Games User Research (GUR)

How to understand your players

Prof. Perttu Hämäläinen

2023 (1st iteration of the course, work in progress)

Course contents

Today:

- Basic qualitative & quantitative user research methods
- AI-assisted data analysis

Tue-Wed:

- GUR experiment: Test/research your game with one or more of the methods

Thursday:

- AI-assisted data generation lecture & exercises (playtesting with AI agents)

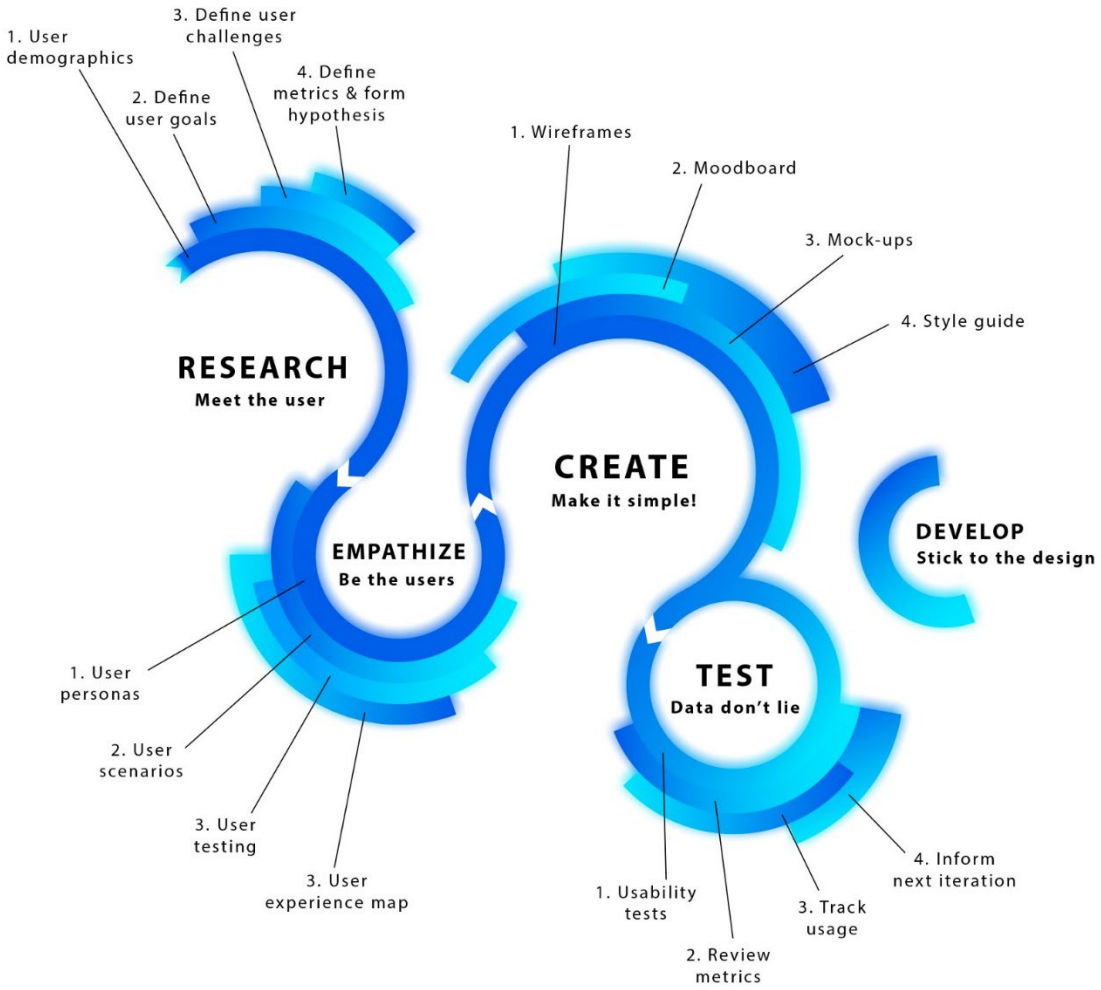
Friday afternoon:

