keer, and Ruben Isenborghs. 2014. Playfully learning visual perspective taking skills with sifteo cubes. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play - CHI PLAY '14*. ACM Press. https://doi.org/10.1145/2658537.2658706

[63] * Thomas A. Goldman, Frank J. Lee, and Jichen Zhu. 2014. Using video games to facilitate understanding of attention deficit hyperactivity disorder. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play - CHI PLAY '14*. ACM Press. https://doi.org/10.1145/2658537.2658707

[64] * Antti Granqvist, Tapio Takala, Jari Takatalo, and Perttu Hämäläinen. 2018. Exaggeration of Avatar Flexibility in Virtual Reality. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. ACM Press. https://doi.org/10.1145/3242671.3242694

[65] Sander Greenland, Stephen J. Senn, Kenneth J. Rothman, John B. Carlin, Charles Poole, Steven N. Goodman, and Douglas G. Altman. 2016. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology* 31, 4 (apr 2016), 337–350. https://doi.org/10.1007/s10654-016-0149-3

[66] * Nathan Navarro Griffin, James Liu, and Eelke Folmer. 2018. Evaluation of Handsbusy vs Handsfree Virtual Locomotion. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. ACM Press. https://doi.org/10.1145/3242671.3242707

[67] * Carl Gutwin, Rodrigo Vicencio-Moreira, and Regan L. Mandryk. 2016. Does Helping Hurt?. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*. ACM Press. https://doi.org/10.1145/2967934.2968101

[68] * Stuart Hallifax, Audrey Serna, Jean-Charles Marty, Guillaume Lavoué, and Elise Lavoué. 2019. Factors to Consider for Tailored Gamification. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3311350.3347167

[69] * Erik Harpstead, Thomas Zimmermann, Nachiappan Nagapan, Jose J. Guajardo, Ryan Cooper, Tyson Solberg, and Dan Greenawalt. 2015. What Drives People. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*. ACM Press. https://doi.org/10.1145/2793107.2793114

[70] * John Harris, Mark Hancock, and Stacey D. Scott. 2016. Leveraging Asymmetries in Multiplayer Games. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*. ACM Press. https://doi.org/10.1145/2967934.2968113

[71] * Jennefer Hart, Ioanna Iacovides, Anne Adams, Manuel Oliveira, and Maria Margoudi. 2017. Understanding Engagement within the Context of a Safety Critical Game. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116633

[72] * Kieran Hicks, Kathrin Gerling, Patrick Dickinson, and Vero Vanden Abeele. 2019. Juicy Game Design. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3311350.3347171

[73] Joseph Hilgard, Christopher R. Engelhardt, Bruce D. Bartholow, and Jeffrey N. Rouder. 2017. How much evidence is p >.05? Stimulus pre-testing and null primary outcomes in violent video games research. *Psychology of Popular Media Culture* 6, 4 (oct 2017), 361–380. https://doi.org/10.1037/ppm0000102

[74] * Britton Horn, Seth Cooper, and Sebastian Deterding. 2017. Adapting Cognitive Task Analysis to Elicit the Skill Chain of a Game. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116640

[75] Kasper Hornbæk. 2011. Some whys and hows of experiments in human-computer interaction. *Foundations and Trends in Human-Computer Interaction* 5, 4 (2011), 299–373. https://doi.org/10.1561/1100000043

[76] Kasper Hornbæk, Søren S. Sander, Javier Bargas-Avila, and Jakob Grue Simonsen. 2014. Is once enough? on the extent and content of replications in human-computer interaction. *Conference on Human Factors in Computing Systems - Proceedings* (2014), 3523–3532. https://doi.org/10.1145/2556288.2557004

[77] * Ioanna Iacovides, Anna Cox, Richard Kennedy, Paul Cairns, and Charlene Jennett. 2015. Removing the HUD. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*. ACM Press. https://doi.org/10.1145/2793107.2793120

[78] * John Porter III, Matthew Boyer, and Andrew Robb. 2018. Guidelines on Successfully Porting Non-Immersive Games to Virtual Reality. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. ACM Press. https://doi.org/10.1145/3242671.3242677

[79] John P. A. Ioannidis. 2005. Why Most Published Research Findings Are False. *PLoS Medicine* 2, 8 (aug 2005), e124. https://doi.org/10.1371/journal.pmed.0020124

[80] * Aliya Iskenderova, Florian Weidner, and Wolfgang Broll. 2017. Drunk Virtual Reality Gaming. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116618

[81] * Colby Johanson, Carl Gutwin, Jason T. Bowey, and Regan L. Mandryk. 2019. Press Pause when you Play. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3311350.3347195

[82] * Colby Johanson, Carl Gutwin, and Regan L. Mandryk. 2017. The Effects of Navigation Assistance on Spatial Learning and Performance in a 3D Game. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116602

[83] Leslie K. John, George Loewenstein, and Drazen Prelec. 2012. Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychological Science* 23, 5 (2012), 524–532. https://doi.org/10.1177/0956797611430953

[84] * Daniel Johnson, Christopher Watling, John Gardner, and Lennart E. Nacke. 2014. The edge of glory. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play - CHI PLAY '14*. ACM Press. https://doi.org/10.1145/2658537.2658694

[85] Daniel Johnson, M. John Gardner, and Ryan Perry. 2018. Validation of two game experience scales: The Player Experience of Need Satisfaction (PENS) and Game Experience Questionnaire (GEQ). *International Journal of Human Computer Studies* 118 (oct 2018), 38–46. https://doi.org/10.1016/j.ijhcs.2018.05.003

[86] * Dennis L. Kappen, Pejman Mirza-Babaei, Jens Johannsmeier, Daniel Buckstein, James Robb, and Lennart E. Nacke. 2014. Engaged by boos and cheers. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play - CHI PLAY '14*. ACM Press. https://doi.org/10.1145/2658537.2658687

[87] * Dennis L. Kappen, Pejman Mirza-Babaei, and Lennart E. Nacke. 2017. Gamification through the Application of Motivational Affordances for Physical Activity Technology. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116604

[88] Maurits Kaptein and Judy Robertson. 2012. Rethinking statistical analysis methods for CHI. *Conference on Human Factors in Computing Systems - Proceedings* (2012), 1105–1113. https://doi.org/10.1145/2207676.2208557

[89] * Geoff Kaufman, Mary Flanagan, and Gili Freedman. 2019. Not Just for Girls. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3311350.3347177

[90] Ken Kelley and Joseph R Rausch. 2006. Sample Size Planning for the Standardized Mean Difference : Accuracy in Parameter Estimation Via Narrow Confidence Intervals. 11, 4 (2006), 363–385. https://doi.org/10.1037/1082-989X.11.4.363

[91] Norbert L. Kerr. 1998. HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review* 2, 3 (aug 1998), 196–217. https://doi.org/10.1207/s15327957pspr0203_4

[92] * Mallory Ketcheson, Zi Ye, and T.C. Nicholas Graham. 2015. Designing for Exertion. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*. ACM Press. https://doi.org/10.1145/2793107.2793122

[93] * Soomin Kim, Gyuho Lee, Seo young Lee, Sanghyuk Lee, and Joonhwan Lee. 2019. Game or Live Streaming?: Motivation and Social Experience in Live Mobile Quiz Shows. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3311350.3347187

[94] * Madison Klarkowski, Daniel Johnson, Peta Wyeth, Cody Phillips, and Simon Smith. 2018. Don't Sweat the Small Stuff. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. ACM Press. https://doi.org/10.1145/3242671.3242714

[95] * Boriana Koleva, Peter Tolmie, Patrick Brundell, Steve Benford, and Stefan Rennick Egglestone. 2015. From Front-End to Back-End and Everything In-Between. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*. ACM Press. https://doi.org/10.1145/2793107.2793131

[96] * Andrey Krekhov, Sebastian Cmentowski, Katharina Emmerich, and Jens Krüger. 2019. Beyond Human. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3311350.3347172

[97] * Andrey Krekhov, Sebastian Cmentowski, Katharina Emmerich, Maic Masuch, and Jens Krüger. 2018. GulliVR. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. ACM Press. https://doi.org/10.1145/3242671.3242704

[98] * Andrey Krekhov, Katharina Emmerich, Philipp Bergmann, Sebastian Cmentowski, and Jens Krüger. 2017. Self-Transforming Controllers for Virtual Reality First Person Shooters. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116615

[99] * Sven Krome, Jussi Holopainen, and Stefan Greuter. 2017. AutoGym. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116626

[100] John K. Kruschke. 2013. Bayesian estimation supersedes the T test. *Journal of Experimental Psychology: General* 142, 2 (2013), 573–588. https://doi.org/10.1037/a0029177 arXiv:/dx.doi.org/10.1037/a0029146 [http:]

[101] Daniël Lakens. 2014. The 20% Statistician: Observed power, and what to do if your editor asks for post-hoc power analyses. Retrieved 2020-01-22 from https://daniellakens.blogspot.com/2014/12/observed-power-and-what-to-do-if-your.html

[102] Daniël Lakens. 2016. Improving your statistical inferences - Week 1: Introduction + Frequentist Statistics. Retrieved 2020-03-10 from https://www.coursera.org/learn/statistical-inferences/home/week/1

[103] Daniël Lakens. 2020. The 20% Statistician: What's a family in family-wise error control? Retrieved 2020-03-16 from https://daniellakens.blogspot.com/2020/03/whats-family-in-family-wise-error.html

[104] Daniel Lakens, Federico G. Adolfi, Casper J. Albers, Farid Anvari, Matthew A.J. Apps, Shlomo E. Argamon, Thom Baguley, Raymond B. Becker, Stephen D. Benning, Daniel E. Bradford, Erin M. Buchanan, Aaron R. Caldwell, Ben Van Calster, Rickard Carlsson, Sau Chin Chen, Bryan Chung, Lincoln J. Colling, Gary S. Collins, Zander Crook, Emily S. Cross, Sameera Daniels, Henrik Danielsson, Lisa Debruine, Daniel J. Dunleavy, Brian D. Earp, Michele I. Feist, Jason D. Ferrell, James G. Field, Nicholas W. Fox, Amanda Friesen, Caio Gomes, Monica Gonzalez-Marquez, James A. Grange, Andrew P. Grieve, Robert Guggenberger, James Grist, Anne Laura Van Harmelen, Fred Hasselman, Kevin D. Hochard, Mark R. Hoffarth, Nicholas P. Holmes, Michael Ingre, Peder M. Isager, Hanna K. Isotalus, Christer Johansson, Konrad Juszczyk, David A. Kenny, Ahmed A. Khalil, Barbara Konat, Junpeng Lao, Erik Gahner Larsen, Gerine M.A. Lodder, Jiří Lukavský, Christopher R. Madan, David Manheim, Stephen R. Martin, Andrea E. Martin, Deborah G. Mayo, Randy J. McCarthy, Kevin McConway, Colin McFarland, Amanda Q.X. Nio, Gustav Nilsonne, Cilene Lino De Oliveira, Jean Jacques Orban De Xivry, Sam Parsons, Gerit Pfuhl, Kimberly A. Quinn, John J. Sakon, S. Adil Saribay, Iris K. Schneider, Manojkumar Selvaraju, Zsuzsika Sjoerds, Samuel G. Smith, Tim Smits, Jeffrey R. Spies, Vishnu Sreekumar, Crystal N. Steltenpohl, Neil Stenhouse, Wojciech Świątkowski, Miguel A. Vadillo, Marcel A.L.M. Van Assen, Matt N. Williams, Samantha E. Williams, Donald R. Williams, Tal Yarkoni, Ignazio Ziano, and Rolf A. Zwaan. 2018. Justify your alpha. *Nature Human Behaviour* 2, 3 (2018), 168–171. https://doi.org/10.1038/s41562-018-0311-x

[105] * Matthew Lakier, Lennart E. Nacke, Takeo Igarashi, and Daniel Vogel. 2019. Cross-Car, Multiplayer Games for Semi-Autonomous Driving. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3311350.3347166

[106] * Nicole Lane and Nathan R. Prestopnik. 2017. Diegetic Connectivity. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116630

[107] * Michael Lankes, Jüergen Hagler, Georgi Kostov, and Jeremiah Diephuis. 2017. Invisible Walls. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116609

[108] * Michael Lankes, Thomas Mirlacher, Stefan Wagner, and Wolfgang Hochleitner. 2014. Whom are you looking for?. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play - CHI PLAY '14*. ACM Press. https://doi.org/10.1145/2658537.2658698

[109] * Effie L.-C. Law, Florian Brühlmann, and Elisa D. Mekler. 2018. Systematic Review and Validation of the Game Experience Questionnaire (GEQ) - Implications for Citation and Reporting Practice. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. ACM Press. https://doi.org/10.1145/3242671.3242683

[110] * Pascal Lessel, Maximilian Altmeyer, and Nicolas Brauner. 2019. Crowdjump: Investigating a Player-Driven Platform Game. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3311350.3347168

[111] * Jingya Li, Erik D. van der Spek, Jun Hu, and Loe Feijs. 2019. Turning Your Book into a Game. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3311350.3347174

[112] * Michael Long and Carl Gutwin. 2018. Characterizing and Modeling the Effects of Local Latency on Game Performance and Experience. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. ACM Press. https://doi.org/10.1145/3242671.3242678

[113] * Bernhard Maurer, Ilhan Aslan, Martin Wuchse, Katja Neureiter, and Manfred Tscheligi. 2015. Gaze-Based Onlooker Integration. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*. ACM Press. https://doi.org/10.1145/2793107.2793126

[114] * Mitchell W. McEwan, Alethea L. Blackler, Daniel M. Johnson, and Peta A. Wyeth. 2014. Natural mapping and intuitive interaction in videogames. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play - CHI PLAY '14*. ACM Press. https://doi.org/10.1145/2658537.2658541

[115] William J McGuire. 1997. Creative hypothesis generating in psychology: Some useful heuristics. *Annual review of psychology* 48, 1 (1997), 1–30. https://doi.org/10.1146/annurev.psych.48.1.1

[116] Tsuyoshi Miyakawa. 2020. No raw data, no science: another possible source of the reproducibility crisis. *Molecular brain* 13, 1 (2020), 24. https://doi.org/10.1186/s13041-020-0552-2

[117] Marcus R. Munafò, Brian A. Nosek, Dorothy V.M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie Du Sert, Uri Simonsohn, Eric Jan Wagenmakers, Jennifer J. Ware, and John P.A. Ioannidis. 2017. A manifesto for reproducible science. *Nature Human Behaviour* 1, 1 (2017), 1–9. https://doi.org/10.1038/s41562-016-0021

[118] * John E. Muñoz, M. Cameirão, S. Bermúdez i Badia, and E. Rubio Gouveia. 2018. Closing the Loop in Exergaming - Health Benefits of Biocybernetic Adaptation in Senior Adults. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. ACM Press. https://doi.org/10.1145/3242671.3242673

[119] * Juliana Nazare, Anneli Hershman, Ivan Sysoev, and Deb Roy. 2017. Bilingual SpeechBlocks. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116616

[120] * Joshua Newn, Eduardo Velloso, Fraser Allison, Yomna Abdelrahman, and Frank Vetere. 2017. Evaluating Real-Time Gaze Representations to Infer Intentions in Competitive Turn-Based Strategy Games. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116624

[121] Michèle B. Nuijten. 2016. A spellchecker for statistics. , 151–152 pages. http://blogs.lse.ac.uk/impactofsocialsciences/2018/02/28/statcheck-a-spellchecker-for-statistics/

[122] Open Science Framework. 2016. OSF | Templates of OSF Registration Forms. https://osf.io/zab38/{#}!

[123] * Pablo Ortiz and D. Fox Harrell. 2018. Enabling Critical Self-Reflection through Roleplay with Chimeria. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. ACM Press. https://doi.org/10.1145/3242671.3242687

[124] * Raul Paradeda, Maria José Ferreira, Raquel Oliveira, Carlos Martinho, and Ana Paiva. 2019. The Role of Assertiveness in a Storytelling Game with Persuasive Robotic Non-Player Characters. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3311350.3347162

[125] * Pratheep Kumar Paranthaman and Seth Cooper. 2019. ARAPID: Towards Integrating Crowdsourced Playtesting into the Game Development Environment. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3311350.3347163

[126] * Taiwoo Park, Tianyu Hu, and Jina Huh. 2016. Plant-based Games for Anxiety Reduction. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*. ACM Press. https://doi.org/10.1145/2967934.2968094

[127] * Cale J. Passmore and Regan Mandryk. 2018. An About Face: Diverse Representation in Games. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*

(Melbourne, VIC, Australia) *(CHI PLAY '18)*. Association for Computing Machinery, New York, NY, USA, 365–380. https://doi.org/10.1145/3242671.3242711

[128] * Johannes Pfau, Jan David Smeddinck, and Rainer Malaka. 2018. Towards Deep Player Behavior Models in MMORPGs. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play* (Melbourne, VIC, Australia) *(CHI PLAY '18)*. Association for Computing Machinery, New York, NY, USA, 381–392. https://doi.org/10.1145/3242671.3242706

[129] * Cody Phillips, Daniel Johnson, Madison Klarkowski, Melanie Jade White, and Leanne Hides. 2018. The Impact of Rewards and Trait Reward Responsiveness on Player Motivation. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. ACM Press. https://doi.org/10.1145/3242671.3242713

[130] Laura Pritschet, Derek Powell, and Zachary Horne. 2016. Marginally significant effects as evidence for hypotheses: Changing attitudes over four decades. *Psychological Science* 27, 7 (2016), 1036–1042. https://doi.org/10.1177/0956797616645672

[131] Xiaoying Pu, Matthew Kay, Licheng Zhu, and Frederick Conrad. 2019. Designing for preregistration: A user-centered perspective. *Conference on Human Factors in Computing Systems - Proceedings* (2019), 1–6. https://doi.org/10.1145/3290607.3312862

[132] Daniel S. Quintana. 2019. Synthetic datasets: A non-technical primer for the behavioural sciences to promote reproducibility and hypothesis-generation. (2019). https://doi.org/10.31234/osf.io/dmfb3

[133] * George E. Raptis, Christos A. Fidas, and Nikolaos M. Avouris. 2016. Do Field Dependence-Independence Differences of Game Players Affect Performance and Behaviour in Cultural Heritage Games?. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*. ACM Press. https://doi.org/10.1145/2967934.2968107

[134] * Anke V. Reinschluessel and Regan L. Mandryk. 2016. Using Positive or Negative Reinforcement in Neurofeedback Games for Training Self-Regulation. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*. ACM Press. https://doi.org/10.1145/2967934.2968085

[135] David C. Rheinheimer and Douglas A. Penfield. 2001. The Effects of Type I Error Rate and Power of the ANCOVA F Test and Selected Alternatives Under Nonnormality and

Variance Heterogeneity. *The Journal of Experimental Education* 69, 4 (jan 2001), 373–391. https://doi.org/10.1080/00220970109599493

[136] * Katja Rogers, Matthias Jörg, and Michael Weber. 2019. Effects of Background Music on Risk-Taking and General Player Experience. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3311350.3347158

[137] Robert Rosenthal. 1979. The file drawer problem and tolerance for null results. *Psychological Bulletin* 86, 3 (1979), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

[138] * Rufat Rzayev, Sven Mayer, Christian Krauter, and Niels Henze. 2019. Notification in VR. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3311350.3347190

[139] * Pejman Sajjadi, Edgar Omar Cebolledo Gutierrez, Sandra Trullemans, and Olga De Troyer. 2014. Maze commander. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play - CHI PLAY '14*. ACM Press. https://doi.org/10.1145/2658537.2658690

[140] * Cheryl Savery and Nicholas Graham. 2014. Reducing the negative effects of inconsistencies in networked games. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play - CHI PLAY '14*. ACM Press. https://doi.org/10.1145/2658537.2658539

[141] * Mike Schaekermann, Giovanni Ribeiro, Guenter Wallner, Simone Kriglstein, Daniel Johnson, Anders Drachen, Rafet Sifa, and Lennart E. Nacke. 2017. Curiously Motivated. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116603

[142] Ulrich Schimmack. 2015. Questionable Research Practices: Definition, Detect, and Recommendations for Better Practices. Retrieved 2020-01-22 from https://replicationindex.com/2015/01/24/questionable-research-practices-definition-detect-and-recommendations-for-better-practices/

[143] Emanuel Schmider, Matthias Ziegler, Erik Danay, Luzi Beyer, and Markus Bühner. 2010. Is It Really Robust?: Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology* 6, 4 (2010), 147–151. https://doi.org/10.1027/1614-2241/a000016

[144] Felix D. Schönbrodt. 2016. p-Hacker: Train your p-hacking skills! Retrieved 17.04.2020 from http://shinyapps.org/apps/p-hacker/

[145] * Valentin Schwind and Niels Henze. 2018. Gender- and Age-related Differences in Designing the Characteristics of Stereotypical Virtual Faces. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. ACM Press. https://doi.org/10.1145/3242671.3242692

[146] * Valentin Schwind, Pascal Knierim, Lewis Chuang, and Niels Henze. 2017. "Where's Pinky?". In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116596

[147] * Valentin Schwind, Sven Mayer, Alexandre Comeau-Vermeersch, Robin Schweigert, and Niels Henze. 2018. Up to the Finger Tip. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. ACM Press. https://doi.org/10.1145/3242671.3242675

[148] * Valentin Schwind, Katrin Wolf, Niels Henze, and Oliver Korn. 2015. Determining the Characteristics of Preferred Virtual Faces Using an Avatar Generator. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*. ACM Press. https://doi.org/10.1145/2793107.2793116

[149] * Sven Seele, Sebastian Misztal, Helmut Buhler, Rainer Herpers, and Jonas Schild. 2017. Here's Looking At You Anyway!. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116619

[150] * Hanieh Shakeri, Samarth Singhal, Rui Pan, Carman Neustaedter, and Anthony Tang. 2017. Escaping Together. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116601

[151] Larissa Shamseer, David Moher, Mike Clarke, Davina Ghersi, Alessandro Liberati, Mark Petticrew, Paul Shekelle, Lesley A. Stewart, Douglas G. Altman, Alison Booth, An Wen Chan, Stephanie Chang, Tammy Clifford, Kay Dickersin, Matthias Egger, Peter C. Gøtzsche, Jeremy M. Grimshaw, Trish Groves, Mark Helfand, Julian Higgins, Toby Lasserson, Joseph Lau, Kathleen Lohr, Jessie McGowan, Cynthia Mulrow, Melissa Norton, Matthew Page, Margaret Sampson, Holger Schünemann, Iveta Simera, William Summerskill, Jennifer Tetzlaff, Thomas A. Trikalinos, David Tovey, Lucy Turner, and Evelyn Whitlock. 2015. Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015: Elaboration and explanation. *BMJ (Online)* 349, December 2014 (2015), 1–25. https://doi.org/10.1136/bmj.g7647

[152] * Hitesh Nidhi Sharma, Z. O. Toups, Igor Dolgov, Andruid Kerne, and Ajit Jain. 2016. Evaluating Display Modalities Using a Mixed Reality Game. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*. ACM Press. https://doi.org/10.1145/2967934.2968090

[153] * Mike Sheinin and Carl Gutwin. 2015. Quantifying Individual Differences, Skill Development, and Fatigue Effects in Small-Scale Exertion Interfaces. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*. ACM Press. https://doi.org/10.1145/2793107.2793129

[154] Joseph P. Simmons, Leif D. Nelson, and Uri Simonsohn. 2011. False-Positive Psychology. *Psychological Science* 22, 11 (nov 2011), 1359–1366. https://doi.org/10.1177/0956797611417632

[155] * Kristin Siu and Mark O. Riedl. 2016. Reward Systems in Human Computation Games. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*. ACM Press. https://doi.org/10.1145/2967934.2968083

[156] * Milad Soroush, Mark Hancock, and Vanessa K. Bohns. 2018. Investigating Game Mechanics that Target Players' Self-Control While Maintaining Engagement. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. ACM Press. https://doi.org/10.1145/3242671.3242698

[157] * Sharon T. Steinemann, Elisa D. Mekler, and Klaus Opwis. 2015. Increasing Donating Behavior Through a Game for Change. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*. ACM Press. https://doi.org/10.1145/2793107.2793125

[158] Denes Szucs and John P.A. Ioannidis. 2017. When null hypothesis significance testing is unsuitable for research: A reassessment. https://doi.org/10.3389/fnhum.2017.00390

[159] * Gustavo F. Tondello, Alberto Mora, and Lennart E. Nacke. 2017. Elements of Gameful Design Emerging from User Preferences. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116627

[160] * Gustavo F. Tondello and Lennart E. Nacke. 2019. Player Characteristics and Video Game Preferences. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3311350.3347185

[161] * Gustavo F. Tondello, Rina R. Wehbe, Rita Orji, Giovanni Ribeiro, and Lennart E. Nacke. 2017. A Framework and Taxonomy of Videogame Playing Preferences. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116629

[162] * Z. O. Toups, Nicole K. Crenshaw, Rina R. Wehbe, Gustavo F. Tondello, and Lennart E. Nacke. 2016. "The Collecting Itself Feels Good". In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*. ACM Press. https://doi.org/10.1145/2967934.2968088

[163] * Olivier Tremblay-Savard, Alexander Butyaev, and Jérôme Waldispühl. 2016. Collaborative Solving in a Human Computing Game Using a Market, Skills and Challenges. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*. ACM Press. https://doi.org/10.1145/2967934.2968104

[164] * April Tyack, Peta Wyeth, and Daniel Johnson. 2016. The Appeal of MOBA Games. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*. ACM Press. https://doi.org/10.1145/2967934.2968098

[165] April Tyack and Elisa D Mekler. 2020. Self-Determination Theory in HCI Games Research: Current Uses and Open Questions. *CHI 2020 - Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020). https://doi.org/10.1145/3313831.3376723

[166] April Tyack, Peta Wyeth, and Madison Klarkowski. 2018. Video game selection procedures for experimental research. In *Conference on Human Factors in Computing Systems - Proceedings*, Vol. 2018-April. Association for Computing Machinery, New York, New York, USA, 1–9. https://doi.org/10.1145/3173574.3173760

[167] * Kellie Vella, Daniel Johnson, and Leanne Hides. 2015. Playing Alone, Playing With Others. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*. ACM Press. https://doi.org/10.1145/2793107.2793118

[168] * Kellie Vella, Christopher James Koren, and Daniel Johnson. 2017. The Impact of Agency and Familiarity in Cooperative Multiplayer Games. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116622

[169] Kellie Vella, Madison Klarkowski, Daniel Johnson, Leanne Hides, and Peta Wyeth. 2016. The social context of video game play: Challenges and strategies. In *DIS 2016 - Proceedings*

of the 2016 ACM Conference on Designing Interactive Systems: Fuse. Association for Computing Machinery, Inc, New York, New York, USA, 761–772. https://doi.org/10.1145/2901790.2901823

[170] Chat Wacharamanotham, Lukas Eisenring, Steve Haroz, and Florian Echtler. 2020. Transparency of CHI Research Artifacts : Results of a Self-Reported Survey. (2020). https://doi.org/10.31219/osf.io/3bu6t

[171] * Guenter Wallner. 2015. Sequential Analysis of Player Behavior. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*. ACM Press. https://doi.org/10.1145/2793107.2793112

[172] * Guenter Wallner and Simone Kriglstein. 2016. Visualizations for Retrospective Analysis of Battles in Team-based Combat Games. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*. ACM Press. https://doi.org/10.1145/2967934.2968093

[173] * Justin D. Weisz, Maryam Ashoori, and Zahra Ashktorab. 2018. Entanglion. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. ACM Press. https://doi.org/10.1145/3242671.3242696

[174] * Lindsay Wells, Aran Cauchi-Saunders, Ian Lewis, Lorenzo Monsif, Benjamin Geelan, and Kristy de Salas. 2016. Mining for Gold (and Platinum). In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*. ACM Press. https://doi.org/10.1145/2967934.2968112

[175] * Matthew Alexander Whitby, Sebastian Deterding, and Ioanna Iacovides. 2019. "One of the baddies all along". In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3311350.3347192

[176] * Laura A. Whitlock, Anne Collins McLaughlin, William Leidheiser, Maribeth Gandy, and Jason C. Allaire. 2014. Know before you go. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play - CHI PLAY '14*. ACM Press. https://doi.org/10.1145/2658537.2658703

[177] Jelte M. Wicherts, Coosje L.S. Veldkamp, Hilde E.M. Augusteijn, Marjan Bakker, Robbie C.M. van Aert, and Marcel A.L.M. van Assen. 2016. Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid P-hacking. *Frontiers in Psychology* 7, NOV (2016), 1–12. https://doi.org/10.3389/fpsyg.2016.01832

[178] * Graham Wilson and Mark McGill. 2018. Violent Video Games in Virtual Reality. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. ACM Press. https://doi.org/10.1145/3242671.3242684

[179] Jacob O. Wobbrock and Julie A Kientz. 2016. Research contribution in human-computer interaction. *interactions* 23, 3 (apr 2016), 38–44. https://doi.org/10.1145/2907069

[180] * Donghee Yvette Wohn, Peter Jough, Peter Eskander, John Scott Siri, Masaho Shimobayashi, and Pradnya Desai. 2019. Understanding Digital Patronage. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3311350.3347160

[181] * Priscilla N.Y. Wong, Jacob M. Rigby, and Duncan P. Brumby. 2017. Game & Watch. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116613

[182] * Daniel Yule, Bonnie MacKay, and Derek Reilly. 2015. Operation Citadel. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*. ACM Press. https://doi.org/10.1145/2793107.2793135

[183] * Anna Zamansky, Dirk van der Linden, Sofya Baskin, and Vitaliya Kononova. 2017. Is My Dog "Playing" Tablet Games?. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17*. ACM Press. https://doi.org/10.1145/3116595.3116634

[184] * Majed Al Zayer, Sam Tregillus, and Eelke Folmer. 2016. PAWdio. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*. ACM Press. https://doi.org/10.1145/2967934.2968079

[185] * David Zendle, Paul Cairns, and Daniel Kudenko. 2015. Higher Graphical Fidelity Decreases Players' Access to Aggressive Concepts in Violent Video Games. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*. ACM Press. https://doi.org/10.1145/2793107.2793113

[186] * Hao Zhang, Qiong Wu, Chunyan Miao, Zhiqi Shen, and Cyril Leung. 2019. Towards Age-friendly Exergame Design. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM. https://doi.org/10.1145/3311350.3347191

* Reference included in literature review.

Paper II

"My Soul Got a Little Bit Cleaner": Art Experience in Videogames

Julia A. Bopp, Jan B. Vornhagen, & Elisa D. Mekler

The paper has been published in the
Proceedings of the ACM on Human-Computer Interaction Vol. 5, pp. 237:1–237:19, 2021.
https://doi.org/10.1145/3474664

The layout has been revised.

Abstract

Videogames receive increasing acclaim as a medium capable of artistic expression, emotional resonance, and even transformative potential. Yet while discussions concerning the status of games as art have a long history in games research, little is known about the player experience (PX) of games as art, their emotional characteristics, and what impact they may have on players. Drawing from Empirical Aesthetics, we surveyed 174 people about whether they had an art experience with videogames and what emotions they experienced. Our findings showcase the prominence of epistemic emotions for videogame art experiences, beyond the negative and mixed emotional responses previously examined, as well as the range of personal impacts such experiences may have. These findings are consistent with art experience phenomena characteristic of other art forms. Moreover, we discuss how our study relates to prior research on emotions and reflection in PX, the importance of games' representational qualities in art experiences, and identify lines of further inquiry. All data, study materials, and analyses are available at https://osf.io/ryvt6/.

II.1 Introduction

Art holds a special role in the human experience [21]: It wields the power to astonish, move, or disturb us [66] — or leave us indifferent [58]. We may have complex and even conflicting opinions on an artwork [40], to the point that art can change our beliefs or even who we are [59, 58, 68]. The game industry has long argued for videogames to be thought of as art, often based on their capacity to afford profound and varied emotional experiences [17, 73, 76]. As early as the 1980s, for example, Electronic *Arts* famously raised the question of whether a computer could make people cry [20]. Later, Sony dubbed the CPU of their then new Playstation 2 *Emotion Engine* in reference to its ability to render faces and emotional expressions in real time [22]. Irked by this label some art critics declared that videogames do not have the capacity to evoke deep emotions, and could therefore not be considered art [37]. This subsequently sparked a series of rebuttals from game scholars, ranging from games being considered a *lively* art that has been unfairly disparaged as pulp [34], likening videogames to theater acting [54], or discussing how some videogames already meet many philosophical and aesthetic definitions of art [73, 49, 76].

Nowadays, few would argue against videogames' artistic potential. Videogames have increasingly attained public recognition as a medium capable of artistic expression and emotional impact, as evidenced by museums like the Smithsonian dedicating exhibitions to "the art of video games" [72], the emergence of art games [67], and the critical acclaim that mainstream titles such as *Journey* [77] and *Hellblade: Senua's Sacrifice* [50] have garnered for their 'Artistic Achievements' [10, 11]. In turn, player experience (PX) research has become increasingly invested in exploring

games' potential to emotionally move and perturb players [e.g., 3, 13, 14, 29, 60], challenge personal convictions [e.g., 5], and even change views and behavior in the outside world [e.g., 84] - responses that are traditionally associated with art [55, 59, 68]. Yet despite this growing body of work, scant attention has been paid to *players' experience* of games as art [cf. 18].

Consequently, PX research currently lacks empirical and conceptual insights into what players consider an "art experience" with videogames, what emotions characterize such experiences, and to what extent their transformative potentials resemble those attributed to other art forms [e.g., 59, 68]. A better understanding of such experiences could help expand our understanding of the emotional spectrum games presently evoke, and what sort of experiences remain as of now untapped [43]. Moreover, due to art's purported power to positively impact individuals and societies [59, 58], studying these experiences may contribute new insights for designing reflective [36, 44] and transformative games [64, 84]. Together, these insights may also help clarify the notion of eudaimonic game experience [14, 19] and contribute to theory-building in player-computer interaction.

To address these gaps, we surveyed 174 participants about a recent "art experience" they had with a videogame. Specifically, drawing from Empirical Aesthetics – a subfield of psychology concerned with the emotional and cognitive processes underlying human experiences with art – we examine the salient emotional responses that characterize the art experience with games, and explore the impacts and downstream effects these experiences have on players.

Our contribution is threefold: First, our results provide a comprehensive view of the range of emotional responses players associate with videogame art experiences, beyond the negative and mixed emotional responses examined in PX research to date [e.g., 3, 5, 29]. For instance, we observed that prototypical aesthetic and epistemic emotions, such as fascination, enchantment and interest were particularly salient. Second, our findings suggest that videogame art experiences often leave a pronounced personal impact – shaping players' views on games as an artistic medium, changing their experience of the game, or altering their self-understanding – extending our understanding of games' potential for transformative reflection [44, 64, 84]. Finally, we contribute a rich qualitative and quantitative dataset – as well as all study materials – for future PX research to use and build towards a more comprehensive basis for empirical game aesthetics [2].

II.2 Related Work

Before we provide an overview of related work, we first clarify what we mean by "videogame art experience". We understand "art experience" in terms of players' aesthetic experience[1], that is, the sensations, emotions, and meaning-making processes that occur during and after gameplay [49]. Art games [67], art created via videogames [67], and games about art [e.g., 61, 62] are not the focus of the present work, although they may undoubtedly also afford aesthetic experiences [21].

Discussions around the status of videogames as art have a long tradition in games scholarship. While popular discourse has revolved around the question whether games can be art [see 53] – sometimes to the point of parody [e.g., 31] – academic discourse has largely moved beyond the simplistic art/not-art dichotomy. Instead, games scholars have been discussing the properties that characterize a videogame as art [76] or "artful" [83]. Notably, some scholars have suggested looking to properties that videogames share with other art forms [2, 49, 76]. Consequently, works have discussed how videogames resemble theater [54], movies [73], or the status of game developers as artists [47]. Most pertinent to the present work, Tavinor [76] draws from existing cluster theories of art [24, 28] that outline a set of conditions which an object *might* meet to be considered art, although meeting these conditions is neither necessary nor sufficient for a game to be considered art. Such conditions include, for instance, a game's capacity for direct sensory and kinaesthetic pleasure, affording or expressing emotion, and/or conveying complex meaning. More recently, Bateman called for "empirical game aesthetics" [2], although his considerations remain mostly restricted to player-satisfaction models and taxonomies of player types, which arguably share few commonalities with the aforementioned cluster theory of art put forth by Tavinor, and do not account for the spectrum of emotional experiences that games may afford.

Indeed, within player-computer interaction, empirical PX studies have become increasingly interested in exploring videogame's potential to afford emotionally complex and reflective experiences. Emotionally moving moments in games, for instance, were found to be particularly characterized by feelings of sadness [3], while uncomfortable game experiences evoked feelings of anxiety, helplessness, guilt, and disgust in players [29]. In both studies, these strong emotional experiences appeared to also act as catalysts for self-reflection and moral contemplation. Curiously, many works originating in player-computer interaction seem to carry the implicit assumption that strong, and particularly, negatively valenced emotional experiences are somehow artistic [see also 43]. Perhaps due to game industry notions which have historically tied games' artistic value to their emotional potential [20]. Cole et al. [13], for instance, argued that the value

[1]Note that in this paper we use the terms "aesthetic experience" and "art experience" interchangeably, following conventions in Empirical Aesthetics [e.g., 58] – but see [71] for a counter-argument.

of “avant-garde“ indie games lies in the novelty, intensity, and quality of the emotional experience they afford. This becomes even more apparent in the study of Bopp et al. [5], where one participant described their emotionally challenging experience as ‘probably one of my earliest memories of seeing true Art (with a capital A) in anything‘" [5, p.7]. However, this implied association between emotional experience and players’ perception of a game as "art" has so far remained unexplored.

The work-in-progress by Craveirinha and Roque [18] describes the first – and to our knowledge, to date only – empirical study explicitly focused on players’ "artistic" experience. They investigated how the absence vs presence of typical, rule-based game elements (e.g, goals, score) affected players’ experience and evaluation of a game’s meaning and aesthetic qualities. However, their results were largely inconclusive, which might be due to the limited statistical power given their sample size, their choice of measures (i.e., the Game Experience Questionnaire [32] and an ad-hoc measure compiled of statements based on Tavinor’s cluster theory of art [76]), or the fact that participants did not get to choose the game themselves. As such, it remains unclear whether players actually encounter videogame experiences they consider "art" in their everyday life, and if they do, what characterizes such experiences.

Outside of games, the field of Empirical Aesthetics has a long history of investigating perceptual, emotional, cognitive and neural processes underlying the art experience [see 58, for an overview]. While a comprehensive account of Empirical Aesthetics lies outside the scope of the present work, we summarize here a few key tenets of particular relevance: First, recent work suggests that the art experience can be broadly classified into five distinct experiential outcomes, each characterized by different emotions and psychological processes [59, 58]. (1) Facile outcomes, where the artwork has no clear relevance to the viewer, resulting in an absence or very subdued emotional response, as well as little to no deliberations of the artwork’s meaning. (2) Artworks that afford surprise and insight experiences due to confronting viewers with novel, albeit personally non-relevant content. (3) Harmonious and emotionally moving experiences, where the content of an artwork resonates with the viewer’s self-image and value, but does not require much cognitive effort. Lastly, if an artwork is of high relevance yet potentially threatening to the viewer’s sense of self (e.g., challenging their worldviews), it (4) often triggers a so-called abort experience, characterized by anger, anxiety, and the desire to discontinue the experience. However, such experiences may also result in (5) transformation, where the viewer confronts their feelings of discrepancy, gains new self-relevant insights, and subsequently changes a negative experience into a positive one. Transformative art experiences also have the potential to lastingly shape viewers’ attitudes and behavior, long after they have encountered the artwork [59, 58].

Second, while the range of emotions that may accompany or follow an art experience is vast

[66], each experiential art outcome outlined above tends to be characterized by specific emotional responses. For example, it has been suggested that transformative art experiences are particularly characterized by feelings of awe and a sense of connectedness [58], or feeling like crying upon coming to new self-relevant insights [59]. Moreover, artworks depicting self-relevant content typically afford more intense emotional responses from viewers [56]. As such, Empirical Aesthetics provides a potentially useful conceptual basis to explore how videogame art experiences relate to players' emotional responses, as well as how this relates to the transformative impact games may have.

II.3 Method

Our aim was to explore the art experience of videogames. Specifically, we address two research questions (RQs): (RQ1) What are the salient emotional responses that characterize art experiences with videogames?; (RQ2) How do art experiences impact players? To do so, we draw from previous studies in PX [e.g., 5] and Empirical Aesthetics research [16, 57], and employed a mixed method approach, collecting both qualitative and quantitative data.

All study materials (complete survey, data, analyses) are available at https://osf.io/ryvt6/

II.3.1 Participants

We aimed for a sample size of 200, which is comparable to other exploratory survey studies in HCI games research [e.g., 5, 4]. A total of 655 responses were collected over three weeks in September 2020, of which 183 completed at least the main survey (see Chapter II.3.2 for details). We then excluded nine responses from participants who had not provided informed consent (5) or were under the age of 18 (4). Given the richness of the qualitative data [8, 42], we decided to stop recruiting after a final sample size of $N = 174$[2], which is comparable to previous mixed-method online surveys in PX research [e.g., 3, 29, 84].

[2]Note that following related work in Empirical Aesthetics [15, 57], we originally intended to perform a Latent Class Analysis to identify different types of art experience. To do so, we first conducted a inductive reflexive thematic analysis [7, 9] to explore why participants thought the experience was art. However, after data collection we noted that our sample size was below recommended thresholds for the amount of potential clusters we identified in the thematic analysis (with variations in their sizes), and the potential clusters not being well-separated as the themes were not mutually exclusive [51]. Therefore, we refrained from conducting the Latent Class Analysis. For the sake of readability, we also omit the findings of the thematic analysis from the manuscript. That said, the statistical script and results of the preliminary Latent Class Analysis, as well as the thematic maps and documentations of the thematic analysis are available in the OSF repository.

Participants were aged 18 to 57 ($m = 26.43, sd = 7.61$), 105 identified as men, 51 as women, 12 as non-binary, 4 participants self-reported, and 2 preferred not to disclose their gender.

Overall, participants rated themselves as highly experienced players ($m = 78.72, sd = 22.12$), having played games for $m = 18.1$ years ($sd = 8; Min, Max[0, 40]$) and playing $m = 14.71$ hours per week ($sd = 13.01; Min, Max[0, 80]$) on average. The four most popular game genres were adventure (140), RPGs (126), strategy (111) and puzzle games (100). Among participants' favorite games were *The Elder Scrolls 5: Skyrim* (25), *The Legend of Zelda: Breath of the Wild* (22), *Minecraft* (18), and *The Witcher 3: Wild Hunt* (14).

Nine participants indicated that their profession was or had at some point been related to games in some way (e.g., as programmer, designer, critic or academic). Thirty-three participants had professional experience in art-related domains (e.g., artist, curator, critic), and 16 were or had been active in professions that bridged arts and games.

II.3.2 Survey Design and Procedure

The study protocol was approved by the research ethics council of the authors' university. The online survey was prepared and conducted in LimeSurvey (v. 3.16.1+190314; [41]). The complete survey materials, including the exact wording of the questions, are available in the OSF repository.

Participants were recruited via social media, including Twitter, snowball sampling, and via the Reddit r/samplesize and r/arthistory subreddits. The survey was advertised as "did you ever experienced a videogame as art?", to recruit both participants who had and have *not* had an art experience involving games. On average, the survey took 23.4 minutes ($SD = 13.99, Min, Max[3.43, 81.32]$) to complete. Participants did not receive any compensation for completing the survey.

Upon clicking the survey link, participants were introduced to the study and asked for consent. After providing informed consent, they were asked to provide demographic information, indicate their English skill level, as well as whether they ever experienced a videogame as art. Six participants had no such experience to report, and were subsequently asked to describe an art experience involving another medium (if any), as well as elaborate on whether they thought digital games could be art. The survey concluded thereafter.

The remainder of the study procedure pertains to the 168 participants who indicated to have experienced a game as art. Following a guided recall process, participants were first asked to recount their most recent games as art experience, explain why they considered it "art", as well as indicate how long ago the experience had taken place. No definition of "art" was provided, as we were interested in participants' own understanding of "art".

Next, participants rated their experience in terms of their emotional responses (see Chapter II.3.3). Finally, they were asked to provide information on their gaming habits and preferences, as well as whether their profession was related to games and/or arts. This was followed by a series of optional open-ended questions about the impact of participants' gaming experience. These questions were adapted from previous studies in Empirical Aesthetics [57] and user experience research [79], respectively. Specifically, participants were asked to describe what the meaning of their experience was, and whether it had changed their perspective on how they viewed the game, their opinion on digital games as a medium, as well as whether it had impacted them personally in any way. Finally, participants were asked whether they thought the experience was intended by the developer. As pilot testing suggested the survey to be rather time-consuming and taxing for participants, we opted to leave this part of the survey optional. In total, 140 participants completed the impact questions.

II.3.3 AESTHEMOS

Given the central role of emotions in player experience [33, 73] and art experience research [66, 46, 70], we employed the Aesthetic Emotions Scale (AESTHEMOS). The AESTHEMOS was developed and validated by Schindler et al. [66] to account for the variety of emotional responses to different aesthetic artefacts (e.g., music, painting, and architecture). We decided to use the AESTHEMOS over other scales previously used in PX research on emotional gaming experiences (e.g., the appreciation scale [52] used in [3, 5, 60], which assesses to what extent experiences are perceived as moving and thought-provoking), because it provides a more comprehensive and granular measure of aesthetic experience. Moreover, the AESTHEMOS allows us to situate whether participants' experiences correspond to art experience outcomes previously discussed in Empirical Aesthetics [58].

Specifically, the questionnaire consists of 42 items compiled into 21 subscales and organized into seven superfactors (Negative Emotions, Prototypical Aesthetic Emotions, Epistemic Emotions, Animation, Nostalgia/Relaxation, Sadness, and Amusement). Example items include "I found it beautiful", "Was enchanted", "Felt deeply moved", "Felt a sudden insight", "Was unsettling to me". Statements were rated on a 5-point scale ranging from 1 ("not at all") to 5 ("very"). The full questionnaire and reliability scores (Cronbach's α) for each subscale are listed in Table II.3 and the OSF repository.

II.3.4 Qualitative Content Analysis

To examine in what ways the reported experiences had impacted players (RQ2), we conducted a qualitative content analysis over all open-ended answers [35]. We chose this approach over other qualitative analysis approaches (e.g., thematic analysis), as we could draw from existing work in Empirical Aesthetics [27, 57] on the effects of art experiences to generate an initial code book.

Two of the authors applied the initial set of codes to a randomly selected set of 60 experiences[3], after which we expanded the code book to cover additional aspects (e.g., meaning of the game's content/message/interpretation, lasting impression). Some of the additional codes were generated based on the data set, whereas others were derived from Bopp et al.'s [3] code book on emotionally moving game experiences. The finalized code book consisted of 15 categories, further divided into 36 subcategories. One author then applied the final set of codes to the entire data set (n=168), including the optional questions. Sentences formed the smallest coding unit, which could be assigned several codes.

Opinions on the need to assess interrater agreement reliability vary, especially in case a coding unit (here a sentence) can be assigned multiple codes and when two coders do not have comparable levels of experience with coding [12]. We opted for percentage agreement as an index for interrater reliability as it is deemed suitable for both natural text segments and coders of different knowledge levels [12, 39], and thus lends itself well to our exploratory research aim. As recommended by Campbell et al. [12], we calculated the interrater reliability twice, once after the random coding of 20 experiences (range 60% - 100%; $Mean = 0.83; SD = 0.13$) and once after resolving disagreements through discussion, revising the code book, and re-coding of the same 20 experiences (range 65% - 100%; $Mean = 0.87; SD = 0.12$; the agreement percentage of each codes can be found in the OSF repository).

Next, the subcodes *insight, meaning of the game's content/message/interpretation*, and *healing experience* were split further into subsubcodes for more in-depth analysis. For instance, the subcode meaning of the game's content/message/interpretation was divided into thoughts on the game, game genre, games as a medium, the games industry, and game developers (see the OSF repository).

Note that responses from the six participants who did not report an art experience with games were excluded from the qualitative content analysis.

[3]Each of the two first authors coded a set of 40 experiences: the same 20 experiences were coded by both authors to ensure interrater reliability; Two separate sets of different 20 experiences were coded by each authors to look out for aspects not covered by the initial code book. This resulted in total of 60 experiences analysed (i.e., 20 by both, 20 by coder 1, 20 by coder 2).

II.4 Results

The largest proportion of reported experiences took place within 2 months before study participation (54 of 168). A few participants recounted their first art experience involving games, instead of their most recent one, reasoning that they "can remember it much more vividly than anything i watched or played in the recent past" (ID137, *Grand Theft Auto: Vice City*). While participants' experiences covered a wide variety of games and genres, *The Legend of Zelda: Breath Of The Wild* ($n = 8$), *Journey* ($n = 7$), *NIER: Automata* ($n = 6$) and *Red Dead Redemption 2* ($n = 6$) were among the most frequently mentioned games.

When describing their art experience, the 25 participants with professions related to games tended to make note of the expressive and thought-provoking potential of games. In contrast, participants with no background in games, emphasized the beauty and emotional intensity of their experience, as well as praised the game developers' artisanship. However, given the small sample size of participants with a professional background in games, we decided not to pursue analysis further (but refer to the OSF repository for more detail).

In the following, we report on the emotional responses that constituted art experiences with videogames (RQ1), as well as how these experiences impacted players (RQ2). We do not report on the content of the individual experiences, as these are closely intertwined and somewhat redundant with RQ2. Again, we refer interested readers to our OSF repository, which contains all conducted analyses, descriptive statistics, and data.

Illustrative participant quotes are kept in their original wording, where brackets contain the unique ID numbers assigned to each participant[4]. All quotes are printed verbatim, including grammar and spelling unless denoted otherwise.

II.4.1 Emotional Responses

To address RQ1, we examined participants' emotional responses via their AESTHEMOS ratings. Due to the exploratory nature of our study, we refrain from computing hypothesis tests [81] and instead report descriptive statistics (means M, pooled standard deviations pSD[5]) and confidence intervals $CIlow$ and $CIhigh$ [23]. We also refer readers to the OSF repository for violin plots of the individual items and emotion subscales, which visualize the distribution of the AESTHEMOS items more clearly.