- Review and discussion of GUR experiment results & general post-mortem

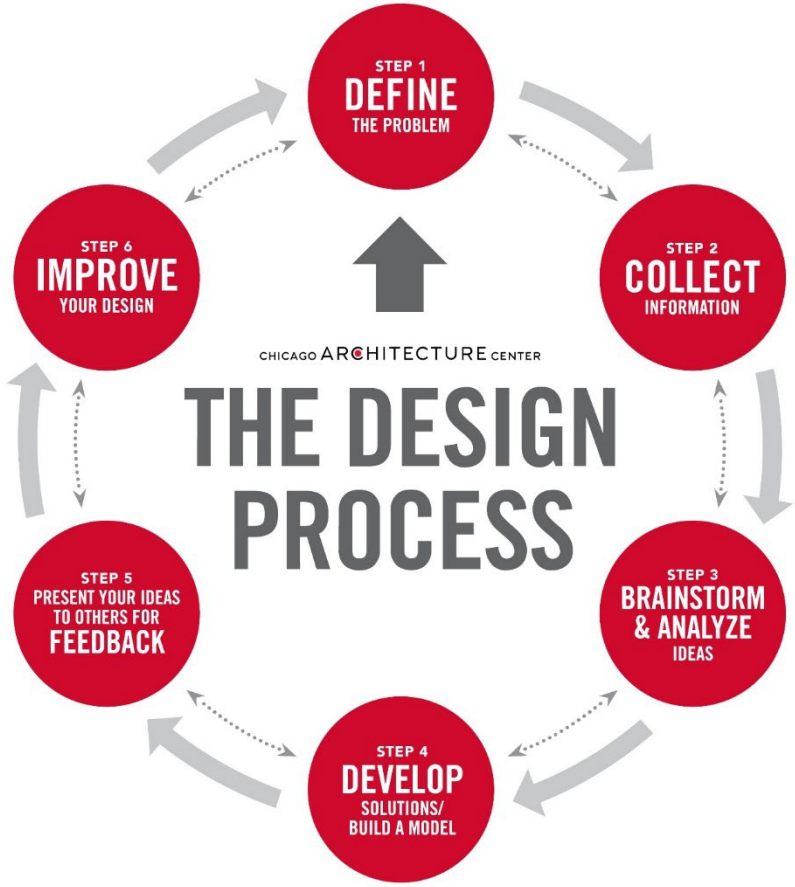
This is primarily a GUR course for *game developers.* If you'd like to become a full-time professional GUR person, need to go deeper to topics such as *human-computer interaction, perception & action, cognitive psychology.*