[4]Note that ID numbers range non-continuously from 8 to 674. LimeSurvey assigns an ID number to each person clicking the survey link, which is why ID numbers do not match the actual number of participants.

[5]combined standard deviation (SD) of the two items per subscale

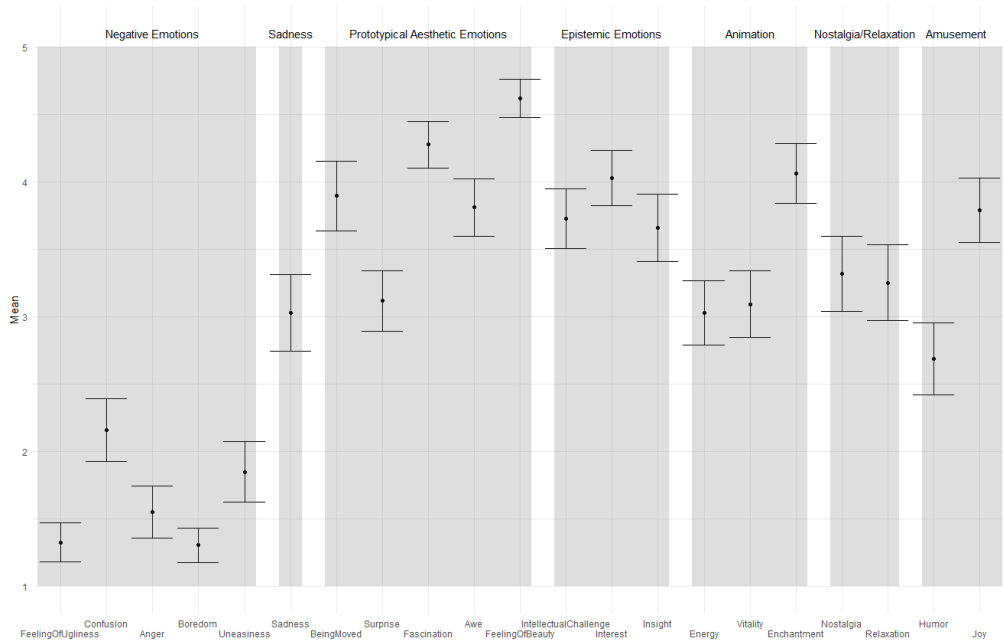

Fig. II.1: Means and Bonferroni corrected 95% CIs for 21 CI ($z = 2.82$ [6]) over all participants who reported an art experience with games ($n = 168$).

Players rated their experiences rather highly on prototypical aesthetic emotions (see Figure II.1). Overall, the Feeling of Beauty (which included Liking) had the highest mean score ($M = 4.62; pSD = 0.75$) and second smallest distribution ($CIlow = 4.48; CIhigh = 4.76$, see also Table II.1), indicating that *most* participants rated their experience as very beautiful. Participants also reported rather pronounced feelings of fascination, awe, and being moved during their experience – although the latter exhibited a broader distribution, indicating that these emotionally moving experiences varied to a greater extent.

Participants scored their experiences also rather highly in terms of Epistemic Emotions, rating their experiences as interesting ($M = 4.03; pSD = 1.1$), intellectually challenging ($M = 3.73; pSD = 1.01$) and insightful ($M = 3.66; pSD = 1.3$).

Art experiences scored relatively highly on feelings of Enchantment ($M = 4.06; pSD = 1.15$), whereas Nostalgia ($M = 3.32; pSD = 1.44$), Relaxation ($M = 3.25; pSD = 1.4$), Energy ($M = 3.25; pSD = 1.4$), and Vitality ($M = 3.09; pSD = 1.32$) were less pronounced. With regards to Amusement emotions, experiences scored rather low on Humor ($M = 2.69; pSD =$

Table II.1: Descriptive statistics of all 21 AESTHEMOS subscales, including the Mean, pooled standard deviation (pSD), confidence interval limit low (CIlow) and high (CIhigh). For the descriptives of each individual item, please refer to Table II.3 and the OSF repository.

Subscale	Mean	pSD	CI low	CI high	Subscale	Mean	pSD	CI low	CI high
Feeling of Ugliness	1.33	0.67	1.18	1.47	Intellectual Challenge	3.73	1.01	3.51	3.95
Confusion	2.16	1.08	1.92	2.39	Interest	4.03	0.94	3.82	4.23
Anger	1.55	0.9	1.36	1.75	Insight	3.66	1.15	3.41	3.91
Boredom	1.31	0.59	1.18	1.43	Energy	3.03	1.1	2.79	3.27
Uneasiness	1.85	1.04	1.62	2.07	Vitality	3.09	1.13	2.85	3.34
Sadness	3.03	1.3	2.74	3.31	Enchantment	4.06	1.03	3.84	4.28
Being Moved	3.9	1.19	3.64	4.15	Nostalgia	3.32	1.29	3.04	3.6
Surprise	3.12	1.03	2.89	3.34	Relaxation	3.25	1.3	2.97	3.53
Fascination	4.27	0.8	4.1	4.45	Humor	2.69	1.22	2.42	2.95
Awe	3.81	0.98	3.6	4.02	Joy	3.79	1.1	3.55	4.03
Feeling of Beauty	4.62	0.65	4.48	4.76					

1.34), whereas most participants reported average to strong feelings of Joy ($M = 3.8; pSD = 1.21$).

Overall, low scores were observed for *negative emotions*: Boredom ($M = 1.31; (pSD) = 0.74$) and feelings of Ugliness ($M = 1.327; pSD = 0.81$) scored lowest. In contrast, sadness was almost uniformly distributed over the scale ($M = 3.03; pSD = 1.46$; note the violin plot in the OSF repository), indicating that for some participants sadness was an essential part of the experience, while for others it hardly featured.

II.4.2 Impact of Game Experience

To answer RQ2, we first examined how many participants had indicated whether the experience had changed their perspective on the game, themselves, or games as a medium. Following this, we analyzed the entirety of open-ended answers via qualitative content analysis to further specify the various impacts the reported experiences had on participants.

Overall, most participants stated that the experience had changed their perspective in some way (see Table II.2). The largest proportion indicated that the experience had changed their opinion of the game, usually for the better (e.g., "Before I played through the story, I thought it was just a violent shooter game, but I realized how beautiful, artistic, and complex it really is." ID37; Red Dead Redemption 2). A substantial number of participants also indicated that the experience had impacted their self in some way (92/140), whereas less than half noted a change in perspective on games as a medium. Note, however, that many participants who reported no changes on this point already considered games as a form of art (e.g., ID8, ID12, ID67, ID87, ID116).

In the following, we report the various impacts of the recounted experiences in more detail.

Table II.2: Absolute and relative frequencies of participants who noted a perspective change. Note that $n = 140$, as these questions were optional in the survey.

Changed Perspective on Game:	Changed Perspective on Self:	Changed Perspective on Medium:
120 (85.71%)	92 (65.71%)	61 (43.57%)

Emotional Resonance and Healing.

In line with the above findings, participants emphasized the emotional impact of their experience, describing how it resonated for hours and days after playing (n=23): "I felt the feelings from this game for the next day at least, as the emotions were particularly powerful." (ID194, *Dear Esther*). This emotional resonance was deeply intertwined with participants' memories, "warm(ing) my heart [...] whenever I remember the game." (ID266, *Unravel Two*), even "Though I may not remember the specifics of Oxenfree, nor how it ended, I vividly recall what I felt." (ID155, *Oxenfree*).

Several participants (n=14) highlighted the healing and relaxing impact of their experience, which provided "a few moments of pure bliss" (ID627, *Red Dead Redemption 2*), or "affects me in that it makes me feel better about myself" (ID348, *Animal Crossing*).

In these instances, participants typically praised the visuals and beauty of the game world: "There was something very serene and beautiful about being elevated and getting to see the rest of this particular level of the game." (ID205, *Sekiro: Shadows Die Twice*), and would revisit the game "just to wander around and enjoy the atmosphere even though I've already completed the main story " (ID321, *The Legend of Zelda: Breath of the Wild*). However, a few participants emphasized healing qualities of the game narrative, such as ID367 who " kept quotes from the game in my everyday life [...] (which) I remind myself when my mental health begins to fade." (*The Last of Us*).

Thoughts on the Game and its Developers.

Several participants (n=38) reported how the experience made them ponder the meaning of the game's narrative or its intended message: "I think it was trying to communicate that existence, particularly human existence as a sentient and intelligent life form, is terrifying but beautiful nonetheless." (ID398, *Disco Elysium*). These interpretations of a game's possible meanings were often linked to it being likened to art, as in the case of one participant's experience with *Shadow of the Colossus*, where "winning does not feel like victory. I think that is exactly what the game tries to tell to the players. And that made me feel the gaming experience like art" (ID512).

Moreover, the recounted experiences sometimes altered participants' impression of the game (n=38): "I started off thinking it was just an aesthetically pleasing metroidvania. Now I view it as

a beautiful, thought-provoking piece of art." (ID75, *Hollow Knight*). These realizations sometimes even shaped the participant's playstyle: "I started off playing to destroy things, but I started wanting to just admire those things instead, and I eventually wanted to create more of those things, although I couldn't, which was frustrating." (ID174, *Grand Theft Auto V*).

A few participants also commented how the experience made them feel a sense of connection with the developers, "that I and the developers share something, a common viewpoint or a lived experience or something similar." (ID87, *In the Pause Between the Ringing*). Others speculated on the developers' approach to the game's production: "I believe the developer Toby Fox had a vision of creating a feelgood experience and being a one man developer team; was able to fulfill this vision to an accuracy that would be difficult for teams of a larger size." (ID557, *Undertale*).

Changed Playstyles and Views on other Games.

Participants' art experience also colored their approach to other games (n=14), with changes in playstyle carrying over into other games: "it helped me realise that the personal journey with a game is the most important thing when playing it. I used to focus more on getting things right in a game, getting the perfect ending but I've done a complete one-eighty now. I just want my story to be mine." (ID563, *Life is Strange*).

This increased awareness of the artistic potential of games made participants "expect a higher standard from the games I play." (ID222, *The Witcher 3: Wild Hunt*). Participants now "enjoy being more critical and questioning of works and stories." (ID451, *NieR: Automata*). However, one participant also remarked that "this game made me have higher expectations about similar videogames, making it harden to enjoy them" (ID152, Call of Duty: Black Ops).

Views on Games as a Medium.

Many participants described how their experience changed their view on games as an artistic medium (n=61); for instance, "open(ing) my mind to experience more games as art." (ID66, *Rain*). Emphasis often lay on how the game in question pushed the boundaries of the medium.

For instance, the storytelling of *NieR: Automata* was described as "extremely innovative [...] and thus matur(ing) the entire medium." (ID451, *NieR: Automata*).

A participant further explained how their experience heightened their appreciation for the craft of game development: "It opened my eyes to game developers being able to massively influence a person's opinion through thoroughly crafted storytelling. It's so subtle that most people who play the games wouldn't even realize it." (ID389, *The Elder Scrolls Online*).

Yet, despite a general consensus among participants on games' status as an artistic medium, some participants remained skeptical of how others perceive games: "Video games aren't treated

as what they are, as art. People see video games as some mindless stupid thing for kids to enjoy. But really it can be deeper than that. Which unfortunately makes us surprised more often than not." (ID674, *Portal 2*).

Sparking Interest for Non-Game Topics.

Several participants explained how their experience had sparked their curiosity in topics outside of games, but which the game had touched upon (n=12), including Chinese history, Carl Jung's psychological theories, science fiction stories, or "motivated me to become to read more English works" (ID431, *Photopia*) when English was not their native tongue. Others became inspired to engage in artistic activities themselves; for instance, citing the music in *The Legend of Zelda: Ocarina of Time* as the reason they "[d]eveloped an interest in composition (music)" (ID584).

Better Self-Understanding.

Several participants noted how the experience lead to a better understanding of their self (n=30); for instance, describing it as "having a mirror held up, the game was an experience to vent/explore those feelings that I had otherwise repressed." (ID301, *Firewatch*). Similarly, one participant emphasized how *Oxenfree* invited exploration of their feelings, which "helped me better understand my emotions." (ID155) and realize that "it is about better understanding myself and how my emotional side sees the world.".

Participants also explained how the experience made them reflect on current and past life experiences: "Through the whole path, I was locked in a reflective state on what made me 'me', and what doesn't. [...] When I walked through the door to say I was a loner, I really did pause and think on how, despite past anguish, Id chosen that reality myself. I can't change that, but there are upsides to the choice." (ID537, *The Secret World Legends*).

Others explained how the experience brought to the fore dearly held values: "I think my emotional experience of that game stemmed from my appreciation for continued companionship and emotional commitment in the face of strife and danger. The game depicted the beauty of that interdependence perfectly and gave me a sense of profound warmth and tenderness. It made me realize just how much I appreciate those things." (ID266, *Unravel Two*).

Personal Transformation.

Finally, a few participants described how the experience had changed them personally in some way (n=17). Some explained that the experience had contributed to their becoming a "better" or more authentic version of themselves – that "for a while at least, I felt like my soul got a

little bit cleaner." (ID512, *Shadow of the Colossus*). For instance, playing through *Portal 2* made ID674 realize that "the way I was before. I just masked it. [...] I became more like myself in the end, and much like progression in video games more of me becomes open and unlocked for more I play. I become more of who I am." (ID674, *Portal 2*). Others considered themselves "a slightly softer person now, with more understanding and appreciation for myself." (ID667, *Ori and the Blind Forest*), and "slightly more likely to express affection and show those close to me that I love them." (ID206, *Legend of Zelda: Breath of The Wild*). For some, this heightened sense of self-understanding also related to increased self-acceptance, when the experience "made me [...] slightly more forgiving for my own mental health. It showed me that strength can still be found it this 'perceived weakness'." (ID393, *Hellblade: Senua's Sacrifice*).

Finally, several participants explained how their experience allowed them to be more comfortable and accepting in the light of suffering:

> "For me personally, it meant the world to me – i had just lost a close family member and was stuck in a really dark place, pretty much having given up on any hope of positive change. That is why i felt so connected to these characters who were stuck in a grim and hopeless world, in an endless loop of struggle and death. When in the end, these characters all survived, thanks to the actions of another player – in a strictly singleplayer game with no online functionality what so ever outside of this experience – i, too, felt that there was hope. Hope for humanity as a whole, hope for better days. Hope for myself." (ID232, *NieR: Automata*).

Negative Impact.

Five participants reported how the experience had impacted them negatively in some way. For instance, ID294 recounted that playing *What Remains of Edith Finch* impaired their mental health, as they were not "mentally prepared to play" and found "the game off putting because of it".

Others reported more conflicting experiences, for example, stating that "The Last of Us 2 was art in that it was engaging, thought-provoking, very distinct and often unenjoyable – a work of art that I appreciate aesthetically yet also find numerous flaws with" (ID609, *The Last of Us 2*). The same participant further added that they "felt indifferent after finishing it". In another example, ID500 described how " (*NieR: Automata*) annoyed me overall and I didn't quite get why so many people recommended it so I looked up some articles/podcasts afters to try and understand this more. Overall I found the game rather pretentious and difficult to understand". Yet "even if I didn't like it too much I still see it as an example art due to the amount of creativity energy and thought that went into it – plus it attempted to do more than 'just' entertain". Nevertheless, ID500 also remarked that they "might be more likely to avoid similar titles!" in the future.

II.5 Discussion

In the last decades many debates have revolved around whether videogames could be considered "art". These arguments are rooted in game industry visions [20], art criticism [37, 76], as well as philosophical accounts of aesthetics [2, 69]. In contrast, PX research has to date largely overlooked the videogame "art experience" [but see 18], despite a growing interest in games' capacity to afford emotionally complex and reflective experiences [e.g., 3, 29, 84]. We set out to explore the types of game experiences that players identify as "art", the emotions that characterize them, and how these experiences impact players. Indeed, our findings indicate that – at least when prompted – people can easily recall and recount game experiences they consider art. As such, our study provides the first empirical account of the videogame art experience, as encountered by players in their everyday lives. In the following, we discuss our most notable findings and their implications.

One feature typically ascribed to art [24] – as well as notions of games as art [49, 76] – is the capacity to engender emotions in people. In our study, we found that prototypical aesthetic (e.g., feeling of beauty, fascination) and epistemic emotions (e.g., interest, intellectual challenge) were among the most salient emotions characterizing videogame art experiences. As indicated by the open-ended responses, participants valued games for the direct sensory pleasure they provided [49, 76], and for intellectual engagement with a game's meaning [76], further suggesting that reflection can form an important part of the player experience [44, 84]. In line with previous work on emotional gaming experiences [3, 43, 84], our findings further showcase that the feelings afforded by the experience seem to resonate and linger with players – sometimes long after playing.

Note that the AESTHEMOS subscale for feeling of beauty includes the item "I liked (the experience)" (see Table II.3). While this may be considered a potential confound of the beauty measure, it also suggests that participants overall value videogame art experiences. Relatedly, emotions associated with negative art experiences (i.e., abort outcomes [58]), such as ugliness, anger, and boredom were scored low, suggesting that the reported art experiences were considered positive by most participants.

While the above findings are not unexpected, as they correspond to emotional responses characteristic of harmonious and transformative art engagement [58], we were surprised that negative emotions and sadness did not figure more prominently. While emotionally challenging [5, 13, 60] and serious game experiences [43] have been linked to notions of art [5] and the avant-garde [13], sadness or discomfort do not seem inextricably linked to videogame art experiences [29]. Conversely, while epistemic emotions are core to Empirical Aesthetics [59, 58, 66], they have received relatively little direct attention within PX research – but refer to [38, 65, 78] for

investigations into curiosity, a well-established aesthetic emotion [25]. The high scores in our study suggest that epistemic emotions may be a particularly interesting avenue for PX research, both to gain a better understanding of the videogame art experience, and player experience more generally. For instance, while curiosity has been linked to certain types of uncertainty [38], intellectual challenge is likely central to games that demand active interpretation from players [1, 14], and insight may be more indicative of transformative reflection [44, 84].

Moreover, our results suggest that videogame art experiences appear to have a pronounced impact on players. The majority of participants reported how their experience made them reconsider their opinion on the game, or come to a sudden realization about the game's themes and meaning. These instances seem to correspond to the insight experiences observed in Empirical Aesthetics [56], as well as notions of dialogical and endo-transformative reflection in games [44, 84].

Notably, while previous works on reflection and perspective-challenging moments in games observed few instances of exo-transformative reflection [44, 84] – moments of transformative reflection that affect players' beliefs or actions outside play [84, p. 342] – our findings suggest that videogame art experiences appear to hold considerable transformative potential. Similar to studies on mixed and negative emotional game experiences [3, 5, 29], our participants reported how their experiences led them to ponder matters of identity and self-understanding, explore their feelings, or even simply reconsider their playstyle and approach to games as a medium. Again, these findings are in line with our expectations, as these contemplations are characteristic of art experiences with other media [27, 68]. Of course, not all emotional experiences provoke deep reflection, nor need reflection and insight be tied to strong emotional experiences, whether in the traditional arts [58] or in games [5, 44, 84]. Nevertheless, together with previous work [3, 5, 29], our study suggests that the emotionality of player experience is, to some extent, linked to videogames' transformative potential [64]. As such, models from Empirical Aesthetics [e.g., 59, 58] may answer calls towards a research agenda for "Empirical Game Aesthetics" [2] and be productively applied to player-computer interaction, to identify psychological mechanisms and facilitating circumstances in which emotions experienced during play may more effectively support transformative reflection.

II.5.1 Limitations and Future Work

First, the reported art experiences with videogames should not be seen as definitive or exhaustive accounts. For instance, participants' accounts concerned mainly representational aspects of videogames (i.e., audio, visual elements, and narrative). Few participants emphasized a game's mechanics or gameplay. This may be due to our participants' associations with the term "art",

in that they may have recalled experiences that *resemble* art experiences with other media, or conventional notions of the art experience. Indeed, most experiences described in our study seem to correspond to the harmonious and transformative art experiences detailed in Empirical Aesthetics research [58]. As such, the present study was also unlikely to find purely uncomfortable gaming experiences, such as the ones reported by Gowler and Iacovides [29], which may correspond to the abort outcomes discussed in Empirical Aesthetics [58].

Relatedly, our findings featured relatively many accounts of exo-transformation, compared to previous studies on reflective and perspective-changing player experiences [44, 84]. This may be due to the wording of our survey questions (i.e., directly asking about different impacts) having rendered such instances particularly salient to participants. Conversely, given the prominence of transformative experiences in the arts [58], exo-transformation may be particularly characteristic of videogame art experiences, especially when accompanied by epistemic emotions [58].

Second, while our findings primarily concern representational aspects of videogames, game scholars have pointed towards the aesthetics of agency [48] as well as the kinaesthetic pleasures of videogame play [75, 76], "the qualities of the physical interaction with the gaming device and the physical world it depicts" [76, p. 181]. It seems evident that kinaesthetic pleasure is often central to players' aesthetic experiences (e.g., when navigating Link through the landscapes of *The Legend of Zelda: Breath of the Wild*) and poses a promising avenue for future work.

Conversely, previous works have suggested that game rules (e.g., goals, win states) and dexterity-based gameplay may be antagonistic or even incompatible with the art experience, as they distract players from a game's aesthetic qualities [13, 18]. As noted, few participants in our study explicitly mentioned game mechanics, and one participant (ID82) even mentioned regularly playing a game for its aesthetic qualities, despite disliking the gameplay. More work is needed to examine how games' unique qualities shape the art experience and players' understanding thereof. While Empirical Aesthetics does not provide conceptual or methodological paradigms that readily apply to the functional properties of games, the AESTHEMOS [66] may nevertheless prove useful to evaluate how game mechanics affect "players' perception of a video game's artistic value" [18, p.3].

Third, while confusion ratings were overall low, they raise interesting questions for future work. Silvia [70] argues that confusion is an indicator of an "unsuccessful" art experience, where the perceiver fails to draw meaning from an ambiguous stimulus, whereas interest arises from successful meaning-making. Likewise, Pelowski [59] argues that confusion is a sign of experienced discrepancy that can either result in an abort outcome or a transformative experience, pending resolution or acceptance of this confusion. Consequently, the role of confusion in the player experience could be of interest for both PX researchers and game designers looking to

elicit transformative experiences [63, 64].

The relation between sadness and being moved poses another promising avenue for future work. Past research has linked the two emotions for both experiences with videogames [3, 19], and with other media [30, 82, 45]. In our data, however, participants reported all possible values of sadness, while being moved was consistently rated highly, indicating its greater salience during the experience (see Figure II.1). Moreover, while the mean and standard deviation of sadness in our sample is rather unremarkable, the violin plot demonstrates its almost uniform distribution (see the OSF repository). This finding suggests that sadness and being moved are not always equivalent in the context of videogame art experiences, as well as that for some videogame art experiences sadness was an important factor. The concept of Kama Muta – a form of being moved characterized by an increased sudden awareness of closeness towards or among other beings [26] – may be instrumental to explicating the sadness-moved relation. Further studies on the interrelation of aesthetic emotions and the game characteristics that shape them would extend our understanding of emotionally complex player experiences [3, 29].

Fourth, our ancillary analyses showed that players with different professional backgrounds seem to focus on different aspects when reporting their art experience. In Empirical Aesthetic, the viewer's expertise and interest in art-related topics and art-making play a significant role in determining their art experiences [e.g., 74]. It remains to be seen to what extent players with differing expertise actually *experience* games differently, or whether their familiarity with gaming tropes [80] primarily shapes how they describe and verbalize their art experiences with games. Interviews might be particularly suited to gain more in-depth insights into how players of varying backgrounds articulate art experiences with games, and help clarify the role of expertise.

Fifth, given the present study's exploratory approach, no conclusions may be drawn with regards to whether and to what extent videogame art experiences *cause* changes in players' perspectives, opinions, or self-understanding. Research on the socio-epistemic impact of art [see 68, for an overview] may benefit future PX research aiming to identify potential causal mechanisms.

Finally, our dataset might yield additional insight into the player experience. We invite readers to explore our data set in the supplementary material and hope that games researchers will find the open data and materials useful for their own future works.

II.6 Conclusion

The status of videogames as art has been a point of contention. While many discussions have revolved around why videogames should or could be thought of as art, whether players ever experience games as art has not been empirically examined yet. In this paper, we introduce

Empirical Aesthetics as a paradigm to help expand our understanding of games not only as entertainment products, but as an art form capable of affording the entire spectrum of aesthetic emotions. We find that game "art experiences", as reported by our participants, are accompanied by aesthetic and epistemic emotions, which have so far not been considered in PX research. Moreover, our findings suggest that art experiences with videogames appear to bear considerable transformative potential.

Data Availability Statement

All data and analysis can be found at https://osf.io/ryvt6/.

Author Contribution

EDM and JBV conceptualized the study. JBV designed and constructed the study. JAB and EDM recruited participants. JAB and JBV conceptualized and conducted the Qualitative Content Analysis. JAB analysed the qualitative results; JBV analysed the quantitative results. JBV and JAB wrote the first draft. All authors contributed to writing the final version.

Acknowledgements

Special thanks to Barbara Keller and Roosa Piitulainen for contributing additional explorations of the dataset. We also thank April Tyack for input on the text and concepts of this paper. Jan Vornhagen was supported by Aalto University Department of Computer Science internal funding.

References

[1] Jonne Arjoranta. 2018. Interpretive Challenges in Games. In *Proceedings of Digital Games Research Association (DiGRA) 2018*. SocArXiv, 1–4.

[2] Chris Bateman. 2014. Empirical game aesthetics. *Handbook of digital games* (2014), 411–443. https://doi.org/10.1002/9781118796443.ch15

[3] Julia Ayumi Bopp, Elisa D Mekler, and Klaus Opwis. 2016. Negative Emotion, Positive Experience?: Emotionally Moving Moments in Digital Games. In *Proceedings of the 2016*

CHI Conference on Human Factors in Computing Systems. ACM, 2996–3006. https://doi.org/10.1145/2858036.2858227

[4] Julia Ayumi Bopp, Livia J. Müller, Lena Fanya Aeschbach, Klaus Opwis, and Elisa D. Mekler. 2019. Exploring Emotional Attachment to Game Characters. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Barcelona, Spain) *(CHI PLAY '19)*. Association for Computing Machinery, New York, NY, USA, 313–324. https://doi.org/10.1145/3311350.3347169

[5] Julia Ayumi Bopp, Klaus Opwis, and Elisa D Mekler. 2018. "An Odd Kind of Pleasure": Differentiating Emotional Challenge in Digital Games. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 41. https://doi.org/10.1145/3173574.3173615

[6] Jürgen Bortz and Nicola Döring. 2007. *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler: Limitierte Sonderausgabe* (4th ed.). Springer-Verlag. Google-Books-ID: 13GbPUYAUHsC.

[7] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[8] Virginia Braun and Victoria Clarke. 2019. To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative Research in Sport, Exercise and Health* (2019), 1–16. https://doi.org/10.1080/2159676X.2019.1704846

[9] Virginia Braun and Victoria Clarke. 2020. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology* (2020), 1–25. https://doi.org/10.1080/14780887.2020.1769238

[10] British Academy of Film and Television Arts. 2013. Games – Artistic Achievement in 2013. http://awards.bafta.org/award/2013/games/artistic-achievement, Last accessed on 2020-06-08.

[11] British Academy of Film and Television Arts. 2018. Games – Artistic Achievement in 2018. http://awards.bafta.org/award/2018/games/artistic-achievement, Last accessed on 2020-06-08.

[12] John L Campbell, Charles Quincy, Jordan Osserman, and Ove K Pedersen. 2013. Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and

agreement. *Sociological Methods & Research* 42, 3 (2013), 294–320. https://doi.org/10.1177/0049124113500475

[13] Tom Cole, Paul Cairns, and Marco Gillies. 2015. Emotional and Functional Challenge in Core and Avant-garde Games. In *CHI PLAY '15*. ACM, 121–126. https://doi.org/10.1145/2793107.2793147

[14] Tom Cole and Marco Gillies. 2019. Thinking and Doing: Challenge, Agency, and the Eudaimonic Experience in Video Games. *Games and Culture* (2019), 1555412019881536. https://doi.org/10.1177/1555412019881536

[15] Katherine N Cotter, Alyssa N Prince, Alexander P Christensen, and Paul J Silvia. 2019. Feeling like crying when listening to music: Exploring musical and contextual features. *Empirical Studies of the Arts* 37, 2 (2019), 119–137. https://doi.org/10.1177/0276237418805692

[16] Katherine N Cotter, Paul J Silvia, and Kirill Fayn. 2018. What does feeling like crying when listening to music feel like? *Psychology of Aesthetics, Creativity, and the Arts* 12, 2 (2018), 216. https://doi.org/10.1037/aca0000108

[17] Rui Craveirinha and Licínio Roque. 2010. Looking for the Heart of Interactive Media: Reflections on Video Games' Emotional Expression. In *Proceedings of the 3rd International Conference on Fun and Games* (Leuven, Belgium) *(Fun and Games '10)*. Association for Computing Machinery, New York, NY, USA, 8–17. https://doi.org/10.1145/1823818.1823819

[18] Rui Craveirinha and Licinio Roque. 2019. Impact of Game Elements in Players Artistic Experience. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–6. https://doi.org/10.1145/3290607.3313049

[19] Rowan Daneels, Nicholas D Bowman, Daniel Possler, and Elisa D Mekler. 2021. The 'eudaimonic experience': A scoping review of the concept in digital games research. *Media and Communication* 9, 2 (2021), 178–190.

[20] Davis, Kendall Deacon. 2017. Can a Computer Make You Cry? https://medium.com/bestcompany/can-a-computer-make-you-cry-afc76bd27784, Last accessed on 2021-01-26.

[21] John Dewey. 2005. *Art as experience*. Penguin.

[22] Keith Diefendorff. 1999. Sony's Emotionally Charged Chip. *Microprocessor report* 13, 5 (1999), 1–6.

[23] Pierre Dragicevic. 2016. Fair Statistical Communication in HCI. In *Modern Statistical Methods for HCI*. 291–330. https://doi.org/10.1007/978-3-319-26633-6_13 arXiv:arXiv:1011.1669v3

[24] Denis Dutton. 2006. A naturalist definition of art. *The Journal of Aesthetics and Art Criticism* 64, 3 (2006), 367–377.

[25] Kirill Fayn, Paul J Silvia, Yasemin Erbas, Niko Tiliopoulos, and Peter Kuppens. 2018. Nuanced aesthetic emotions: Emotion differentiation is related to knowledge of the arts and curiosity. *Cognition and Emotion* 32, 3 (2018), 593–599. https://doi.org/10.1080/02699931.2017.1322554

[26] Alan Page Fiske, Beate Seibt, and Thomas Schubert. 2019. The sudden devotion emotion: Kama muta and the cultural practices whose function is to evoke it. *Emotion Review* 11, 1 (2019), 74–86. https://doi.org/10.1177/1754073917723167

[27] Alf Gabrielsson and Siv Lindström Wik. 2003. Strong experiences related to music: adescriptive system. *Musicae scientiae* 7, 2 (2003), 157–217. https://doi.org/10.1177/102986490300700201

[28] Berys Gaut. 2000. 'Art'as a Cluster Concept. *Theories of Art Today* (2000), 25–44.

[29] Chad Phoenix Rose Gowler and Ioanna Iacovides. 2019. "Horror, guilt and shame"–Uncomfortable Experiences in Digital Games. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. 325–337. https://doi.org/10.1145/3311350.3347179

[30] Julian Hanich, Valentin Wagner, Mira Shah, Thomas Jacobsen, and Winfried Menninghaus. 2014. Why we like to watch sad films. The pleasure of being moved in aesthetic experiences. *Psychology of Aesthetics, Creativity, and the Arts* 8, 2 (2014), 130. https://doi.org/10.1037/a0035690

[31] Houlden, Sophie. 2012. Can Art be Games? https://sophiehoulden.com/can-art-be-games/, Last accessed on 2021-01-27.

[32] Wijnand A IJsselsteijn, Yvonne AW de Kort, and Karolien Poels. 2013. The game experience questionnaire. *Eindhoven: Technische Universiteit Eindhoven* 46, 1 (2013).

[33] Katherine Isbister. 2016. *How games move us: Emotion by design*. Mit Press.

[34] Henry Jenkins. 2005. Games, the new lively art. In *Handbook of computer game studies*. IEEE Comput. Soc, 175–189. https://doi.org/10.1109/MASCOT.2002.1167092

[35] Bryan Jenner, Uwe Flick, Ernst von Kardoff, and Ines Steinke. 2004. *A companion to qualitative research*. Sage.

[36] Rilla Khaled. 2018. Questions over answers: Reflective game design. In *Playful disruption of digital media*. Springer, 3–27.

[37] Kroll, Jack. 2000. Emotion Engine? I Don't Think So. https://www.newsweek.com/emotion-engine-i-dont-think-so-156675, Last accessed on 2021-02-14.

[38] Shringi Kumari, Sebastian Deterding, and Jonathan Freeman. 2019. The Role of Uncertainty in Moment-to-Moment Player Motivation: A Grounded Theory. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. 351–363. https://doi.org/10.1145/3311350.3347148

[39] Karen S Kurasaki. 2000. Intercoder reliability for validating conclusions drawn from open-ended interview data. *Field methods* 12, 3 (2000), 179–194. https://doi.org/10.1177/1525822X0001200301

[40] Helmut Leder and Marcos Nadal. 2014. Ten years of a model of aesthetic appreciation and aesthetic judgments : The aesthetic episode - Developments and challenges in empirical aesthetics. *British Journal of Psychology* 105, 4 (Nov. 2014), 443–464. https://doi.org/10.1111/bjop.12084

[41] Limesurvey GmbH. 2020. LimeSurvey.

[42] Kirsti Malterud, Volkert Dirk Siersma, and Ann Dorrit Guassora. 2016. Sample size in qualitative interview studies: guided by information power. *Qualitative health research* 26, 13 (2016), 1753–1760. https://doi.org/10.1177/1049732315617444

[43] Tim Marsh and Brigid Costello. 2013. Lingering serious experience as trigger to raise awareness, encourage reflection and change behavior. In *International Conference on Persuasive Technology*. Springer, 116–124. https://doi.org/10.1007/978-3-642-37157-8_15

[44] Elisa D Mekler, Ioanna Iacovides, and Julia Ayumi Bopp. 2018. "A Game that Makes You Question..." Exploring the Role of Reflection for the Player Experience. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. 315–327. https://doi.org/10.1145/3242671.3242691

[45] Winfried Menninghaus, Valentin Wagner, Julian Hanich, Eugen Wassiliwizky, Milena Kuehnast, and Thomas Jacobsen. 2015. Towards a psychological construct of being moved. *PloS one* 10, 6 (2015), e0128451. https://doi.org/10.1371/journal.pone.0128451

[46] Winfried Menninghaus, Valentin Wagner, Eugen Wassiliwizky, Ines Schindler, Julian Hanich, Thomas Jacobsen, and Stefan Koelsch. 2019. What are aesthetic emotions? *Psychological review* 126, 2 (2019), 171. https://doi.org/10.1037/rev0000135

[47] Katharine Neil. 2007. My game in your gallery? *the user manual* (2007).

[48] C. Thi Nguyen. 2020. *Games: Agency As Art* (1 ed.). Oxford University Press. https://doi.org/10.1093/oso/9780190052089.001.0001

[49] Niedenthal, Simon. 2009. What We Talk About When We Talk About Game Aesthetics. Digital Games Research Association (DiGRA) Online Library. http://urn.kb.se/resolve?urn=urn:nbn:se:mau:diva-11023, Last accessed on 2021-02-08.

[50] Ninja Theory. 2017. *Hellblade: Senua's Sacrifice*. Videogame [Multiplatform]. Ninja Theory, Cambridge, England.

[51] Karen Nylund-Gibson and Andrew Young Choi. 2018. Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science* 4, 4 (2018), 440. https://doi.org/10.1037/tps0000176

[52] Mary Beth Oliver and Tilo Hartmann. 2010. Exploring the role of meaningful experiences in users' appreciation of "good movies". *Projections* 4, 2 (2010), 128–150. https://doi.org/10.3167/proj.2010.040208

[53] Felan Parker. 2018. Roger Ebert and the Games-as-Art Debate. *Cinema Journal* 3 (2018), 25. https://doi.org/10.1353/cj.2018.0032

[54] J. R. Parker. 2013. Games are art: Video games as theatrical performance. In *2013 IEEE International Games Innovation Conference (IGIC)*. IEEE, Vancouver, BC, Canada, 203–208. https://doi.org/10.1109/IGIC.2013.6659148

[55] Matthew Pelowski and Fuminori Akiba. 2011. A model of art perception, evaluation and emotion in transformative aesthetic experience. *New Ideas in Psychology* 29, 2 (2011), 80–97. https://doi.org/10.1016/j.newideapsych.2010.04.001

[56] Matthew Pelowski, Michael Forster, Pablo P. L. Tinio, Maria Scholl, and Helmut Leder. 2017. Beyond the lab: An examination of key factors influencing interaction with 'real' and museum-based art. *Psychology of Aesthetics, Creativity, and the Arts* 11, 3 (Aug. 2017), 245–264. https://doi.org/10.1037/aca0000141

[57] Matthew Pelowski, Young-Jin Hur, Katherine N Cotter, Tomohiro Ishizu, Alexander P Christensen, Helmut Leder, and IC McManus. 2019. Quantifying the if, the when, and the what of the sublime: A survey and latent class analysis of incidence, emotions, and distinct varieties of personal sublime experiences. *Psychology of Aesthetics, Creativity, and the Arts* (2019). https://doi.org/10.1037/aca0000273

[58] Matthew Pelowski, Patrick S Markey, Michael Forster, Gernot Gerger, and Helmut Leder. 2017. Move me, astonish me... delight my eyes and brain: The Vienna integrated model of top-down and bottom-up processes in art perception (VIMAP) and corresponding affective, evaluative, and neurophysiological correlates. *Physics of Life Reviews* 21 (2017), 80–125. https://doi.org/10.1016/j.plrev.2017.02.003

[59] Matthew John Pelowski. 2015. Tears and transformation: Feeling like crying as an indicator of insightful or "aesthetic" experience with art. *Frontiers in Psychology* 6 (2015), 1006. https://doi.org/10.3389/fpsyg.2015.01006

[60] Xiaolan Peng, Jin Huang, Alena Denisova, Hui Chen, Feng Tian, and Hongan Wang. 2020. A Palette of Deepened Emotions: Exploring Emotional Challenge in Virtual Reality Games. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, 94. https://doi.org/10.1145/3313831.3376221

[61] Pippin Barr. 2011. *The Artist Is Present*. Videogame [Browser]. Pippin Barr, Montréal, Canada.

[62] Pippin Barr. 2020. *The Artist Is Present 2*. Videogame [Browser]. Pippin Barr, Montréal, Canada.

[63] Doris C. Rusch. 2017. *Making deep games: designing games with meaning and purpose*. CRC Press, Taylor & Francis Group, an Informa business, Boca Raton, FL.

[64] Doris C Rusch and Andrew M Phelps. 2020. Existential Transformational Game Design: Harnessing the "Psychomagic" of Symbolic Enactment. *Frontiers in Psychology* 11 (2020), 3021. https://doi.org/10.3389/fpsyg.2020.571522

[65] Mike Schaekermann, Giovanni Ribeiro, Guenter Wallner, Simone Kriglstein, Daniel Johnson, Anders Drachen, Rafet Sifa, and Lennart E Nacke. 2017. Curiously Motivated: Profiling Curiosity with Self-Reports and Behaviour Metrics in the Game" Destiny". In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. 143–156. https://doi.org/10.1145/3116595.3116603

[66] Ines Schindler, Georg Hosoya, Winfried Menninghaus, Ursula Beermann, Valentin Wagner, Michael Eid, and Klaus R Scherer. 2017. Measuring aesthetic emotions: A review of the literature and a new assessment tool. *PLoS One* 12, 6 (2017), e0178899. https://doi.org/10.1371/journal.pone.0178899

[67] John Sharp. 2015. *Works of game: On the aesthetics of games and art*. mit Press.

[68] Aleksandra Sherman and Clair Morrissey. 2017. What is art good for? the socio-epistemic value of art. *Frontiers in Human Neuroscience* 11 (2017), 411. https://doi.org/10.3389/fnhum.2017.00411

[69] Eugénie Shinkle. 2005. Feel It, Don't Think: the Significance of Affect in the Study of Digital Games. *Digital Games Research Association (DiGRA) 2005: Changing Views: Worlds in Play* (2005), 7.

[70] Paul J. Silvia. 2010. Confusion and interest: The role of knowledge emotions in aesthetic experience. *Psychology of Aesthetics, Creativity, and the Arts* 4, 2 (2010), 75–80. https://doi.org/10.1037/a0017081

[71] Martin Skov and Marcos Nadal. 2020. A farewell to art: Aesthetics as a topic in psychology and neuroscience. *Perspectives on Psychological Science* 15, 3 (2020), 630–642. https://doi.org/10.1177/1745691619897963

[72] Smithsonian American Art Museum. 2012. The Art of Video Games. https://americanart.si.edu/exhibitions/games, Last accessed on 2021-02-14.

[73] Aaron Smuts. 2005. Are Video Games Art? *Contemporary Aesthetics* 3, 1 (2005), 6.

[74] Eva Specker, Michael Forster, Hanna Brinkmann, Jane Boddy, Matthew Pelowski, Raphael Rosenberg, and Helmut Leder. 2018. The Vienna Art Interest and Art Knowledge Questionnaire (VAIAK): A Unified and Validated Measure of Art Interest and Art Knowledge. *Psychology of Aesthetics, Creativity, and the Arts* October (Oct. 2018). https://doi.org/10.1037/aca0000205

[75] Steve Swink. 2008. *Game feel: a game designer's guide to virtual sensation*. CRC Press.

[76] Grant Tavinor. 2009. *The art of videogames*. John Wiley & Sons.

[77] thatgamecompany. 2012. *Journey*. Videogame [Multiplatform]. Sony Interactive Entertainment, Tokyo, Japan.

[78] Alexandra To, Safinah Ali, Geoff F Kaufman, and Jessica Hammer. 2016. Integrating Curiosity and Uncertainty in Game Design.. In *Proceedings of the First International Joint Conference of Digital Games Research Association and Foundations of Digital Games*.

[79] Alexandre N. Tuch and Kasper Hornbæk. 2015. Does Herzberg's Notion of Hygienes and Motivators Apply to User Experience? *ACM Trans. Comput.-Hum. Interact.* 22, 4, Article 16 (June 2015), 24 pages. https://doi.org/10.1145/2724710

[80] April Tyack and Elisa D. Mekler. 2021. Off-Peak: An Examination of Ordinary Player Experience. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 115, 12 pages. https://doi.org/10.1145/3411764.3445230

[81] Jan B Vornhagen, April Tyack, and Elisa D Mekler. 2020. Statistical Significance Testing at CHI PLAY: Challenges and Opportunities for More Transparency. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. 4–18. https://doi.org/10.1145/3410404.3414229

[82] Jonna K Vuoskoski and Tuomas Eerola. 2017. The pleasure evoked by sad music is mediated by feelings of being moved. *Frontiers in psychology* 8 (2017), 439. https://doi.org/10.3389/fpsyg.2017.00439

[83] Ge Wang. 2018. *Artful Design: Technology in Search of the Sublime, A MusiComic Manifesto*. Stanford University Press.

[84] Matthew Alexander Whitby, Sebastian Deterding, and Ioanna Iacovides. 2019. "One of the baddies all along" Moments that Challenge a Player's Perspective. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. 339–350. https://doi.org/10.1145/3311350.3347192

A Appendix: Tables

Table II.3: Full AESTHEMOS questionnaire, descriptive statistics and Cronbach's α for each AESTHEMOS dimension ($n = 168$).

Item Name	Item ID	M	SD	Cronbach's α
I found it distasteful	Ugliness_1	1.18	0.59	.54
I found it ugly	Ugliness_2	1.48	0.98	
I felt confused	Confusion_1	2.17	1.18	.65
It was unsettling to me	Confusion_2	2.14	1.31	
It made me aggressive	Anger_1	1.68	1.12	.73
It made me angry	Anger_2	1.42	0.89	
I felt indifferent	Boredom_1	1.21	0.54	.46
It bored me	Boredom_2	1.40	0.89	
It worried me	Uneasiness_1	1.93	1.25	.58
Felt oppressive	Uneasiness_2	1.77	1.23	
I felt deeply moved	BeingMoved_1	3.83	1.29	.87
It touched me	BeingMoved_2	3.96	1.23	
It baffled me	Surprise_1	2.47	1.36	.52
It surprised me	Surprise_2	3.76	1.15	
I was impressed	Fascination_1	4.32	0.94	.6
It fascinated me	Fascination_2	4.23	0.96	
I liked it	Beauty/Liking_1	4.69	0.61	.65
I found it beautiful	Beauty/Liking_2	4.54	0.87	
It challenged me intellectually	Int. Challenge_1	3.27	1.31	.68
Was mentally engaged	Int. Challenge_2	4.18	1.00	
I found it sublime	Awe_1	3.96	1.11	.45
I felt Awe	Awe_2	3.65	1.33	
Made me curious	Interest_1	3.78	1.24	.62
Sparked my interest	Interest_2	4.27	0.95	
I felt a sudden insight	Insight_1	4.01	1.26	.75
I sensed a deeper meaning	Insight_2	3.30	1.34	
It motivated me to act	Energy_1	2.96	1.40	.56
It energized me	Energy_2	3.10	1.24	
It spurred me on	Vitality_1	3.11	1.35	.65
It invigorated me	Vitality_2	3.07	1.28	
I felt something wonderful	Enchantment_1	4.01	1.17	.74
I was enchanted	Enchantment_2	4.11	1.13	
It made me feel nostalgic	Nostalgia_1	2.96	1.52	.74
It made me feel sentimental	Nostalgia_2	3.67	1.36	
It calmed me	Relaxation_1	3.27	1.38	.84
It relaxed me	Relaxation_2	3.23	1.42	
It made me sad	Sadness_1	3.10	1.45	.73
It made me feel melancholic	Sadness_2	2.96	1.48	
It amused me	Humor_1	3.11	1.38	.79
It was funny to me	Humor_2	2.27	1.31	
It delighted me	Joy_1	3.80	1.16	.8
It made me happy	Joy_2	3.78	1.25	

Paper III

"I'm the leader and I'm going to save the world": Characterizing Empowering and Disempowering Game Experiences

Jan B. Vornhagen, Dan Bennett, Dooley Murphy, & Elisa D. Mekler

The paper has been published in the

Proceedings of the ACM on Human-Computer Interaction Vol. 7, pp. 1330–1360, 2023.
https://dl.acm.org/doi/10.1145/3611071

The layout has been revised.

Abstract

Empowering people through technology is a core concern of HCI, yet little is known about how players experience empowerment or disempowerment through videogames. We surveyed 250 participants about their dis-/empowering videogame experiences, investigating why they felt dis-/empowered, and how these experiences related to core player experience constructs including emotion and basic needs satisfaction. While empowering experiences were often positive, and disempowering experiences often negative, we found meaningful exceptions to this, and a surprising complexity in player accounts. We capture this diversity in seven themes. These range from "heroic victories" which follow long periods of failure, to positive experiences of disempowerment, which were appreciated for their narrative meaning. By articulating these complex experiences, and relating them to quantitative measures we provide a foundation for understanding of the role of dis-/empowerment in player experience, and highlight avenues for future work. Data and analyses are available at https://osf.io/zhtu8/

III.1 Introduction

Empowerment is considered a desirable goal for HCI, with researchers and tech companies employing the term to signify their intent to improve people's lives through technology [72, p.1]. Given the positive connotations of empowerment, and its apparently "simple inherent meaning" [91, p.2], it is unsurprising that empowerment has been adopted by many different strands of (HCI) research: Empowerment has been used, for example, as a synonym for having superpowers [43], to explore power differentials reinforced by technology [e.g., 19] and to motivate AI agents [40]. In HCI games research, empowerment has been used in relation to both playing [e.g., 43, 27], and making games [e.g., 6, 81]; it has been used as a lens to study participatory culture [24], and gender equality in game communities [58]; and level of empowerment has been argued to predict player experience (PX) [31].

This plurality of uses and understandings speaks to the promise of empowerment: both as a worthwhile goal of technology and as a fruitful research avenue. However, it also illustrates the challenges involved in defining the term. For some time, researchers have argued that both "empowerment", and its conceptual opposite "disempowerment" are used inconsistently [e.g., 21, 72, 19]. Some have argued that the terms seem to overlap significantly with other theoretical constructs, and some forms of empowerment discussed in HCI games research [e.g., psychological empowerment; 96] may share much conceptual overlap with core player experience constructs [e.g. the PXI; 1], as well as theoretical frameworks used in HCI games research, such as self-determination theory [84] and attribution theory [23].

At the same time, HCI games research often seems to treat empowerment as a synonym for broadly positive experience, or else as a vague gesture towards improvements in players' lives. Taken together, this contributes to a lack of clarity in our engagement with the construct of empowerment — which in turn impedes researchers' ability to discuss these phenomena, share findings and develop a coherent knowledge base [83]. Exacerbating this, it appears that in HCI games research "empowerment" is usually attributed "from the outside" by authors — rarely by the players themselves. Researchers will observe a positive player experience or successful intervention, and then describe this as a matter of having empowered the players [e.g., 6, 31]. It remains unclear whether players considered *themselves* empowered and what their experience of empowerment entailed. This is an important issue, as some accounts of empowerment emphasize contextual factors such as person-environment fit [96] which would point to significant variation between people and cases [95].

In this work, we seek to address these questions around conceptualizations of empowerment by collecting and analyzing players' own accounts of empowering and disempowering player experiences. Specifically, we conducted an online study ($n = 250$) in which participants were asked to describe an empowering or disempowering videogame experience, and to rate their experienced in terms of need satisfaction and frustration, constructs of the Player Experience Inventory [PXI, 1], and attribution [23].

Our contribution is three-fold: First, we present empirical evidence of the different kinds of experiences players themselves understand as dis-/empowering. In particular, our findings showcase how an experience may come to be dis-/empowering, going far beyond conditions which affect players perceived or real abilities and achievements. We reflect on what these experiences mean for players, designers and researchers.
Second, we differentiate dis-/empowerment from related PX concepts. This helps clarify the concepts of dis-/empowerment as player experience, and strengthen lateral connections with past research on failure, frustration and emotional challenge.
Third, we further understandings of how and when videogame experiences come to be enjoyed and appreciated, irrespective of adversity and surface level emotional valence. We discuss a number of directions for HCI games research suggested by these understandings.

III.2 Related Work

The concept of empowerment has found wide adoption in HCI research and games discourse alike. In their review of empowerment in HCI, Schneider et al. suggest that for both researchers and practitioners empowering people *"is an unconditionally positive mission no one would argue*

against"[72, p.1], while game designer Warren Spector recently described empowerment as the "ultimate success criterion" [78]. However, the understanding and use of the term can differ widely. Empowerment can mean giving differently-abled players the ability to play and compete with one-another [27, 29]. In other work it can mean to simply exaggerate players' movements in a mixed reality game [43] giving the in-game appearance of "super-powers". Elsewhere, researchers have sought to empower children by providing coding-free development tools, to consequently foster long term interest in the activity [6].

This latter goal of *persistent* (player) empowerment [see 72] is often emphasized by *serious games* [see 50] research. A use of "empowerment" which is analogous to its use in sociology [see 77, 63, 95, 91]. Serious games do not seek to empower players only within the game, but rather seek to empower people in regards to certain societal and personal issues [see 86]. For example, serious game developers seek to change the perception of mental health problems[e.g., 32], empower players to develop emotional regulation skills [88] or sensitize them towards risk taking behavior [30].

Another quite different understanding of empowerment can be seen in a line of work spanning AI and player experience [40, 41, 31]. In contrast with the life-situated work in serious games, this line of work treats empowerment as a *"context-free"* [p.1 41] *universal function*. Here empowerment is the quantification of *"an agent's potential perceivable influence"*[31, p.3]. While this account is developed for AI models, the authors nonetheless argued that the drive to increase this influence is a unifying theme of all beings. In line with this assumption of universality, recent work has transposed this context-free account back into humans: Guckelsberger et al. [31] linked player experience to their awareness of the amount of potential actions they can perform — a measure which seems closely related to what has elsewhere been called "agency" [e.g. 36, 80].

Notably, these examples all emphasize empowerment as a sought after *outcome*. Little work addresses players' *experience* of empowerment, and player experience of empowerment is rarely assessed in HCI games studies. This is despite the fact that empowerment during the play experience is considered a highly useful lens for game design [78], and seems likely to serve as a fruitful avenue for PX research. Some papers capture user experience measures such as self-esteem [e.g., 27], or basic need satisfaction [e.g., 43], and some interpret qualitative accounts of player experience (such as the perceived ability to "squish others") as markers of empowerment [e.g., 43]. However, we have found no work which asks players directly about whether they actually feel dis-/empowered, nor about the qualities of this feeling of dis-/empowerment.

In many conceptualizations, empowerment seems likely to make for good player experience. Zimmerman's "Psychological Empowerment" [96] conceptualizes empowerment as an individual, context-dependent phenomenon, grounded in intrapersonal, interactional and behavioral

components. I.e., to be empowered a person has to believe that they can change, understand, and control their current situation [95]. This is closely related to concepts and theories already in use by HCI games research such as SDT [68] [for an overview see 84], attribution theory [23], and compound PX measures such as the PXI [1], which all put a premium on the player feeling like they are in control, can act freely, and have impact. In fact, Zimmerman argues that psychological empowerment seems to be closely related to — among others — self-efficacy, self-esteem and competence [95, p.590]. This points to a promising direction forward for clarifying understandings of empowerment experience, and one which to date has not been explored in HCI games research. At the same time, however, this account does not capture one complex aspect of empowerment in positive gaming experiences: that in the right circumstances, positive experiences can be supported by *disempowerment*.

The majority of videogames involve some aspect of player failure. While players generally want to win, they also seem to like games which make them lose. Juul called this the *Paradox of Failure* [35], pointing out that at least the credible threat of failure is needed to enjoy a game. This notion has since been supported, with other works pointing out that the struggle to succeed can lead to more highly appraised experiences [59, 25].

In extension, disempowerment in games can also be a subversive act, against the medium's usual reverence of agency and control [24]. For example, Benford et al. [7] deliberately disempowered players by having them surrender control to a machine or even a total stranger as a way to induce *discomfort*; attempting "not to cause long term suffering or pain, but rather to underpin positive design values related to entertainment, enlightenment and sociality" [7, p.2011].

Of course such experiences of disempowerment can have a negative impact on players: Gowler and Iacovides [28] describe gaming experiences which players strongly disliked; noting that "players become uncomfortable because they just don't know what is going to happen next, or if they will be able to cope" [28, p. 333]. Likewise, disempowerment can be actively used to force players into predatory micro-transactions [38]. An important factor in this seems to be the disempowered player's ability to consent to the experience. For example, players of discomforting TRPGs [52], who had consented to such experiences, and were prepared for them, welcomed similar if not more extreme themes, and subsequent feelings of disempowerment.

In summary, prior literature points to quite a degree of conceptual uncertainty: both around understandings of empowerment, and around its relationship to other constructs, and to players' experiences. It is hard to answer when and how *players themselves* experience dis-/empowerment in videogames. Experiences which on the surface seem to be disempowering — such as failure or loss — also seem to have the potential to be highly enjoyable gaming experiences. Likewise, if struggle and failure are seen as key parts of a good gaming experience [35], then perfect empowerment of the player seems unlikely to be enjoyed.

III.3 Methods

The aim of this work is to characterize and understand better the experience of empowerment and disempowerment in games. To this end, we examined two research questions:

(RQ1) How do players understand and describe empowering and disempowering experiences?

(RQ2) Do empowering and disempowering experiences differ on common PX measures, and if so, how?

There is a lack of systematic prior work on *experiences* of empowerment in games research, which would support hypothesis formation. As such we take an exploratory approach to our research questions. We randomly assign participants to one of two conditions, asking them to describe either a videogame experience in which they felt empowered or disempowered, respectively. We deliberately refrained from providing any definitions of dis-/empowerment to not prematurely constrain the types of experiences we would get. To answer RQ1, we used open-ended questions, and analyzed responses with thematic analysis. To answer RQ2, we used self-reported quantitative measures for which we report the descriptive results. The exact measures are described below. The de-identified raw data, analysis scripts and all research artifacts can be found at https://osf.io/zhtu8/.

III.3.1 Participants

We collected 264 full responses. Of these, 35 were recruited via snowball sampling and 229 via Prolific.co. The latter were pre-screened using the Prolific pre-screen options to have an approval rating > 90, be fluent in English, count videogames as their hobbies and report playing videogames for at least 3 hours per week on average. Participants on Prolific were paid £9/h in accordance with Prolific's recommendation as of August 2022.

Of the 264 full responses, we excluded 7 participants who failed one of the three attention checks. We removed two participant whose answers were not recorded (all N/A), one participant who was not 18, one double entry, one participant who answered in Spanish and two participants who submitted nonsensical answers. After exclusions we considered $n = 250$ participants in the analysis. Among these, the participants' median age was 26 ($min = 19; max = 64$), 191 indicated they were a man, 56 indicated they were a woman, 2 indicated they were a non binary person and one person did not report their gender. Participants were randomly assigned to one of two conditions, where they were either asked to describe an experience in which they felt empowered ($n = 110$) or disempowered ($n = 140$).

III.3.2 Procedure

We invited participants to take part in a study about dis-/empowering gaming experiences. After giving informed consent and confirming they were over 18, participants were asked their age, gender and gaming habits. They were then randomly assigned to one of the two conditions. Participants were asked to bring to mind a videogame experience in which they felt either disempowered or empowered depending on condition. On the next page, we asked the participants to describe this experience in at least 50 words, then to explain why they felt dis- or empowered during the experience, and lastly to describe the emotions they felt during the experience. Participants then filled out several qualitative PX questionnaires (see Chapter III.3.3) and optionally gave comments on their quantitative answers. Finally, we gave participants the option to give an overall comment about the study, before they were thanked and debriefed.

III.3.3 Player Experience Measures

Since we are interested in understanding how dis-/empowering experiences relate to common PX concepts (RQ2), we employ several quantitative measures. All the following scales were presented on a seven point Likert scale (-3 strongly disagree - 0 neither disagree or agree - 3 strongly agree). However, scales were internally recorded as 1 through 7 to facilitate interpretation. Hence a score of 1 indicates strong disagreement, 4 indicates neither agreement, nor disagreement, and 7 indicates strong agreement.

Player Experience Inventory. The Player Experience Inventory (PXI) was only recently published [1], but has since seen wide adoption [e.g., 76, 64, 26]. It utilizes Means-End theory to propose that positive PX is dependent on *functional* and *psychosocial consequences*. The former is thereby primarily derived from the attributes of the game, while the latter incorporates how players values align with the game. Each consequence contains five dimensions: The functional consequences contain *ease of control, progress feedback, audiovisual appeal, goals and rules, and challenge*; The psychosocial consequences contain *mastery, curiosity, immersion, autonomy, and meaning*.

Game-Specific Attribution Questionnaire The Game-Specific Attribution Questionnaire (GSAQ; [23]) measures how players attribute in-game success. It contains four dimensions: (1) Internality/externality - Is the success due to a characteristic of the player, or the environment; (2) stability — could the player repeat the success; (3) controllability — Is the player in control over what caused the success; and (4) globality — are the reasons for success applicable in

different situations. Understanding players' attributions can be key to understanding players experience of feeling dis- or empowered, as this feeling can hinge on the player believing the success was their own ability or e.g. due to the game helping them [23] ([see also 27]).

Need Satisfaction and Frustration. We employ a quantitative, self-report measure for basic psychological need satisfaction and frustration. Several standardized questionnaires are in use in HCI games research to assess need satisfaction. One often used questionnaire is the Player Experience Need Satisfaction questionnaire [PENS, 69]. While PENS is well-established in HCI games research [84], it is proprietary and concerns have been raised about its validity [33, 34]. Similarly the Ubisoft Perceived Experience Questionnaire [UPEQ, 3] has come into question regarding its factor structure [37].

Needs *frustration* — which may be of importance in disempowerment — is rarely measured in HCI games research, despite need satisfaction and frustration being shown to be distinct dimensions [74, 46, 82]. Due to this lack of a validated questionnaire measuring need satisfaction and frustration in gaming experiences we opted to use a compound questionnaire which features distinct items for need satisfaction and frustration, and which has been used successfully in a previous project. The full questionnaire and prior performance metrics can be found in the OSF repository and an overview can be found in Table III.1. Note that we reused the PXI scales Autonomy and Mastery as autonomy satisfaction and competence satisfaction respectively in the SDT questionnaire. We decided to employ the scales twice to ensure participants answered these questions in the context of the other items maintaining the psychometrics of the measures.

Appreciation and Enjoyment. To understand whether and to what extent participants regarded dis-/empowering videogame experiences as positive we measured appreciation and enjoyment. We used the scale derived from Oliver et al. [54], which has been used in prior work about mixed-affect videogame experiences [e.g., 12, 13, 57]. We chose to measure both enjoyment and appreciation, as past research has shown appreciation to constitute a distinct facet of the player's game experience; especially so when the experience is not "fun" in the traditional sense [12, 55].

III.3.4 Emotions

We asked participants to report the emotions they felt during the experience they described. The free answers were manually corrected for spelling before being classified using an R implementation [89] of the Geneva Affect Label Coder (GALC [71]). The GALC is a collection of word stems and synonyms of 38 emotion groups. For example, "admir*", "ador*", "fascina*" etc. are

grouped together as "Admiration/Awe", while stems like "dissapoint*", "disgruntl*", "frustrat* etc. are grouped together as "Disappointment". Emotion words that were not included in the GALC - such as "excitement" - were marked and retained without further grouping.

III.3.5 Open Questions and Thematic Analysis

There is a host of questionnaires to quantify and operationalize empowerment (mostly in medical fields [e.g., 75, 44, 2, 79, 51, 61, 66, 87, 48]). However, the measurement of empowerment is highly context dependent [p. 596 95]. As there is no extant work on dis-/empowering experience in games, we aimed to understand how players themselves experienced dis-/empowerment to provide a basis for future work. We therefore opted to use open questions to collect information on the dis-/empowering experiences of players, and analyze these with thematic analysis.

We used two questions: (1) "Please describe the videogame experience which made you feel dis-/empowered. Give as much detail as possible, so we understand what about this experience was dis-/empowering or made you feel dis-/empowered."; and (2) In your opinion, why did this experience make you feel dis-/empowered? Please give as much detail as you need in order to explain this clearly.". These questions were adapted to address disempowerment or empowerment depending on the condition participants were assigned to.

To analyze participants' written accounts, we employed reflexive thematic analysis (TA), following best practices described by Braun and Clarke [16, 17, 18]. The analysis was performed on a reduced dataset only including the open questions, group membership, age, gender, emotions and any comments the participants made.

For the analysis, the first author read all participant responses to familiarize themselves with the data and generated an initial set of codes. During the second step, authors one, two, and four coded the first ten participants, and discussed the generated codes to decide the granularity of the coding process. Author one then employed a complete coding approach [17, p. 210], i.e., they systematically worked through each participant's answer, coding every part that was relevant to the research question. As we did not want to make any a priori assumptions about the qualities and properties of disempowering and empowering experiences, responses in both conditions were coded together and could share the same codes.

In a second coding round, codes were adjusted and clustered based on similarities. Based on these code clusters, the first author then grouped the codes and generated preliminary themes. These themes were then again discussed and adjusted among authors one, two, and four. The first author then completed a third complete coding round before deciding on the final themes. Participants could be assigned to more than one theme. The final themes are reported in the results section below. Participants' experience reports and intermediate files documenting the

analysis process can be found in the OSF repository.

III.4 Results

Quantitative analyses were conducted in RStudio [67], using R 4.2.1 [62] and the packages tidyverse [93], Cairo [85], introdataviz [53], patchwork [56], ltm [65], readxl [94] and GALCR [89]. As this study had an exploratory aim, we only describe the data and refrain from reporting inferential statistics.

Most experiences reported were recent, taking place less than two months ago ($n = 84$), followed by experiences which took place 2-11 months ago ($n = 62$), 1-2 years ago ($n = 46$), 2-10 years ago ($n = 45$) and over 10 years ago ($n = 13$). Most participants reported experiences in multiplayer games, with *League Of Legends* ($n = 32$) mentioned most frequently, followed by the *FIFA* series ($n = 12$), the *Dark Souls* Series ($n = 9$), *GTA* (series and online: $n = 9$), and *Counter Strike: Global Offensive* ($n = 8$).

III.4.1 PX Ratings

Overall, empowering experiences scored higher on appreciation, enjoyment, need satisfaction, as well as all the GSAQ and PXI constructs. Participants in the disempowered condition reported higher levels of need frustration. Moreover, with the exception of relatedness satisfaction, all scale ratings in the disempowerment condition exhibited greater standard deviations, indicating a wider range of responses. Responses in the empowerment condition were comparably more uniform (see Table III.1). We interpret the quantitative findings both with regards to the mean score M and standard deviation SD, to assess to what extent PX ratings were distinct or common to dis-/empowering experiences, respectively.

Empowering game experiences scored consistently high on enjoyment ($M(SD) = 6.56(0.58)$; see Table III.1), although disempowering experiences were characterized by moderate enjoyment ($M(SD) = 4.33(1.95)$). Participants in the empowered condition also appreciated the experience more ($M(SD) = 4.92(1.45)$).

With regards to psychological needs, participants in the empowered condition reported overall high autonomy and competence satisfaction, as well as low corresponding frustration. Participants in the disempowered condition in turn reported considerably lower satisfaction and higher — but not overly high frustration. This indicates that overall there was more of an absence of need satisfaction than active frustration. Regarding relatedness, participants in the empowered condition likewise reported low frustration, but also reported rather moderate relatedness

Table III.1: Descriptive results for the different PX scales overall and per group. Scales are taken from the sources provided. * indicates an adjustment of the scale. Refer to the OSF repository for details.

Variable	Source	Cronbach's α	Overall M(SD)	Empowering M(SD)	Disempowering M(SD)
n			250	110	140
Appreciation	[54]	0.81	4.25(1.74)	4.92(1.46)	3.74(1.76)
Enjoyment	[54]	0.95	5.29(1.89)	6.57(0.58)	4.32(1.96)
Basic Psychological Needs					
Autonomy Satisfaction	[1]	0.88	4.32(1.86)	5.61(1.18)	3.35(1.68)
Autonomy Frustration	[20]*	0.74	3.79(1.6)	2.97(1.39)	4.42(1.47)
Competence Satisfaction	[1]	0.91	4.77(1.89)	6.11(0.85)	3.75(1.83)
Competence Frustration	[20]*	0.83	3.61(1.85)	2.35(1.3)	4.56(1.63)
Relatedness Satisfaction	[3]*	0.89	4.16(1.85)	4.66(1.79)	3.78(1.81)
Relatedness Frustration	[74]*	0.84	3.18(1.52)	2.22(1.11)	3.91(1.37)
PXI					
Autonomy	[1]	0.93	4.64(1.81)	5.72(1.18)	3.82(1.78)
Audiovisual Appeal	[1]	0.86	5.84(1.26)	6.31(0.94)	5.48(1.36)
Challenge	[1]	0.79	4.46(1.52)	5.1(1.09)	3.98(1.62)
Clarity of Goals	[1]	0.8	5.93(1.17)	6.29(0.68)	5.66(1.37)
Curiosity	[1]	0.9	5.17(1.64)	5.63(1.38)	4.81(1.73)
Ease of Control	[1]	0.71	5.37(1.23)	5.84(0.89)	5.01(1.33)
Immersion	[1]	0.66	5.38(1.22)	5.81(0.92)	5.05(1.32)
Meaning	[1]	0.89	5.11(1.59)	5.93(0.98)	4.48(1.68)
Mastery	[1]	0.92	4.81(1.86)	6.07(0.85)	3.86(1.86)
Progress Feedback	[1]	0.76	5.15(1.33)	5.58(1.1)	4.82(1.39)
GSAQ					
Internality	[23]	0.76	4.81(1.32)	5.43(0.94)	4.35(1.36)
Stability	[23]	0.71	4.46(1.33)	4.92(1.01)	4.13(1.43)
Globality	[23]	0.65	3.6(1.33)	4.01(1.28)	3.29(1.28)
Controllability	[23]	0.79	5.59(1.26)	5.99(0.87)	5.27(1.4)

satisfaction. Participants in the disempowered condition in turn scored almost as high on relatedness satisfaction ($M(SD) = 3.78(1.81)$) as on frustration ($M(SD) = 3.92(1.37)$).

With regards to the PXI, functional consequences (i.e., ease of control, progress feedback, audiovisual appeal, clarity of goals) were rated rather positively in both conditions, with considerable overlap in-between. One exception is the challenge scale which participants in the disempowered condition rated rather low ($M(SD) = 3.89(1.62)$), indicating a potential mismatch between their skill and the obstacles in the game. More pronounced differences between conditions can be observed in the psychosocial consequences, particularly for autonomy and mastery, followed by meaning and curiosity. As expected, participants in the empowering condition reported considerably higher mastery and autonomy scores.

Lastly, the GSAQ also showed a considerable overlap between conditions. Participants in the disempowered condition report overall slightly less controllability, globality, and stability (see Table III.1), with internality ($M(SD) = 4.35(1.36)$) showing the biggest difference to the participants in the empowered group ($M(SD) = 5.42(0.94)$). On all four scales, participants in the disempowered condition reported more widely distributed scores, featuring comparably more mixed attribution than the empowered condition.

III.4.2 Emotions During Empowering and Disempowering Experiences

Table III.2: Frequencies of the top ten most commonly mentioned emotions overall and per condition. Emotion words were coded using the Geneva Affect Label Coder for R (GALCR; [89]), * refers to emotion words not originally included in the GALC [71]. The full list of emotion words is available in the supplementary material.

Overall (n=245)		Empowerment (n=106)		Disempowerment (n=139)	
Emotion	Count (%)	Emotion	Count (%)	Emotion	Count (%)
Anger	88 (35.92)	Happiness	65 (61.32)	Anger	80 (57.55)
Sadness	77 (31.42)	Excitement*	44 (41.51)	Sadness	69 (49.64)
Disappointment	68 (27.76)	Pride	27 (25.47)	Disappointment	65 (46.76)
Happiness	67 (27.35)	Joy	25 (23.58)	Irritation	18 (12.95)
Excitement*	48 (19.59)	Contentment	18 (16.98)	Boredom	16 (11.51)
Pride	27 (11.02)	Relief	14 (13.21)	Anxiety	12 (8.36)
Joy	25 (10.2)	Surprise	12 (11.32)	Helplessness*	12 (8.36)
Anxiety	22 (8.98)	Anxiety	10 (9.43)	Desperation	12 (8.36)
Irritation	19 (7.76)	Empowerment*	9 (8.49)	Confusion*	10 (7.19)
Contentment	18 (7.35)	Powerfulness*	9 (8.49)	Hope	10 (7.19)

Participants reported a wide range of emotions, with considerable variation between conditions. Overall, participants in the disempowerment condition reported a somewhat narrower range of emotions with $n = 86$ unique entries, compared to the $n = 95$ unique entries in the empowerment condition. In the empowerment condition, participants reported mostly positive emotions such as happiness ($n = 65$) and excitement ($n = 44$). Feelings of pride ($n = 27$) and relief ($n = 14$) were also common. Some participants in the empowerment condition did report negative emotions, though this was quite rare: The most commonly mentioned negative emotions were anxiety ($n = 10$), sadness ($n = 8$), anger ($n = 8$), and tension/stress ($n = 8$).

In the disempowerment condition, participants primarily reported negative emotions — mostly anger ($n = 80$), followed by sadness ($n = 69$) and disappointment ($n = 65$). Unsurprisingly, positive emotions were less common ($n = 16$). The most frequently mentioned positive emotions were hope ($n = 10$) and excitement ($n = 4$). This indicates a much heavier skew in emotional valence (here towards negative valence) compared to the empowerment condition, where mixed emotional experiences were more common.

III.4.3 Types of Empowering and Disempowering Experiences

In the following we first present the seven themes we generated during the thematic analysis: (1) The Joys of Control; (2) Heroic Victory; (3) Being the Worst; (4) Out of Their hands; (5) Sexist Harassment; (6) Seeing Themselves in the Game; and (7) Understanding the Message. We then compare them in terms of PX ratings (see also Table III.3 and Chapter A). Note that themes were not mutually exclusive, and participant accounts could be assigned to multiple themes. Participant quotes are reported unchanged, in their original spelling, accompanied by participant ID (as listed in the OSF repository), age, gender and game title.

The Joys of Control

This theme centers participants' enjoyment of having control over the game world, both at the level of individual actions and choices, but also with regards to other players. Participants emphasized their own abilities and skill, and what this allowed them to do and achieve. For example, how a game requires *"knowledge, strategy, and quick thinking,"* and how *"[a]fter winning a hard match I really do feel like I can do anything"* (506; F; 20; League Of Legends). Despite the emphasis on control, not all cases focused on the participants' own efficacy and skill. In some cases, the sense of power and control was augmented by the game, for example, the sensory feedback that accompanied player action: *"My character's [...] attacks have really nice, smooth animations with water particles flowing here and there. I enjoy running around, using attacks that make my character move faster, and attack enemies." (ID 433; M; 19; Black Desert Online)*.

Table III.3: PX ratings per theme and overall. Note that numbers do not add up to N = 250, as 28 participants were assigned two themes. Additional visualizations are available in the Appendix.

Variable M(SD)	The Joys Of Control	Heroic Victory	Being The Worst	Out Of Their Hands	Sexist Harassment	Seeing Themselves	Understanding The Message	Overall
n	37	62	36	94	8	21	20	250
From Prolific (Percent)	35 (94.59%)	55 (88.71%)	30 (83.33%)	87 (92.55%)	5 (62.5%)	16 (76.19%)	16 (80%)	219 (87.6%)
Empowered (Percent)	37 (100%)	61 (98.39%)	0 (0%)	1 (1.06%)	0 (0%)	13 (61.9%)	7 (35%)	110 (44%)
Appreciation	4.63 (1.4)	4.67 (1.54)	3 (1.71)	3.59 (1.61)	2.71 (1.33)	4.98 (1.77)	6.22 (0.85)	4.25 (1.75)
Enjoyment	6.61 (0.49)	6.52 (0.64)	3.58 (1.85)	4.23 (1.99)	3.62 (2.04)	5.95 (1.33)	6.48 (0.7)	5.31 (1.88)
BPN								
Autonomy Satisfaction	5.99 (0.98)	5.46 (1.2)	2.7 (1.34)	3.28 (1.68)	3.33 (1.86)	4.63 (1.6)	4.65 (1.96)	4.34 (1.87)
Autonomy Frustration	2.58 (1.44)	3.04 (1.39)	4.76 (1.27)	4.5 (1.41)	3.83 (1.94)	3.37 (1.66)	3.77 (1.8)	3.77 (1.61)
Competence Satisfaction	6.31 (0.81)	6.05 (0.95)	2.32 (1.62)	4 (1.83)	3.83 (1.75)	4.95 (1.69)	4.93 (1.4)	4.79 (1.9)
Competence Frustration	1.89 (0.87)	2.51 (1.44)	5.79 (1.3)	4.41 (1.65)	3.88 (1.5)	3.49 (1.7)	3.33 (1.56)	3.58 (1.85)
Relatedness Satisfaction	4.95 (1.67)	4.09 (1.73)	3.18 (1.57)	3.71 (1.86)	2.54 (1.1)	5.13 (1.69)	5.7 (1.31)	4.16 (1.84)
Relatedness Frustration	2.12 (1.11)	2.12 (1.09)	3.89 (1.49)	3.79 (1.36)	4.63 (1.55)	3.24 (1.46)	3.8 (1.42)	3.18 (1.52)
PXI								
Autonomy	6.12 (0.94)	5.55 (1.25)	3.13 (1.64)	3.67 (1.75)	4.46 (1.46)	4.94 (1.64)	5.03 (1.65)	4.67 (1.81)
Audiovisual Appeal	6.3 (1.07)	6.24 (0.9)	5.14 (1.43)	5.37 (1.41)	5.12 (1.26)	6.25 (0.93)	6.72 (0.44)	5.85 (1.25)
Challenge	5.25 (1.02)	4.82 (1.17)	2.97 (1.54)	3.96 (1.66)	4.33 (0.98)	5.06 (1.29)	5.2 (0.95)	4.47 (1.51)
Clarity of Goals	6.45 (0.52)	6.17 (0.9)	5.13 (1.76)	5.77 (1.14)	6.58 (0.58)	6.08 (1.08)	5.93 (1.05)	5.93 (1.17)
Curiosity	5.5 (1.58)	5.48 (1.27)	3.88 (1.94)	4.68 (1.73)	4.29 (1.44)	5.67 (1.44)	6.6 (0.58)	5.18 (1.63)
Ease of Control	6.02 (0.85)	5.72 (1)	4.48 (1.43)	5.06 (1.32)	5.5 (1.1)	5.9 (0.65)	5.55 (0.87)	5.38 (1.22)
Immersion	5.73 (0.83)	5.81 (0.97)	4.72 (1.66)	5.06 (1.2)	4.42 (1.55)	5.52 (1.44)	5.95 (0.94)	5.39 (1.22)
Meaning	5.76 (1.12)	5.89 (0.97)	3.53 (1.69)	4.43 (1.64)	4.42 (1.19)	5.78 (1.29)	6.33 (0.83)	5.12 (1.58)
Mastery	6.26 (0.72)	6.02 (1.02)	2.47 (1.65)	4 (1.86)	4.25 (1.81)	5.03 (1.55)	5.1 (1.34)	4.83 (1.86)
Progress Feedback	5.67 (1.17)	5.53 (1.16)	4.73 (1.52)	4.93 (1.35)	5.17 (1.14)	5.57 (0.96)	4.65 (1.18)	5.16 (1.33)
GSAQ								
Internality	5.49 (0.84)	5.58 (1.02)	4.65 (1.24)	4.1 (1.38)	4.44 (1.44)	4.54 (1.03)	5.2 (1.08)	4.83 (1.33)
Stability	4.92 (1.05)	4.91 (0.95)	3.7 (1.28)	4.2 (1.44)	4.38 (1.7)	4.95 (1.48)	4.58 (1.05)	4.47 (1.32)
Globality	3.73 (1.43)	4.05 (1.27)	3.03 (1.31)	3.32 (1.29)	3.42 (0.97)	4.38 (1.25)	3.6 (1.37)	3.61 (1.32)
Controlability	6.14 (0.7)	6.09 (0.94)	5.47 (1.3)	5.11 (1.51)	5.75 (0.79)	5.33 (0.87)	5.82 (1.04)	5.6 (1.26)

A unifying aspect of this theme is that participants felt confident that they could deal with any threat, describing a *state* of certainty and seeming invulnerability: *"no matter what my friends tried they could not win. Even when the had amazing teams and I had poor ones. I was unbeatable"* (ID 460; M; 40; FIFA 20). This state of empowerment — feeling powerful, in control, having resources or feeling in charge — enabled participants to act with impunity: *"It made me feel empowered because i was better off then the other player so i was able to kill him if i had liked"* (ID 492; M; 37; DayZ). Participants enjoyed *"the amount of choice I had in the story as well as how I developed my character"* (ID 823; M; 34; Dragon Age Series) or being able to *"buy whichever truck I wanted and choose the type of cargo I want to carry* (ID 443; M; 32; Euro Truck Simulator 2). Other participants emphasized how thanks to their abilities, they felt empowered through social capital, for example, when being approached for advice by other players — *"It was the time when I started playing the game much later than my friends but I still catch them up and [...] after some time they started to ask me for a opinion about how to play"* (ID 423; M; 31; Summoners War) — or being a preferred teammate — *"i am the best of my friends and that gives me a power that everyone wants to duo with me" (ID 515; F; 28; League of Legends)*. Consequently, participants reported feeling respected, wanted, and needed: *"I'd love to have all this great power and respect in my life. It would be something out of the ordinary and I would feel greatly empowered. I'm the leader and I'm going to save the world."* (ID 497; M; 36; Tales of

Arise).

Heroic Victory

This theme captures experiences of success preceded by considerable struggle. Situations which tested participants' skill and determination, but where they eventually came out on top in the end, win the tournament, and save the day. These experiences are characterized by short, intense bursts of stress and anxiety, leading up to a moment of success that brought sudden relief, and celebration of their achievement by and with others.

> "We were playing a competitive match [...] against one of the top teams in the scene. We were a good team, but in comparison, we never expected to win. However, the match was really nail biting, and I was performing well and carrying the team. In the last and deciding round, I made some strategic calls during the round and won a final 1v1 standoff against our opponents, securing the win for our team. Even our opponents and people who were spectating the match congratulated me on my performance afterwards." (ID 465; M; 22; Mount & Blade II: Bannerlord)

These experiences could take the form of an archetypal last stand, where the player was the underdog, outnumbered and seemingly outclassed. Yet, because of their heroic actions, they managed to turn the experience around. A common example for this would be a "clutch" or "ace": *"I managed to kill all 5 members of the opposite team by myself, when all of my other team members where dead, so the responsibility (and pressure) fell on me."* (ID629; M; 37; Rainbow Six Siege).

Feeling in control was characteristic of *The Joys of Control*, certain and stable. In the Heroic Victory theme, in contrast, success and control are hard-fought and at first highly uncertain. This could be a victory against all odds as described above, or a long and arduous path towards success — i.e., running into a boss over and over again until finally succeeding. This latter kind of experience was characterized by a long state of frustration with little progress until the player finally succeeded, making the struggle ultimately worthwhile:

> "I died tens of times, without getting even close to defeat. I almost broke the controller by hitting it on the wall. I dropped the game (that I liked) about a week only to return and kill the [Capra Demon ...] The game proceeded to become my favourite of all time." (ID 813; M; 39; Dark Souls)

Often in these experiences, what made the frustration bearable was the player's awareness that, despite failure, they continued to make progress. They remained motivated by this, *"cause its shows that my effort its paying off. those moments its what keep me gaming and playing even when i have bad runs. are those moments that make me realize that i am geting the sense of the game"* (ID

449; M; 28; The Cycle Frontier). Once victory was achieved, however, players felt particularly gratified. Participants repeatedly emphasized self-reliance and feelings of pride: *I usually consult tutorials so being able to do it all alone was an interesting project and really empowering* (ID 540; M; 25; Minecraft). This sometimes was accompanied by a sense of poetic justice, such as when another player *"[...] kept ramming people of the street/cutting people off. In the second to last corner I faked going from the inside and he tried to cut me off, but I went left and he crashed. I finished 1st and he finished 4th. I was able to use my skill and beat a dirty racer, and for that, I felt empowered"* (ID 432; M; 22; Forza Horizon 5).

However, despite this emphasis on self-reliance, heroic victories were not entirely individualistic experiences. In fact, almost all these accounts include mention of shared celebration, whether being celebrated by others — *"everyone in my team was congratulating with me for my wonderful performance, writing "you rock" [...]"* (ID 456; F; 29; Smite) — or sharing the victory with friends, creating an important event for the group: *"Whole match was going pretty good for us but we wasnt sure if we can win [...] we literally destroyed enemy team and we won. We felt like gods that moment. I wont forget about that night ever." (ID 510; M; 20; League of Legends)*.

Lastly, in some instances participants reported that their achievements not only felt good in the moment, but that they led to positive real-life consequences. For example, the sense of improvement and achievement *"helped with my mental health and my confidence in real life too."* (ID 462; F; 26; Dark Souls 3).

Being the Worst

When not succeeding in the game, many participants saw themselves at fault. In particular, this theme focuses on players' perceived lack of proficiency to succeed and consequently finding themselves outclassed by or embarrassed in front of their peers: *"I tried to play a famous e-sports shooter with my brother. [... It] was hard to master at my novice level. I always bottom fragged, i.e, i was always the bottom one in the scoreboard" (ID 484; M; 32; Counter Strike: Global Offensive)*.

Participant descriptions frequently referred to feelings of inferiority, inadequacy, and lack of efficacy. Their actions had no tangible effect on the world, leaving them feeling that "[i]f i were to leave in the last 20 minutes of game there would be no change in the outcome." (ID 487; M; 20; League of Legends). One participant described feeling *"completely useless and powerless, not being able to do anything meaningful and I practically just end up watching everyone else play the game around me"* (ID 840; M; 21; League Of Legends).

The feeling that one's actions had no tangible impact was especially disheartening in multiplayer gaming. One participant reported wanting *"to participate and be a relevant member of the team. If i can't shoot the enemies, i can't help the team win games"*(ID 484; M; 32; CS:GO). Another

reported:

> "It was me being the reason why we were loosing, [it] felt horrible to make those mistakes and then being blamed for them by my team, especially because this time it was my fault. I felt that I had no impact on the game or maybe even was helping the enemy team more than our team [...]" (ID 519; M; 20; League Of Legends)

Participants in this theme saw themselves at fault for their bad performance, where some directly blamed their own inattentiveness — *"The character I played was objectively powerful, but a simple mistake still managed to ruin everything I worked for.* (ID 507; M; 24; The Elder Scrolls V: Skyrim) — or skill — *"[I]t was clear that my skill level was below what it takes to complete certain parts of the game and made me feel discouraged to play at all."* (ID 810; F; 27; Ori and the Will of the Wisps). Notably, although participants reported failing at a task, or not contributing value, they retained the feeling that they could overcome and — in principle — succeed: *"The boss was unbeatable. I know it wasn't impossible as the game is pretty popular, so I knew it was my fault. Did I miss something? [...] Am I just bad at this game? Why can I not get past this boss?"* (ID 61; M; 19; Borderlands 2).

As participants felt to blame, some reported a loss of confidence: *"I'm decent at most videogames. And seen so many people completing this game, but I wasn't able to."* (ID 436; M; 28; Hollow Knight). As a consequence, participants started *"to doubt myself as I regarded myself as a very good player."* (ID 768; M; 50; Gran Turismo) or considered themselves *"not "worthy" to play with those teammates"* (ID 414; M; 40; Overwatch 2). Confronted with the fact that "[e]very time I would try to do an action I would be punished or otherwise not able to get through." (ID 36; M; 27; Dark Souls 3), several participants reported quitting the game: *"[it] got me feeling like a looser [...] I decided to pause a bit on it and focus on other interesting video games. Then I started playing other videogames."* (ID 403; M; 23; God of War).

Out of Their Hands

This theme relates to participants describing experiences in which they did their best, but still failed for reasons that were 'out of their hands', i.e., participants felt disheartened, treated unfairly, and had no chance to turn the situation in their favor. *"[It] was simply unfair and I had no control over the situation.*" (ID 805; M; 21; Escape From Tarkov). At the core of this theme was a feeling of not being in control of the situation. In this sense, these experiences pose a counterpoint to the *Joys of Control* theme.

One participant, for example, reported feeling hopeless in situations in which *"no matter how much I would try to replay the match, the game was already set that for that game I have to lose"* (ID 692; M; 26; FIFA 16). Likewise, participants disliked having to rely on others, as *"You cannot control what other players do"* (ID 818; M; 30, Lost Ark).

Oftentimes, these experiences left participants feeling let down. As one participant noted, *"i felt really disempowered because all of my choices, which this game is literally made of, didn't really matter"* (ID791; F; 32; Life is Strange). In other cases, participants felt disappointed as the game did not provide the experience they had hoped for or felt they had been promised. After looking forward to a game's release, one participant found that *"the game was full of bugs and it didntbwork properly. I felt cheated. So much so i returned it. All that excitement for nothing"* (ID 808; M; 24; Cyberpunk 2077). Another participant reported feeling let down by a game that turned out to be more difficult than expected, leading them to conclude that *"[...] this was designed very badly in my opinion. I was interested in the story but I wanted to enjoy the game, not suffering"* (ID 764; M; 31; Mortal Shell).

Another important distinction between themes is that the *Joys of Control* were about players attributing the success internally, whereas "Out of Their Hands" experiences usually centered external causes. In some cases, frustration stemmed from participants believing that they had played well, and should have won, if only — for example — other players had pulled their weight: *" [...] you give your maximum, you try as hard as possible, you look through different guides to get better, but that might only be you"* (ID 535; M; 19; League of Legends). One participant felt disempowered when other players ignored their advice, feeling they would have won if only they had received the respect and compliance they expected:

> "I may not be the best player but I feel like im a really good one and can direct other players into doing good strategies to win over multiplayer competitive games, but If they dont want to listen, thats not up to me." (ID 440; M; 23; Apex Legends)

Participants also felt that the situation was 'Out of their hands', when they were confronted with abusive language or sabotaging behavior (griefing), but felt they had no recourse against the harassment:

> "The enemies being toxic exacerbated the negative experience as it felt like they were rubbing in our faces that we were worse even though we didn't even do anything to incite such toxicity. It felt as if someone was damaging my ego but I couldn't do anything to them because they were anonymous." (ID 480; M; 32; Valorant)

The perpetrators of such behavior were not only opponents, but often teammates. One participant described how *"people in game were saying I was bad, almost insult someone they don't know just because they felt superior [...]"* (ID 583; F; 28;). Especially when a game disincentivized players from leaving early, participants felt forced to endure the abuse: *"I couldn't enjoy the game, yet I had to stay and play the game to avoid punishment"* (ID 841; M; 23; League Of Legends). Crucially, participants believed the developers to be at fault, lameting, for example, that teammates *"[...]*

do whatever they want without any consequences cuz the game devs don't give an actual shit about the players" (ID 526; M; 25; League of Legends).

Some participants also blamed developers for sacrificing gameplay in favor of monetization. As one participant wrote, *"everything was setup against me and to abuse my enjoyment of the game for the monetary gain of the company and that alone. everything was just so blatantly exploitative [...]"* (ID 839; M; 25; YuGiOh Duel Links). Other participants noted that *"the game currently feels like it wasn't made for us to have fun anymore just feels like it is just to make money"*(ID 815; M; 22; FIFA 22). Other participant noted that bugs and glitches rendered the game unplayable unless the developers fix them:*"I can't take any actions fixing what is wrong with it or getting someone from the programmers of the game to do something about it"* (ID 819; F; 27; FIFA 22).

Sexist Harassment

This theme captures the experiences of participants who suffered griefing, insults and harassment, because they had been identified as women.

> "I have a feminine username and play female characters and roles a lot and sometimes there's people who notice and call me names because of it. So there was this guy who kept laughing at me in chat every time I made mistakes, calling me all sorts of derogatory names and telling me things like 'go back to the kitchen'. I muted him but it still hurt a lot and made me feel like I can never be a part of this community unless I pretend I'm not a woman." (ID 481; F; 25; League of Legend)

While other themes also featured instances of griefing and harassment, accounts in this theme were about participants becoming aware that they were targeted based on their gender. Several participants described that the harassment started as soon as they had been identified as women: *"Someone in the voicechat noticed that I'm a woman and startet berating and attacking me pre and during the game"* (ID 49; F; 27; Overwatch). Moreover, one participant emphasized that she was harassed specifically*"[b]ecause I was female doing well and the other team felt the need to gang up on me because of that and not let me play properly"* (ID 775; F; 29; Call of Duty).

As a result of the targeted attacks, several participants noted that they largely withdrew from online spaces or had to hide their gender identity. One participant wrote: *"I rarely turn on my mic because of that, and after each experience with horrible, sexist people leaves me feeling disempowered and disgusted. "[...] I do not deserve to be treated this way only because I do not have a penis"* (ID 43; F; 25; Valorant). Another wrote *"I can never be a part of this community unless I pretend I'm not a woman"* (ID 481; F; 25; League Of Legends).

Seeing Themselves in the Game

This theme relates to experiences in which participants could relate their personal experiences to the game narrative, or saw something in the game as representative of aspects of their identity. As noted in the previous theme (*Sexist Harassment*), women repeatedly reported that they did not feel welcome in gaming spaces — facing harassment and exclusion. Against this backdrop, some women described feeling empowered by gaming environments that were more welcoming to them. One reported:

> "Having been harrassed, cajoled, disrespected and discounted because of my gender, to have a game where I could not only enjoy the story, the game play and the world building, but also not feel like, once again, I was immersed in a man's hobby was incredibly empowering." (ID 81; F; 35; Dragon Age: Origins)

Some participants said they specifically enjoyed in-game opportunities to break gender stereotypes which they felt subjected to in everyday life, for example *"women tend to be portrayed in the real world as frail, thin and dumb human beings. They play the role of damselle in distress. [... This game] makes me feel powerful, strong and capable every time I solve a puzzle and reach a new level that was super hard"* (ID 797; F; 31; Tomb Raider).

While noting that representation in videogames has improved, participants still considered the variety of women characters limited, and appreciated instances of positive representation:

> "[W]hile nowadays there is a fair share of female protagonists and people have come to make complex and strong female characters and so on, it's so rare to see women in video games that actually look as strong as they seem to be. Seeing strong muscular women in media always sends a rush of power through me, and that's what I understand by 'empowerment'." (ID 860; F; 27; The Last Of Us: Part 2)

However, for some participants recognizing aspects of themselves in the game was not a positive experience. One participant described feeling disempowered when they recognized themselves in an in-game character, only to be forced by the game to refer to the character as "weird" or a "pervert":

> "I didn't want to pick either of those options. I don't think he's weird or a pervert, he's just a bit different. I'm autistic and so it felt like I was being insulted by the developers, like if they met me would they think I was a weird pervert?" (ID 41; NB; 38; Persona 5)

Other participants found it disempowering to perceive aspects of themselves in the game which they associated with personal difficulties and fears. One participant reported that they regularly felt disempowered by games that offered rigid and limited character customization

options, as this left them looking at a body which *"reminds me of my dysphoria and that I'm stuck in a body that's not mine. I hate it."* (ID 96; M; 30; Any game with a limited character creator). Yet other participants reported discomfort over being confronted with their aging process, which prompted fears that *"I wouldnt be able to keep up the pace I was used & maybe start not enjoying games so much"* (ID 466; M; 41; Ninja Gaiden). Similarly, one participant considered their game experience as:

> "an allegory for real life, ie I couldn't expect to compete with younger people in terms of stamina, reactions and also in the job market. It also made me felt that I didn't b elong any more in the world of online gaming, because I wasn't good enough anymore." (ID 477; M; 54; Call Of Duty: Black Ops 3)

Understanding the Message

This theme focuses on experiences where participants interpreted and related what they perceived as the intended message of the game. At its most basic, "understanding the message" could involve a moment of epiphany, where players came to appreciate the craft or elegance involved in a game's design. When describing a puzzle in Baba is You, for example, one participant mentioned: *"[it] was really clever and yet simple at the same time so finding that solution was very rewarding, because of 'ohhhhh now I understand, that's cool' moment."* (ID 850; 39; M; Baba is You).

Several accounts focused on narrative or character elements, where players described a sense of closeness and shared experience with the player-character, or personally related to existential questions raised in the game. One participant wrote:

> "[I] really liked to start as a robot, trying to play as if I were a robot and what I would do, until a point I felt the need to take more emotional decisions in the game, and noticed myself unconsciously starting to make more human decisions pretty much until the end where I noticed I was kinda wanting the robot to succeed and live a normal human life" (ID 395; M; 26; Detroit: Become Human)

Even emotionally difficult choices could lead to a feeling of empowerment, as players experienced the opportunity to act as the character and support them in their journey:

> "During an emotionally driven part of the story, there was the choice of either [...] sparing the innocent guards and scientists, or going forward with violence which based on the rest of the story would have been justified in the eyes of the protagonist. Having the ability to make this emotionally driven decision made me feel empowered." (ID 488; M; 23; The Last of Us)

Not all such moments of understanding were reported in the empowerment condition, however. Some instances were reported in the disempowerment conditions, where players came to appreciate the experience if they understood the underlying reasons for why the game made them feel disempowered. For instance, when the player-character encountered plot-dictated hardships and the player realized that *"Fighting back is useless, and eventually you'll be captured [...I]t feels disempowering because it contributes to the theme of being captured, cutting you off from the game temporarily"* (ID 767; M; 21; Far Cry 5). This appreciation for disempowerement could also follow from game mechanics and skillful elements, if they were perceived to support the themes of the game. In such cases even repeated failure could be appreciated, insofar as it *"helped me understand the message that the game was trying to tell me, and improved the experience when I finally defeated the thing"* (ID 451; M; 25; Elden Ring).

Likewise, participants appreciated restrictions on their freedom and ability to act, if they understood the underlying directorial intentions and felt they were consistent with the themes of the game:

> "The game [...] limits your ability to traverse easily and quickly. In most games, this would be considered a shortcoming. However, the game justifies this by making it the game's main point, mechanically and thematically. [...] I played it at the beginning of the covid pandemic lockdowns. It hit home, as the game is about connecting isolated people in a world that just ended." (ID 155; M; 36; Death Stranding)

In some of these experiences, players felt that the game forced them into bad decisions. In this respect they can be directly and meaningfully contrasted with experiences grouped under *Out of Their Hands*. Both themes involve a lack of (good) choices. However, in *Out of Their Hands*, this was almost always perceived negatively. In *Understanding the Message*, by contrast, participants were able to understand and appreciate the hard choices or bad situations as potent aspects of the game: *"I felt disempowered and many more feelings certainly because the game was so well made [...] enough to make me feel what they wanted to make me feel"* (ID 494; M; 33; Disco Elysium).

III.4.4 PX Patterns among Themes

Differences on PX measures between themes offer further information on the different aspects of dis-/empowerment experienced by participants. A numerical overview is provided in Table III.3 and visualizations are provided in Chapter A. Note that these findings are purely descriptive, and do not allow for causal inferences.

As noted above, empowering experiences tended to be need-satisfying overall, while disempowering experiences were less so. Within the thematic groupings, this pattern held for *Joys of Control* and *Heroic Victory* (both > 98% empowering), *Being the Worst*, and *Out of Their*

Hands (both > 98% disempowering), with *Being the Worst* showing the highest need frustration of any thematic group (Table III.3). Notably, while the *Understanding the Message* theme also occurred among disempowering experiences, these experiences were associated with the highest relatedness-satisfaction: perhaps indicating a felt connection to the in-game characters, empathy towards non-player characters or appreciation of the developers and writers. However, in interpreting results it should be noted that the Relatedness scale asked about other "characters", not other *people*, which might have resulted in the questions being more meaningful for the type of single-player experiences found in the *Understanding The Message* theme (see section III.5.3).

Turning to the PXI, scores for all groups were above the middle of the scale, indicating that the experience was meaningful, supported mastery, etc.. A notable exception to this was the challenge scale, with the *Being the Worst* theme exhibiting particularly low scores ($M(SD) = 2.47(1.65)$), indicating that these participants experienced a notable mismatch between their skill and the challenge the game posed. This was consistent with players' descriptive accounts of their poor performance. Lastly, there was generally medium to high agreement across all four dimensions of the Game Specific Attribution Questionnaire, with globality (whether the cause of the experience applies across multiple situations or circumstances) scoring the lowest (Empowered: $M(SD) = 4(1.27)$; Disempowered: $M(SD) = 3.3(1.28)$). The exception to this was the *Seeing Themselves in the Game* theme, which scored notably higher on globality ($M(SD) = 4.38(1.25)$), suggesting that these kind of experiences involve extrapolation from the gaming experience to a wider context.

III.5 Discussion

While empowerment has been considered an important issue in HCI and HCI games research, previous work on empowerment in these fields has not addressed the experience of dis-/empowerment, and data on players' own experiences of these phenomena are lacking. Our goal in this study was to provide an foundation for understanding dis-/empowerment in terms of actual player experiences.

Previous accounts of empowerment, both in industry and HCI research, give the impression that empowerment in games is uncomplicatedly positive. Schneider et. al's suggestion that empowering users is almost always treated as an "unconditionally positive mission" [72], and Warren Spector recent description of empowerment as "the ultimate success criterion" [78]. This might leave us with the impression that disempowerment, in turn, must be unconditionally negative. Our results, however, reveal a more varied picture. Overall, empowering experiences were experienced more positively than disempowering experiences. In some themes this pat-

tern was particularly strong, such as in the straightforwardly empowering *Joys of Control* theme, where PX measures showed high enjoyment and needs satisfaction. The two most consistently disempowering themes — (*Sexist Harassment*, and *Being the Worst*) — were characterized by low PX scores, and high needs frustration. However, this is only part of the picture. Consistently empowering themes like *Joys of Control* did not score highest for appreciation. Both *Understanding the Message* and *Seeing Themselves in the Game* contained large proportions of disempowering experiences, but were associated with positive PX. Average scores for relatedness satisfaction, meaning, appreciation and enjoyment in these groups were among the highest theme-averages, higher than averages for empowering experiences overall. Context mattered considerably in how dis-/empowerment was perceived, and how it contributed to player experience.

We also found that dis-/empowering experiences were surprisingly diverse. Previous work has suggested that empowerment has a "simple inherent meaning" [91, p. 2], yet we found that empowerment could mean quite different things in different contexts. While previous work has focused on empowerment in terms of players' in-game abilities, or efficacy, we found that players' own accounts of both empowering and disempowering experiences often focused other issues: factors such as identification, understanding, and social inclusion or exclusion.

Our results suggest that the goal of empowering the player must be more carefully unpacked. Straightforward empowerment in terms of maximizing player efficacy should not always be considered the highest goal. Nor should empowerment be narrowly associated with in-game abilities, nor even with power and achievement more generally. In the following, we articulate which factors were important in differentiating experiences of dis-/empowerment in our results. We then discuss the open questions raised by these results, and how they relate to prior work in psychology and PX research, supporting future work on the player experience of dis-/empowerment.

III.5.1 Beyond Efficacy: Ownership, Social Factors, and Understanding

Many accounts of empowerment in HCI games research concern in-game efficacy [e.g., 31, 43]: success, influence and control. Other work focused on supporting perceived efficacy by balancing competition between players [27, 29]. Our results did not contradict the idea that in-game efficacy could be an important aspect of empowerment experience, but they also showcased that it is not forcibly the definitive factor which influenced experiences of empowerment: Player accounts often placed more emphasis on contextual factors. Our results suggest that experiences and outcomes of dis-/empowerment were distinguished by players' sense of ownership of the outcome (or the perceived source of efficacy), social and inter-personal factors around

the play-experience, and players' understanding of the game and its meaning. Participants' PX ratings lend further credence to the importance of contextual factors over efficacy: First, scores for controllability on the Game Specific Attribution Questionnaire were very high across groups (Table III.1), even for the disempowering *Out of Their Hands* theme, which focused on the lack of control. Second, results on the PXI (which focuses on the grounding of high-level player experiences in low-level player actions in game) were not strongly differentiated between empowering and disempowering experiences. This suggests that dis-/empowerment experience may be more related to to *higher-level experiences* [1], than *low-level player actions*.

Some of our themes did show a strong association between efficacy and dis-/empowerment — *The Joys of Control* theme was defined by positive experiences of control, and the disempowering *Being the Worst* theme focused on perceived failure and lack of ability. However, even in these cases we found that it was not only the *degree* of efficacy which was important in characterizing experiences of dis-/empowerment, but also the perceived *source* of efficacy. Meanwhile, *Being the Worst* and *Out of Their Hands* both contained high percentages of disempowering experiences, and both themes centered on lack of efficacy. However, they were also differentiated by other factors. *Being the Worst* focused on perception of *internal* efficacy, with players ascribing the failure or lack of control to themselves, while *Out of Their Hands* players ascribed problems and failures to external factors beyond their control. This difference seems to have impacted on needs satisfaction also, with the two groups having quite different scores for competence satisfaction and frustration. These results are notable, as previous work has tended to focus on the efficacy of designers to empower the user [27, 29] — an external source of efficacy, and not on players' own involvement in empowerment. Future work might further explore the impact of internal and external sources of efficacy in dis-/empowerment experience.

III.5.2 What conditions foster and shape experiences of dis-/empowerment

Our findings indicate that increases or decreases in players' in-game efficacy may not always, by themselves, be a reliable guide to whether an experience will be perceived as empowering or disempowering. Further, whether an experience is empowering, or disempowering, may not always by itself be a reliable guide to whether it will be experienced positively, or negatively. This complex landscape of experiences raises several questions for PX research: when and how designers should seek to empower players? Should designers sometimes seek to disempower players — when is this acceptable, and how can it be achieved in a positive way? Since our work represents the first groundwork on players' dis-/empowerment experiences we can only offer provisional answers to these questions. However, our results point to avenues for future investigation, and to resources in prior research which might ground and guide this investigation.

We found that empowerment experiences centered on the sense of achievement, ability and were associated with enjoyment, though only moderate appreciation. This indicates that, at least for certain kinds of play experiences, the goal of empowering players, and the approach of increasing perception of in-game efficacy, remain valid. Our results suggest that such experiences can be amplified by opportunities for players to celebrate each other.

However, such approaches for supporting experiences of empowerment seem likely to feel a little "one-note". In our results, less direct routes to empowerment outcomes were described and enjoyed by a large number of players, as evidenced by instances of comparably high scores for enjoyment and appreciation where success and a sense of achievement followed a period of struggle, or even frustration and doubt (*Heroic victories*). This theme most closely resembles Juul's paradox of failure [35] and Frommel et al's notion of temporary failure [25]. Appreciation of struggle — sometimes without much focus on a final success — was also found in the largely positive, but both dis- and empowering experiences in *Understanding the Message*. This raises the question: how are such experiences of struggle and frustration differentiated from the largely negative and disempowering experiences described in *Being the Worst* and *Out of Their Hands*. Various factors may be at play here, including whether players feel they are themselves responsible for outcomes and to what extent their sense of disempowerment aligns with their understanding of the game's themes and narrative. However, details of how these factors interact, and how they can be effectively fostered remains an open question for future work.

Our study also addressed experiences of disempowerment — both negative and positive. First, it seems obvious that designers will wish to minimize straightforward forms of negative disempowerment that are due to permanent in-game failure [25], or abuse by others. Such issues and their potential solutions are well documented in previous research [15, 9, 42, 4, 39].

Moreover, our results complement Frommel et al's [25] findings on failure experiences, which were characterized by lower enjoyment, competence, and more likely to be attributed to external and uncontrollable causes. These experiences share some semblance with our *Out of Their Hands* theme, which likewise was characterized by player feelings of helplessness due to external circumstances such as other players. The *Being the Worst* theme, in contrast, centered on failures where participants not only felt a lack of competence, but where they considered themselves responsible for the negative outcome.

That said, our results also indicate that there may be value in supporting positive experiences of disempowerment. In player descriptions, such experiences were associated with a sense of meaning, connection to characters, and reflection on the game and their wider life. PX scores for such experiences showed high levels of appreciation, enjoyment, immersion and needs satisfaction — in particular relatedness-satisfaction was considerably higher than averages for most other themes. Our findings around positive disempowerment further resonate with previous

research on positive experiences of negative emotion [12, 49, 13], where reflectiveness, and the relation to personal memories can positively influence the appreciation of negative emotional experiences in games [12, 14]. This aligns with our own finding that it was insight and understanding that often seemed to distinguish positive experiences of disempowerment from negative. Beyond this agreement, however, our results highlights open questions for research on positive response to adverse or negative game-experiences. For example, what is the relevant content of such insight and understanding: Should it concern particular aspects of the game? The context around play? Issues in the player's wider life, and their relation to the game? All of these showed up in our player descriptions, but it is unclear which were most influential, and under what circumstances. Previous work also leaves open other questions: what conditions foster such understanding, or make space for players to arrive at it by themselves [see also 22]?

One line of research responding to such questions is suggested by recent work by Ballou and Deterding [5]. They argue that the enjoyment of need-frustrating experiences depended on players' expectations: frustration which exceeds levels anticipated by the players will result in needs frustration and negative experience. This resonates with our observation that players appreciated struggle which was congruent with their perceived design intention. A potentially important subtlety here is that Ballou et al. emphasize *expectation*. This suggests that users should arrive at this understanding *ahead* of the experience. Understanding may function differently if it arises in reflection, after the fact. Future work might investigate how experiential outcomes are influenced both by player expectations going into disempowering videogame experiences, and by understandings arrived at after the fact.

Elsewhere Bopp et al. have suggested that there might be skill in facing emotional challenge [13], which concerns the capacity to reflect and create meaning, and that this might support positive responses to adverse gaming experiences. This perspective must be applied carefully to issues of disempowerment: it is clear that some experiences described by our participants — for example those of harassment — were serious and likely systemic. Focusing on emotional "skill" in such cases would be trivializing and beside the point. But in less serious and personal cases, future research may investigate whether aspects of "emotional skill" may explain differences in player responses to negative experiences in games. This may influence whether experiences are perceived as disempowering, and in turn whether this disempowerment is experienced positively or negatively. Experimental work might operationalize emotional skill in terms of psychological constructs such as affective clarity [45, 10], and goal self-concordance [73]. Such work might inform future research on therapeutic games and research on risk factors for negative well-being outcomes in videogame play [47, 60].

III.5.3 Limitations and Future Work

This study sought to explore empowering and disempowering experiences in videogames, making use of thematic analysis alongside descriptive statistics. Thematic analysis offers a structured approach to inductive analysis that does not require prior hypotheses. It can contribute to understanding in many ways, one of which is to provide structure for exploratory work addressing new questions. However, insofar as it involves a process of active interpretation by the researchers it does not support causal claims. As such the relation of our themes to quantitative PX measures must be read with caution. Since both systematic investigation and conceptualization of dis-/empowerment experiences is lacking at present, hypothesis testing would not have been appropriate, and descriptive ground work is called for [70]. Rather than providing evidence for strong claims, the findings described here serve as a preliminary map for coordinating future work, a source of potential hypotheses, and (following prior work on similarly complex experiential concepts) a frame for clarifying commonalities and divergences between different accounts of dis-/empowerment [8]

Secondly, most reported experiences concerned multiplayer games — a contrast with similar prior work [e.g., 11, 92]. Separating by our data sources (Reddit and Prolific) indicated that participants who took part out of interest were more likely to describe single player games. No other noteworthy differences were found between the groups (see data in our OSF repository). Future studies might consider the impact of data sources, since type of game and player may influence the kind of experiences represented. Interested parties might utilize our shared data for further analyses.

Third, caution is also required when interpreting the scales for relatedness satisfaction and frustration. Questions for these referred to "characters" (e.g. "I felt close to a character."). Given the prevalence of multiplayer games, and reports which did not reference in-game characters, these questions were less relevant to the experiences of some participants. Future studies should be careful in selecting relatedness scales based on expected experiences.

Fourth, our work addresses single episodes of play, and not dis-/empowerment over longer time scales. It is an open question whether dis-/empowerment experience might express itself differently over longer periods. For example long experiences of being over-powered, or lacking credible threat might become boring and even disempowering [see 35]. Equally, constant success may alienate players from peers who regularly lose, leading to a different kind of disempowerment. These are open questions for future research. Answering them could provide valuable insight into factors that support longer term engagement.

Finally the only form of targeted identity-based harassment reported in our dataset was gender-based, against women (see *Sexist Harassment*). This is not evidence that sexism is the

only form of identity-based harassment in videogame spaces. One participant in our dataset did mention racial slurs but their description indicated that this was not related to their own identity (ID 480). Other forms of identity-based harassment were not reported. One explanation for this might follow from the larger number of multi-player games in our dataset. Voice-chat communication in these games may result in gender being more easily identifiable than other identity factors. Future work might address how identity group membership impacts on experiences of dis-/empowerment.

III.6 Conclusion

Although empowerment is considered a desirable design goal in HCI and games research, little is known about what constitutes dis-/empowering player experiences. Our analysis of 250 player accounts revealed that experiences of empowerment take in states of high power, transient experiences of celebration after long stretches of struggle, and even moments of understanding of narrative or design intent. We found that disempowerment in gaming is a familiar experience for players, as often related to the actions of other players as to in-game events. Interestingly, disempowerment of the player was not always associated with negative experiences: in the right circumstances it could result in experiences that were as much enjoyed, and even more appreciated, than some empowering experiences. We found that both kinds of experience could be related meaningfully to prior work on emotional challenge and frustration, while also pointing to gaps in such work, and new ways forward. As such, our work offers HCI games research a player-centered starting point to develop dis-/empowerment as a dimension of the player experience: providing a more nuanced view of the concept and how it relates to past works.

Data Availability Statement

De-identified raw data and all analysis scripts available (CC BY 4.0) at https://osf.io/zhtu8/. In case of re-use, please cite as [90].

Author Contributions

JBV: Conceptualization, data curation, formal analysis, investigation, methodology, resources, software, visualization, writing — original draft, writing - review & editing.

DB: Formal analysis, validation, writing — original draft, writing — review & editing.
DM: writing — original draft
EDM: Conceptualization, formal analysis, methodology, resources, validation, writing — original draft, writing — review & editing

Acknowledgment

Funded by the European Union (ERC, THEORYCRAFT, 101043198). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

We would like to thank our participants who shared their insightful, often touching experiences with us. We would like to thank Søren Knudsen for giving input on the visualizations.

References

[1] Vero Vanden Abeele, Katta Spiel, Lennart Nacke, Daniel Johnson, and Kathrin Gerling. 2020. Development and validation of the player experience inventory: A scale to measure player experiences at the level of functional and psychosocial consequences. *International Journal of Human Computer Studies* 135, October 2019 (2020), 102370–102370. https://doi.org/10.1016/j.ijhcs.2019.102370 Publisher: Elsevier Ltd.

[2] R M Anderson, M M Funnell, J T Fitzgerald, and D G Marrero. 2000. The Diabetes Empowerment Scale: a measure of psychosocial self-efficacy. *Diabetes Care* 23, 6 (June 2000), 739–743. https://doi.org/10.2337/diacare.23.6.739

[3] Ahmad Azadvar and Alessandro Canossa. 2018. UPEQ: ubisoft perceived experience questionnaire: a self-determination evaluation tool for video games. In *Proceedings of the 13th International Conference on the Foundations of Digital Games (FDG '18)*. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3235765.3235780

[4] Alexander Baldwin, Daniel Johnson, and Peta A. Wyeth. 2014. The effect of multiplayer dynamic difficulty adjustment on the player experience of video games. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. 1489–1494.

[5] Nick Ballou and Sebastian Deterding. 2022. *'I Just Wanted to Get it Over and Done With': A Grounded Theory of Psychological Need Frustration in Video Games*. preprint. PsyArXiv. https://doi.org/10.31234/osf.io/zehgr

[6] Rahul Banerjee, Jason Yip, Kung Jin Lee, and Zoran Popović. 2016. Empowering Children To Rapidly Author Games and Animations Without Writing Code. In *Proceedings of the The 15th International Conference on Interaction Design and Children (IDC '16)*. Association for Computing Machinery, New York, NY, USA, 230–237. https://doi.org/10.1145/2930674.2930688

[7] Steve Benford, Chris Greenhalgh, Gabriella Giannachi, Brendan Walker, Joe Marshall, and Tom Rodden. 2012. Uncomfortable interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. Association for Computing Machinery, New York, NY, USA, 2005–2014. https://doi.org/10.1145/2207676.2208347

[8] Dan Bennett, Oussama Metatla, Anne Roudaut, and Elisa Mekler. 2023. How does HCI Understand Human Autonomy and Agency? https://doi.org/10.1145/3544548.3580651 arXiv:2301.12490 [cs].

[9] Nicole A. Beres, Julian Frommel, Elizabeth Reid, Regan L. Mandryk, and Madison Klarkowski. 2021. Don't you know that you're toxic: Normalization of toxicity in online gaming. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–15.

[10] Matthew Tyler Boden and Howard Berenbaum. 2011. What you are feeling and why: Two distinct types of emotional clarity. *Personality and Individual Differences* 51, 5 (Oct. 2011), 652–656. https://doi.org/10.1016/j.paid.2011.06.009

[11] Julia Bopp, Jan Benjamin Vornhagen, Roosa Piitulainen, Barbara Keller, and Elisa D. Mekler. 2020. GamesAsArt. (July 2020). https://doi.org/10.17605/OSF.IO/RYVT6 Publisher: OSF.

[12] Julia Ayumi Bopp, Elisa D. Mekler, and Klaus Opwis. 2016. Negative Emotion, Positive Experience? Emotionally Moving Moments in Digital Games. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2996–3006. https://doi.org/10.1145/2858036.2858227

[13] Julia Ayumi Bopp, Klaus Opwis, and Elisa D. Mekler. 2018. "An Odd Kind of Pleasure": Differentiating Emotional Challenge in Digital Games. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173615

[14] Julia A Bopp, Jan B Vornhagen, and Elisa D Mekler. 2021. "My Soul Got a Little Bit Cleaner": Art Experience in Videogames. *Proceedings of the ACM on Human-Computer Interaction* 5, CHI PLAY (2021), 19. https://doi.org/10.1145/3474664

[15] Nicholas David Bowman, Daniel Schultheiss, and Christina Schumann. 2012. "I'm Attached, and I'm a Good Guy/Gal!": How Character Attachment Influences Pro- and Anti-Social Motivations to Play Massively Multiplayer Online Role-Playing Games. *Cyberpsychology, Behavior, and Social Networking* 15, 3 (March 2012), 169–174. https://doi.org/10.1089/cyber.2011.0311 Publisher: Mary Ann Liebert, Inc., publishers.

[16] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[17] Virginia Braun and Victoria Clarke. 2013. *Successful qualitative research: a practical guide for beginners*. SAGE, Los Angeles. OCLC: ocn811733656.

[18] Virginia Braun and Victoria Clarke. 2020. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology* (Aug. 2020), 1–25. https://doi.org/10.1080/14780887.2020.1769238

[19] Philip Brey. 2008. The Technological Construction of Social Power. *Social Epistemology* 22, 1 (Jan. 2008), 71–95. https://doi.org/10.1080/02691720701773551 Publisher: Routledge _eprint: https://doi.org/10.1080/02691720701773551.

[20] Beiwen Chen, Maarten Vansteenkiste, Wim Beyers, Liesbet Boone, Edward L. Deci, Jolene Van der Kaap-Deeder, Bart Duriez, Willy Lens, Lennia Matos, Athanasios Mouratidis, Richard M. Ryan, Kennon M. Sheldon, Bart Soenens, Stijn Van Petegem, and Joke Verstuyf. [n. d.]. Basic psychological need satisfaction, need frustration, and need strength across four cultures. 39, 2 ([n. d.]), 216–236. https://doi.org/10.1007/s11031-014-9450-1

[21] Robert A. Dahl. 1957. The concept of power. *Behavioral Science* 2, 3 (1957), 201–215. https://doi.org/10.1002/bs.3830020303

[22] Alena Denisova, Julia Ayumi Bopp, Thuy Duong Nguyen, and Elisa D Mekler. 2021. "Whatever the emotional experience, it's up to them": Insights from designers of emotionally impactful games. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–9.

[23] Ansgar E. Depping and Regan L. Mandryk. 2017. Why is This Happening to Me? How Player Attribution can Broaden our Understanding of Player Experience. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 1040–1052. https://doi.org/10.1145/3025453.3025648

[24] Eva Patrícia Ribeiro Filipe. 2018. Harnessing Interactive Media Ideological Power: A Disempowerment Model for Video Games. *Proceedings of Play2Learn 2018* (2018), 17.

[25] Julian Frommel, Madison Klarkowski, and Regan L. Mandryk. 2021. The Struggle is Spiel: On Failure and Success in Games. In *The 16th International Conference on the Foundations of Digital Games (FDG) 2021 (FDG'21)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3472538.3472565

[26] Julian Frommel, Cody Phillips, and Regan L. Mandryk. 2021. Gathering Self-Report Data in Games Through NPC Dialogues: Effects on Data Quality, Data Quantity, Player Experience, and Information Intimacy. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3411764.3445411

[27] Kathrin Maria Gerling, Matthew Miller, Regan L. Mandryk, Max Valentin Birk, and Jan David Smeddinck. 2014. Effects of balancing for physical abilities on player performance, experience and self-esteem in exergames. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 2201–2210. https://doi.org/10.1145/2556288.2556963

[28] Chad Phoenix Rose Gowler and Ioanna Iacovides. 2019. "Horror, guilt and shame" – Uncomfortable Experiences in Digital Games. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '19*. ACM Press, New York, New York, USA, 325–337. https://doi.org/10.1145/3311350.3347179

[29] Roland Graf, Pallavi Benawri, Amy E. Whitesall, Dashiell Carichner, Zixuan Li, Michael Nebeling, and Hun Seok Kim. 2019. iGYM: An Interactive Floor Projection System for Inclusive Exergame Environments. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '19)*. Association for Computing Machinery, New York, NY, USA, 31–43. https://doi.org/10.1145/3311350.3347161

[30] Stefanie Größbacher, Peter Judmaier, Lucas Schöffer, Doris Malischnig, Nicole Bilek, and Mylene Kreiger. 2020. All Tomorrow's Parties: Empowerment Game for Young Adults. In

22nd International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '20). Association for Computing Machinery, New York, NY, USA, 1–4. https://doi.org/10.1145/3406324.3410545

[31] Christian Guckelsberger, Christoph Salge, Jeremy Gow, and Paul Cairns. 2017. Predicting Player Experience without the Player.: An Exploratory Study. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '17)*. Association for Computing Machinery, New York, NY, USA, 305–315. https://doi.org/10.1145/3116595.3116631

[32] Weina Jin, Diane Gromala, and Xin Tong. 2015. Serious game for serious disease: Diminishing stigma of depression via game experience. In *2015 IEEE Games Entertainment Media Conference (GEM)*. 1–2. https://doi.org/10.1109/GEM.2015.7377256

[33] Daniel Johnson and John Gardner. 2010. Personality, motivation and video games. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction - OZCHI '10*. ACM Press, Brisbane, Australia, 276. https://doi.org/10.1145/1952222.1952281

[34] Daniel Johnson, M. John Gardner, and Ryan Perry. 2018. Validation of two game experience scales: The Player Experience of Need Satisfaction (PENS) and Game Experience Questionnaire (GEQ). *International Journal of Human-Computer Studies* 118 (Oct. 2018), 38–46. https://doi.org/10.1016/j.ijhcs.2018.05.003

[35] Jesper Juul. 2013. *The Art of Failure: An Essay on the Pain of PLaying Video Games*. The MIT Press. Pages: 173.

[36] Sukran Karaosmanoglu, Katja Rogers, Dennis Wolf, Enrico Rukzio, Frank Steinicke, and Lennart E. Nacke. 2021. Feels like Team Spirit: Biometric and Strategic Interdependence in Asymmetric Multiplayer VR Games. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 443, 15 pages. https://doi.org/10.1145/3411764.3445492

[37] Dominik Kayser, Sebastian Andrea Caesar Perrig, and Florian Brühlmann. 2021. *Measuring Players' Experience of Need Satisfaction in Digital Games: An Analysis of the Factor Structure of the UPEQ*. Technical Report. PsyArXiv. https://doi.org/10.31234/osf.io/zm6gk type: article.

[38] Do Own (Donna) Kim. 2021. "Pay for your choices": Deconstructing neoliberal choice through free-to-play mobile interactive fiction games. *New Media & Society* (May 2021),

14614448211018177. https://doi.org/10.1177/14614448211018177 Publisher: SAGE Publications.

[39] Madison Klarkowski, Daniel Johnson, Peta Wyeth, Mitchell McEwan, Cody Phillips, and Simon Smith. 2016. Operationalising and Evaluating Sub-Optimal and Optimal Play Experiences through Challenge-Skill Manipulation. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5583–5594. https://doi.org/10.1145/2858036.2858563

[40] Alexander S. Klyubin, Daniel Polani, and Chrystopher L. Nehaniv. 2005. All Else Being Equal Be Empowered. In *Advances in Artificial Life (Lecture Notes in Computer Science)*, Mathieu S. Capcarrère, Alex A. Freitas, Peter J. Bentley, Colin G. Johnson, and Jon Timmis (Eds.). Springer, Berlin, Heidelberg, 744–753. https://doi.org/10.1007/11553090_75

[41] Alexander S. Klyubin, Daniel Polani, and Chrystopher L. Nehaniv. 2008. Keep Your Options Open: An Information-Based Driving Principle for Sensorimotor Systems. *PLoS ONE* 3, 12 (Dec. 2008), e4018. https://doi.org/10.1371/journal.pone.0004018

[42] Yubo Kou. 2020. Toxic behaviors in team-based competitive gaming: The case of league of legends. In *Proceedings of the annual symposium on computer-human interaction in play*. 81–92.

[43] Lauri Lehtonen, Maximus D. Kaos, Raine Kajastila, Leo Holsti, Janne Karsisto, Sami Pekkola, Joni Vähämäki, Lassi Vapaakallio, and Perttu Hämäläinen. 2019. Movement Empowerment in a Multiplayer Mixed-Reality Trampoline Game. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '19)*. Association for Computing Machinery, New York, NY, USA, 19–29. https://doi.org/10.1145/3311350.3347181

[44] Donald R. Leslie, Carol M. Holzhalb, and Thomas P. Holland. 1998. Measuring Staff Empowerment: Development of a Worker Empowerment Scale. *Research on Social Work Practice* 8, 2 (March 1998), 212–222. https://doi.org/10.1177/104973159800800205 Publisher: SAGE Publications Inc.

[45] Tanja Lischetzke, Ghislaine Cuccodoro, Anja Gauger, Laure Todeschini, and Michael Eid. 2005. Measuring Affective Clarity Indirectly: Individual Differences in Response Latencies of State. *Emotion* 5, 4 (2005), 431–445. https://doi.org/10.1037/1528-3542.5.4.431

[46] Ylenio Longo, Manuel Alcaraz-Ibáñez, and Alvaro Sicilia. 2018. Evidence supporting need satisfaction and frustration as two distinguishable constructs. *Psicothema* 30, 1 (Feb. 2018), 74–81. https://doi.org/10.7334/psicothema2016.367

[47] Stéphanie Mader, Stéphane Natkin, and Guillaume Levieux. 2012. How to analyse therapeutic games: the player/game/therapy model. In *Entertainment Computing-ICEC 2012: 11th International Conference, ICEC 2012, Bremen, Germany, September 26-29, 2012. Proceedings 11*. Springer, 193–206.

[48] Diego Maiorano, Dishil Shrimankar, Suruchi Thapar-Björkert, and Hans Blomkvist. 2021. Measuring empowerment: Choices, values and norms. *World Development* 138 (Feb. 2021), 105220. https://doi.org/10.1016/j.worlddev.2020.105220

[49] Tim Marsh and Brigid Costello. 2012. Experience in Serious Games: Between Positive and Serious Experience. In *Serious Games Development and Applications*, Minhua Ma, Manuel Fradinho Oliveira, Jannicke Baalsrud Hauge, Heiko Duin, and Klaus-Dieter Thoben (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 255–267.

[50] Tim Marsh and Brigid Costello. 2013. Lingering Serious Experience as Trigger to Raise Awareness, Encourage Reflection and Change Behavior. In *Persuasive Technology*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Shlomo Berkovsky, and Jill Freyne (Eds.). Vol. 7822. Springer Berlin Heidelberg, Berlin, Heidelberg, 116–124. https://doi.org/10.1007/978-3-642-37157-8_15 Series Title: Lecture Notes in Computer Science.

[51] Russell A. Matthews, Wendy Michelle Diaz, and Steven G. Cole. 2003. The organizational empowerment scale. *Personnel Review* 32, 3 (Jan. 2003), 297–318. https://doi.org/10.1108/00483480310467624 Publisher: MCB UP Ltd.

[52] Markus Montola. 2010. The Positive Negative Experience in Extreme Role-Playing. *The Foundation Stone of Nordic Larp* (2010), 153–153. http://www.digra.org/wp-content/uploads/digital-library/10343.56524.pdf

[53] Nordmann, E., McAleer, P., Toivo, W., Paterson, H. & DeBruine, and L. [n. d.]. Data visualisation using R, for researchers who don't use R. ([n. d.]). https://psyteachr.github.io/introdataviz

[54] Mary Beth Oliver and Anne Bartsch. 2010. Appreciation as audience response: Exploring entertainment gratifications beyond hedonism. *Human Communication Research* 36, 1 (2010), 53–81. https://doi.org/10.1111/j.1468-2958.2009.01368.x

[55] Mary Beth Oliver, Nicholas David Bowman, Julia K. Woolley, Ryan Rogers, Brett I. Sherrick, and Mun-Young Chung. [n. d.]. Video games as meaningful entertainment experiences. 5, 4 ([n. d.]), 390–405. https://doi.org/10.1037/ppm0000066

[56] Thomas Lin Pedersen. [n. d.]. *patchwork: The Composer of Plots*. https://CRAN.R-project.org/package=patchwork

[57] Xiaolan Peng, Jin Huang, Alena Denisova, Hui Chen, Feng Tian, and Hongan Wang. 2020. A Palette of Deepened Emotions: Exploring Emotional Challenge in Virtual Reality Games. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3313831.3376221

[58] Kelsea Perry. 2021. Damsels and darlings: decoding gender equality in video game communities. *Feminist Media Studies* 0, 0 (Feb. 2021), 1–18. https://doi.org/10.1080/14680777.2021.1883085 Publisher: Routledge _eprint: https://doi.org/10.1080/14680777.2021.1883085.

[59] Serge Petralito, Florian Brühlmann, Glena Iten, Elisa D. Mekler, and Klaus Opwis. 2017. A Good Reason to Die: How Avatar Death and High Challenges Enable Positive Experiences. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 5087–5097. https://doi.org/10.1145/3025453.3026047

[60] Cody Phillips, Madison Klarkowski, Julian Frommel, Carl Gutwin, and Regan L Mandryk. 2021. Identifying commercial games with therapeutic potential through a content analysis of steam reviews. *Proceedings of the ACM on Human-Computer Interaction* 5, CHI PLAY (2021), 1–21.

[61] Bina Pradhan. 2003. Measuring Empowerment: A methodological approach. *Development* 46, 2 (June 2003), 51–57. https://doi.org/10.1057/palgrave.development.1110445

[62] R Core Team. [n. d.]. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

[63] Julian Rappaport. 1984. Studies in Empowerment. *Prevention in Human Services* (1984). https://doi.org/10.1300/J293v03n02_02 Publisher: Taylor & Francis Group.

[64] Giovanni Ribeiro, Katja Rogers, Maximilian Altmeyer, Thomas Terkildsen, and Lennart E. Nacke. 2020. Game Atmosphere: Effects of Audiovisual Thematic Cohesion on Player Experience and Psychophysiology. In *Proceedings of the Annual Symposium on Computer-Human*

Interaction in Play (CHI PLAY '20). Association for Computing Machinery, New York, NY, USA, 107–119. https://doi.org/10.1145/3410404.3414245

[65] Dimitris Rizopoulos. [n. d.]. ltm: An R package for Latent Variable Modelling and Item Response Theory Analyses. 17, 5 ([n. d.]), 1–25. http://www.jstatsoft.org/v17/i05/

[66] E. Sally Rogers, Ruth O. Ralph, and Mark S. Salzer. 2010. Validating the Empowerment Scale With a Multisite Sample of Consumers of Mental Health Services. *Psychiatric Services* 61, 9 (Sept. 2010), 933–936. https://doi.org/10.1176/ps.2010.61.9.933 Publisher: American Psychiatric Publishing.

[67] RStudio Team. [n. d.]. RStudio: Integrated Development Environment for R. ([n. d.]). http://www.rstudio.com/ Place: Boston, MA.

[68] Richard M. Ryan and Edward L. Deci. 2017. *Self-determination theory: basic psychological needs in motivation, development, and wellness*. Guilford Press, New York.

[69] Richard M. Ryan, C. Scott Rigby, and Andrew Przybylski. 2006. The Motivational Pull of Video Games: A Self-Determination Theory Approach. *Motivation and Emotion* 30, 4 (Dec. 2006), 344–360. https://doi.org/10.1007/s11031-006-9051-8

[70] Anne M. Scheel, Leonid Tiokhin, Peder M. Isager, and Daniël Lakens. 2021. Why Hypothesis Testers Should Spend Less Time Testing Hypotheses. *Perspectives on Psychological Science* 16, 4 (July 2021), 744–755. https://doi.org/10.1177/1745691620966795 Publisher: SAGE Publications Inc.

[71] Klaus R. Scherer. 2005. What are emotions? And how can they be measured? *Social Science Information* 44, 4 (Dec. 2005), 695–729. https://doi.org/10.1177/0539018405058216 Publisher: SAGE Publications Ltd.

[72] Hanna Schneider, Malin Eiband, Daniel Ullrich, and Andreas Butz. 2018. Empowerment in HCI - A Survey and Framework. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3173818

[73] Kennon M. Sheldon. 2014. Becoming Oneself: The Central Role of Self-Concordant Goal Selection. *Personality and Social Psychology Review* 18, 4 (Nov. 2014), 349–365. https://doi.org/10.1177/1088868314538549 Publisher: SAGE Publications Inc.

[74] Kennon M. Sheldon and Jonathan C. Hilpert. 2012. The balanced measure of psychological needs (BMPN) scale: An alternative domain general measure of need satisfaction. *Motivation and Emotion* 36, 4 (Dec. 2012), 439–451. https://doi.org/10.1007/s11031-012-9279-4

[75] Paula M. Short and James S. Rinehart. 1992. School Participant Empowerment Scale: Assessment of Level of Empowerment within the School Environment. *Educational and Psychological Measurement* 52, 4 (Dec. 1992), 951–960. https://doi.org/10.1177/0013164492052004018 Publisher: SAGE Publications Inc.

[76] Tanay Singhal and Oliver Schneider. 2021. Juicy Haptic Design: Vibrotactile Embellishments Can Improve Player Experience in Games. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3411764.3445463

[77] Barbara Bryant Solomon. 1976. *Black empowerment : social work in oppressed communities*. New York : Columbia University Press. http://archive.org/details/blackempowerment00barb_0

[78] Warren Spector. 2022. Player Empowerment as the Ultimate Success Criterion. https://www.gamedeveloper.com/blogs/player-empowerment-as-the-ultimate-success-criterion Section: blogs.

[79] Paul W. Speer and N. Andrew Peterson. 2000. Psychometric properties of an empowerment scale: Testing cognitive, emotional, and behavioral domains. *Social Work Research* 24, 2 (June 2000), 109–118. https://doi.org/10.1093/swr/24.2.109

[80] Alina Striner, Andrew M. Webb, Jessica Hammer, and Amy Cook. 2021. Mapping Design Spaces for Audience Participation In Game Live Streaming. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 329, 15 pages. https://doi.org/10.1145/3411764.3445511

[81] Jaroslav Svelch. 2021. Gaming the Iron Curtain: Making, Playing, and Copying Computer Games in Communist Czechoslovakia. In *Extended Abstracts of the 2021 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '21)*. Association for Computing Machinery, New York, NY, USA, 4. https://doi.org/10.1145/3450337.3486937

[82] Isabeau K. Tindall and Guy J. Curtis. 2019. Validation of the Measurement of Need Frustration. *Frontiers in Psychology* 10 (2019). https://www.frontiersin.org/article/10.3389/fpsyg.2019.01742

[83] Noam Tractinsky. 2018. The Usability Construct: A Dead End? *Human–Computer Interaction* 33, 2 (March 2018), 131–177. https://doi.org/10.1080/07370024.2017.1298038

[84] April Tyack and Elisa D. Mekler. 2020. Self-Determination Theory in HCI Games Research: Current Uses and Open Questions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–22. https://doi.org/10.1145/3313831.3376723

[85] Simon Urbanek and Jeffrey Horner. [n. d.]. *Cairo: R Graphics Device using Cairo Graphics Library for Creating High-Quality Bitmap (PNG, JPEG, TIFF), Vector (PDF, SVG, PostScript) and Display (X11 and Win32) Output*. https://CRAN.R-project.org/package=Cairo

[86] L.M. van der Lubbe, C. Gerritsen, M.C.A. Klein, and K.V. Hindriks. 2021. Empowering vulnerable target groups with serious games and gamification. *Entertainment Computing* 38 (May 2021), 100402. https://doi.org/10.1016/j.entcom.2020.100402

[87] Nanja van Dop, Jan Depauw, and Kristel Driessens. 2016. Measuring Empowerment: Development and Validation of the Service User Psychological Empowerment Scale. *Journal of Social Service Research* 42, 5 (Oct. 2016), 651–664. https://doi.org/10.1080/01488376.2016.1216915 Publisher: Routledge _eprint: https://doi.org/10.1080/01488376.2016.1216915.

[88] Daniela Villani, Claudia Carissoli, Stefano Triberti, Antonella Marchetti, Gabriella Gilli, and Giuseppe Riva. 2018. Videogames for Emotion Regulation: A Systematic Review. *Games for Health Journal* 7, 2 (April 2018), 85–99. https://doi.org/10.1089/g4h.2017.0108 Publisher: Mary Ann Liebert, Inc., publishers.

[89] Jan B. Vornhagen. [n. d.]. *GALCR: Geneva Affect Label Coder (GALC) for R*.

[90] Jan Benjamin Vornhagen and Elisa D. Mekler. 2023. Dis-/Empowerment as PX. (Feb. 2023). https://osf.io/zhtu8/ Publisher: OSF.

[91] Linda Weidenstedt. 2016. Empowerment Gone Bad: Communicative Consequences of Power Transfers. *Socius: Sociological Research for a Dynamic World* 2 (Jan. 2016), 237802311667286. https://doi.org/10.1177/2378023116672869

[92] Matthew Alexander Whitby, Ioanna Iacovides, Sebastian Deterding, and Sebastian Deterding. 2019. Gameplay moments that challenge the player's perspective. https://osf.io/2pg5j/

[93] Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester,

Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. [n. d.]. Welcome to the tidyverse. 4, 43 ([n. d.]), 1686. https://doi.org/10.21105/joss.01686

[94] Hadley Wickham and Jennifer Bryan. [n. d.]. *readxl: Read Excel Files*. https://CRAN.R-project.org/package=readxl

[95] Marc A. Zimmerman. 1995. Psychological empowerment: Issues and illustrations. *American Journal of Community Psychology* 23, 5 (1995), 581–599. https://doi.org/10.1007/BF02506983 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1007/BF02506983.

[96] Marc A. Zimmerman and Julian Rappaport. 1988. Citizen participation, perceived control, and psychological empowerment. *American Journal of Community Psychology* 16, 5 (1988), 725–750. https://doi.org/10.1007/BF00930023 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1007/BF00930023.

A Appendix: Figures

A.1 PX measures per Theme

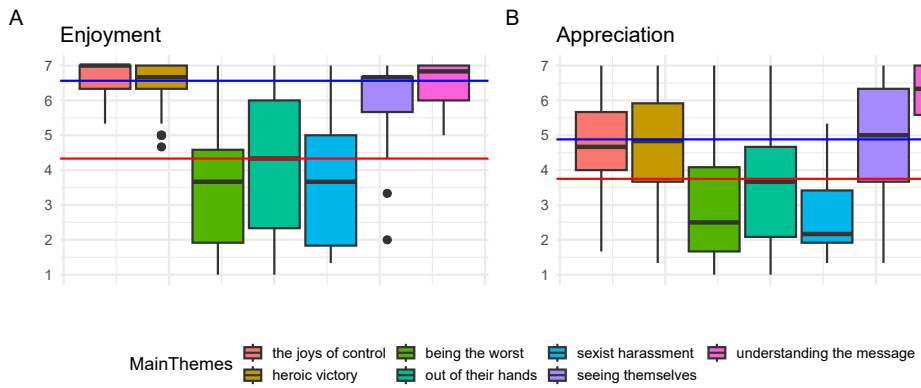

Fig. III.1: Boxplot for Enjoyment (A) and Appreciation (B) separated by theme (color). The red and blue lines represent the mean of the empowered and disempowered group respectively. Particularly notable is the high appreciation of "Understanding the Message", which mostly consists of disempowered experiences.

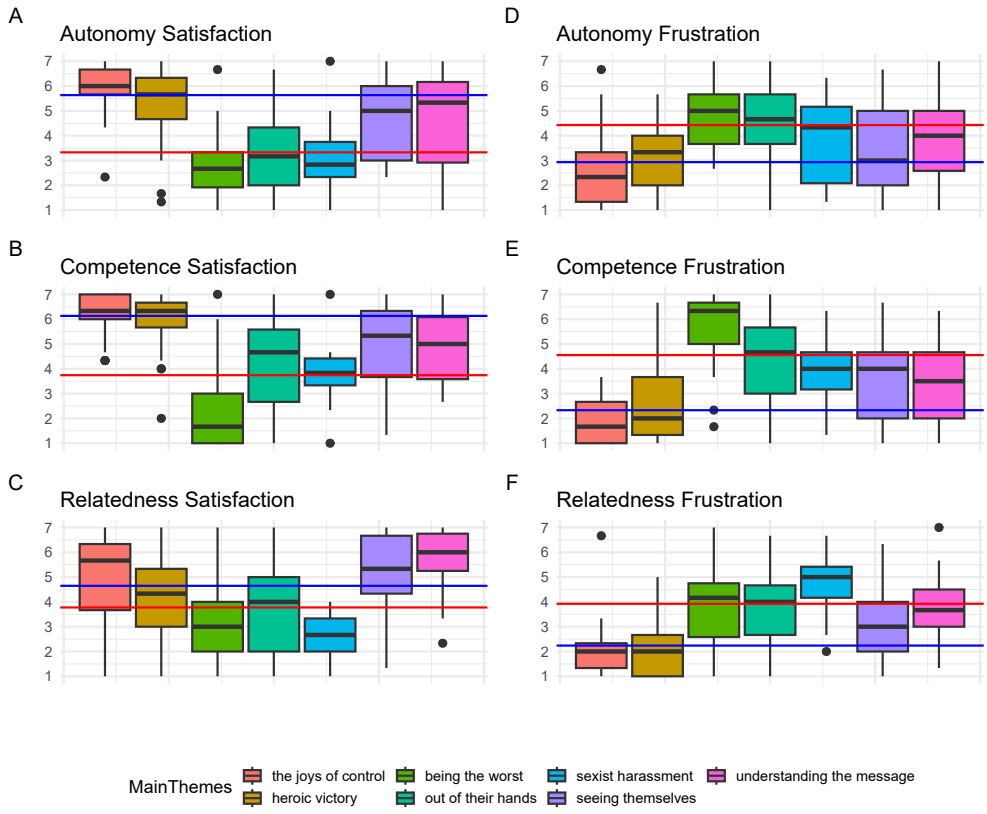

Fig. III.2: Boxplots for need satisfaction (A,B,C) and frustration (D,E,F), separated by themes. The red and blue lines represent the mean of the empowered and disempowered group respectively. Satisfaction follows a somewhat clear trend with empowering experiences scoring high, and disempowering experiences scoring low. However, there is relatively little need frustration. Only competence seemed to be particularly frustrated in the "Being the Worst" theme and relatedness was somewhat frustrated in the "Sexist Harassment" group.

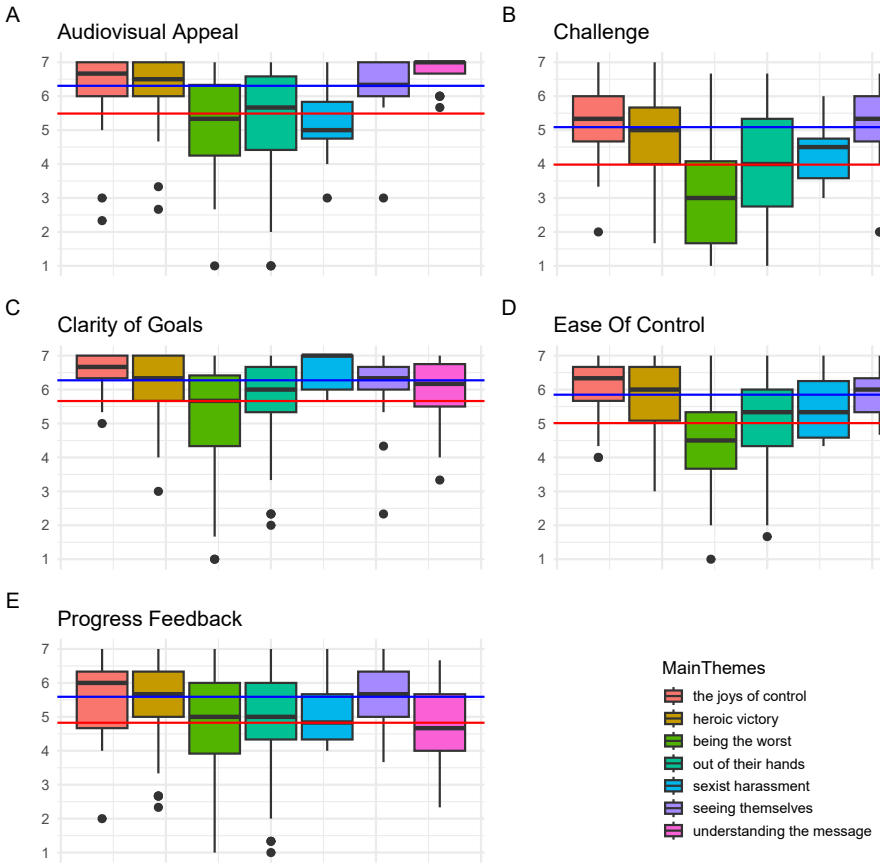

Fig. III.3: Boxplots for the functional consequences of the PXI, separated by themes. The red and blue lines represent the mean of the empowered and disempowered group respectively. It is notable, that all themes score very high on these dimensions. The primary outlier is the challenge scale on which particularly the "Out Of Their Hands" theme scored lower than the others.

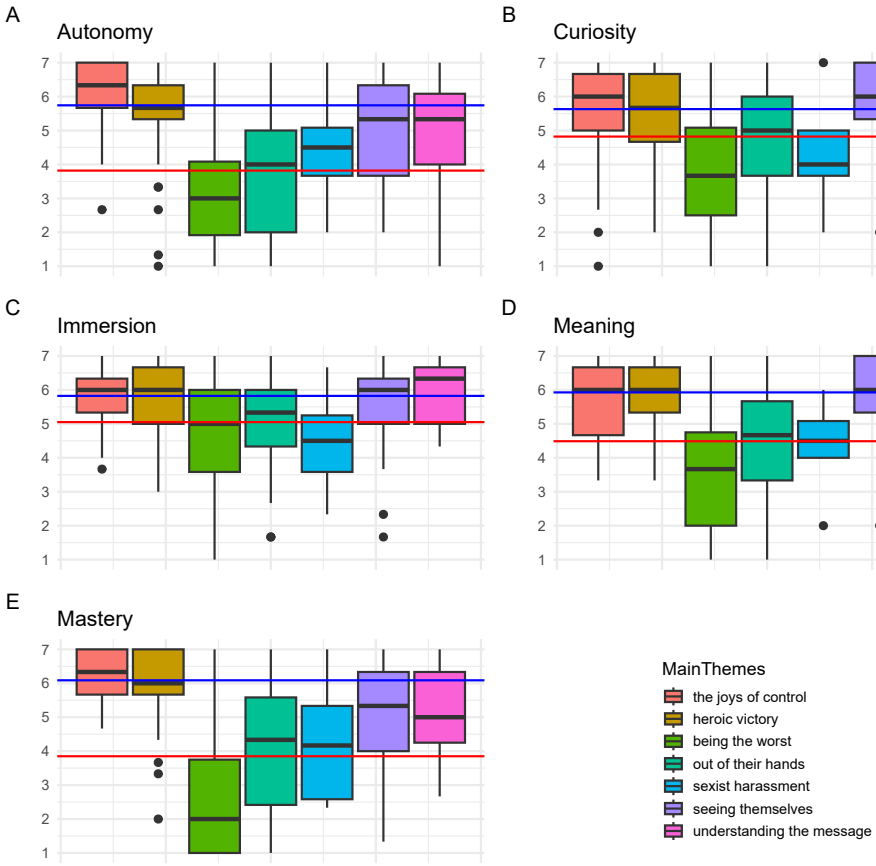

Fig. III.4: Boxplots for the psychosocial consequences of the PXI, separated by themes. The red and blue lines represent the mean of the empowered and disempowered group respectively. Compared to the functional consequences the differences between themes are more pronounced, with particularly curiosity and meaning scoring rather high in the mixed themes ("Seeing Themselves in the Game" and "Understanding the Message").

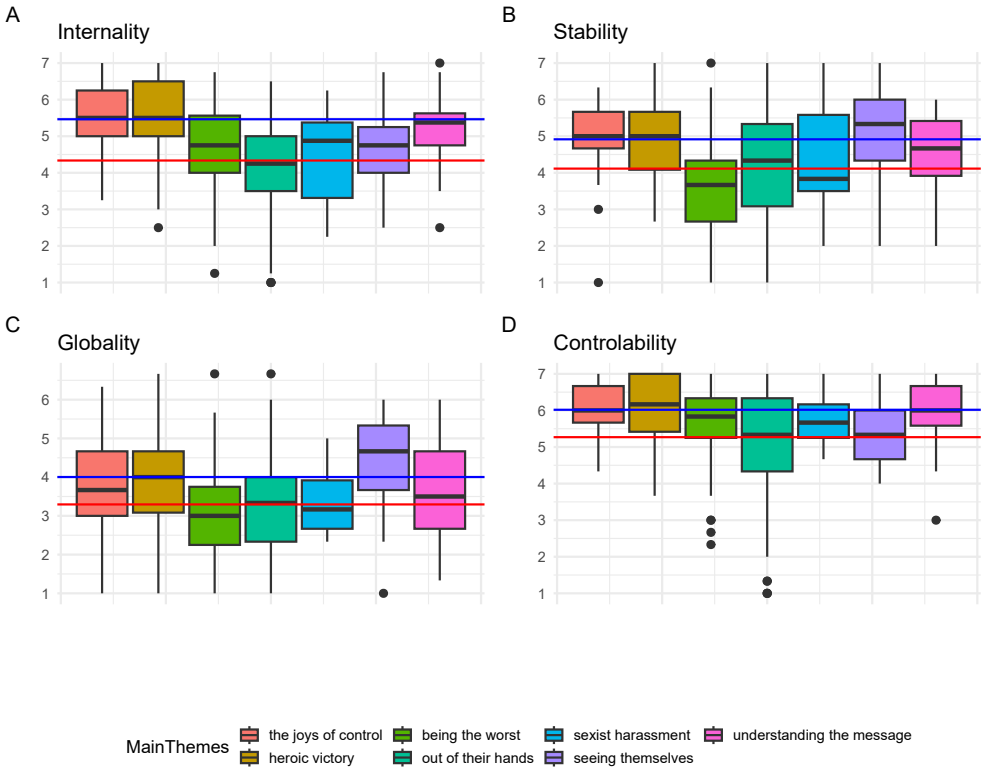

Fig. III.5: Boxplots for the GSAQ, separated by themes. The red and blue lines represent the mean of the empowered and disempowered group respectively. It is noteworthy, that controllability was rated quite high, even for the "Out of Their Hands" theme. Moreover, Globality was the lowest scored dimension overall, with only "Exemplar" rating considerably higher than the other themes.