GUR and game development

Game Design and Development Processes

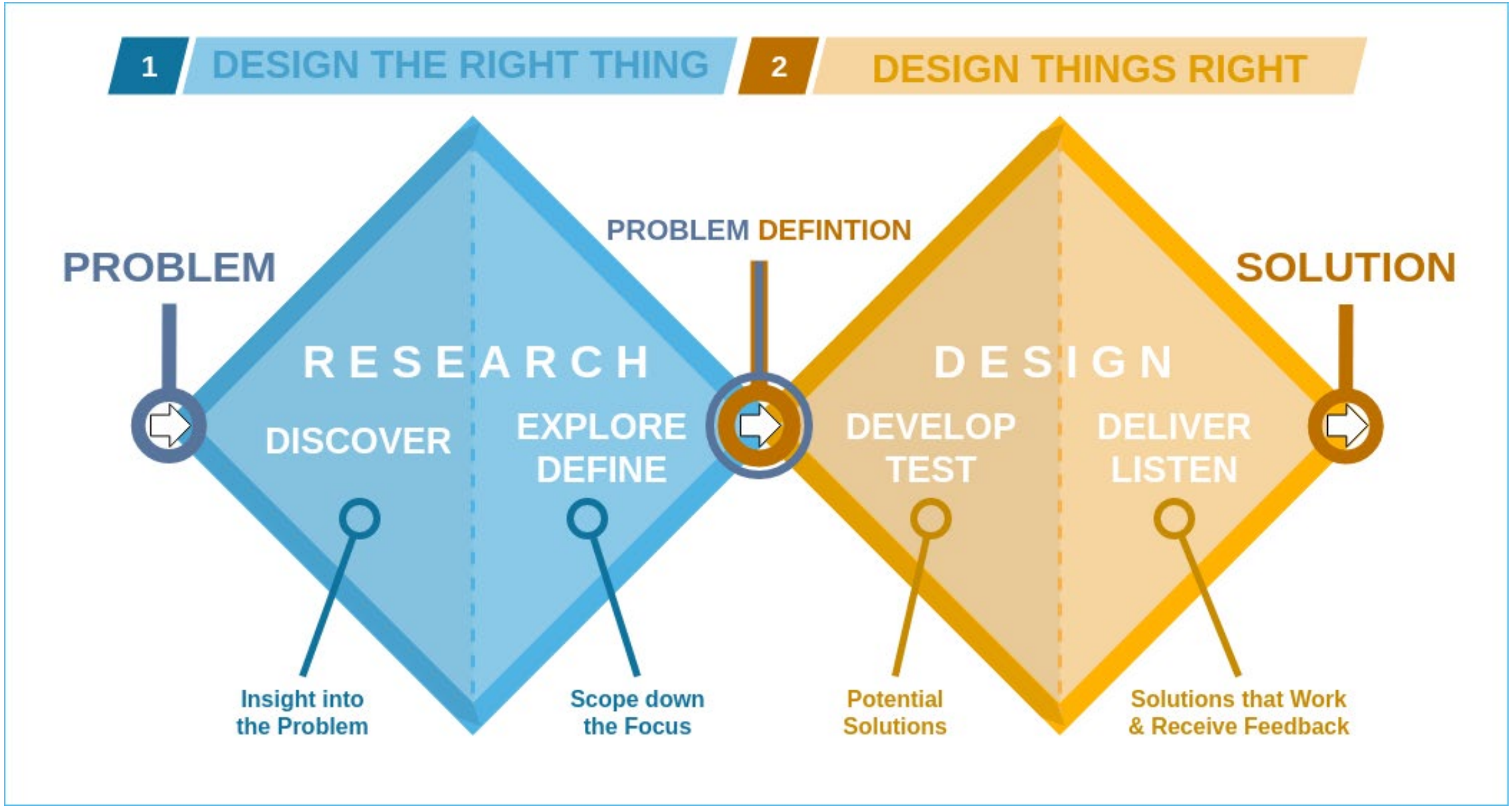


https://medium.com/@alexwyrick/ux-design-process-a66e6796857e



https://www.discoverdesign.org/handbook

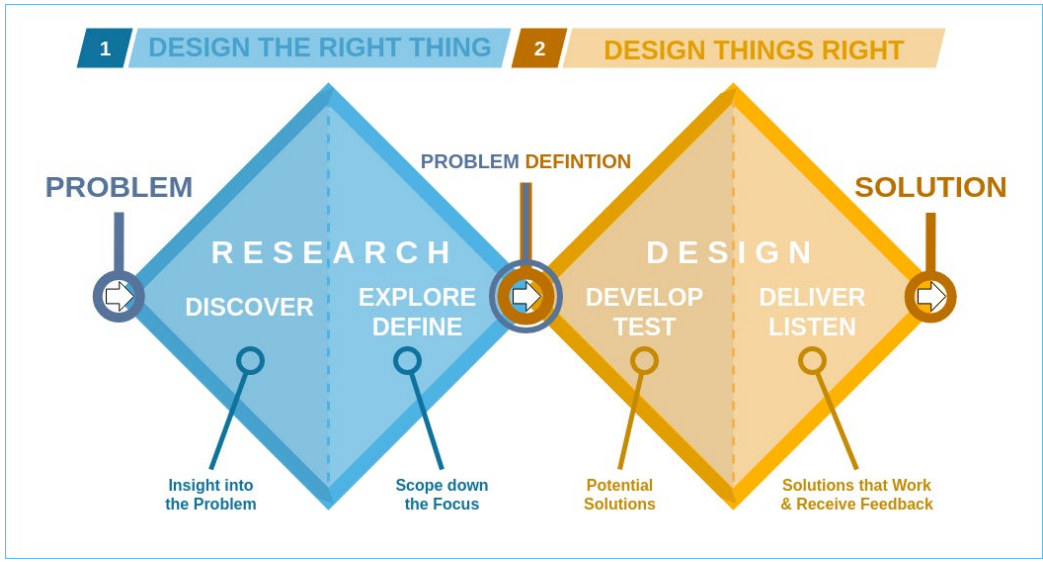
The Double Diamond Design Process Model



Let’s think step by step!

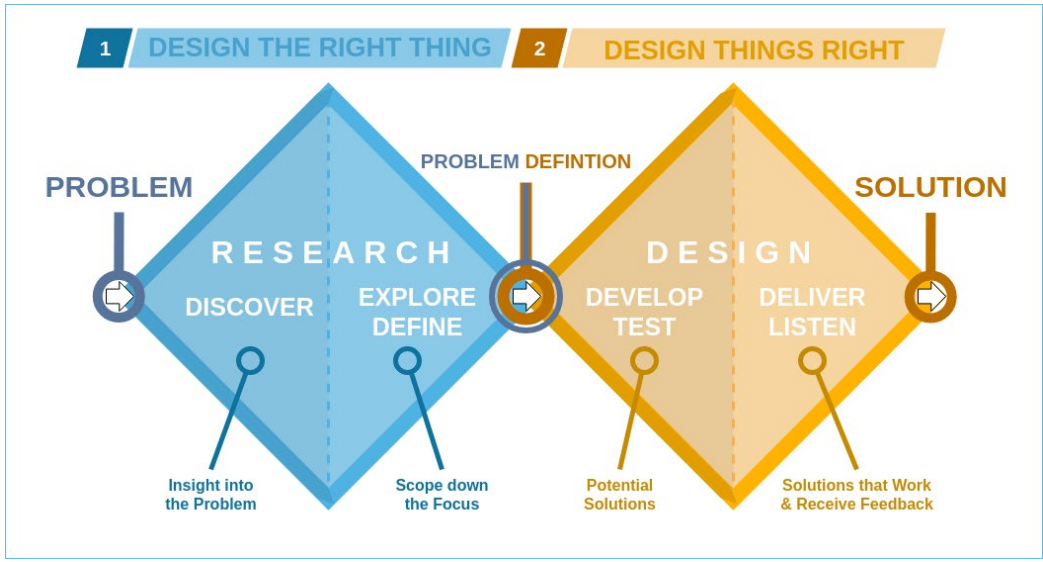
GUR and Double Diamond

- Discover: Generate ideas & sketches.
- Useful GUR data: insights on player preferences, motivations etc. Ideas and improvements suggested by players.
- Data sources: Discord, Reddit etc., playtests and interviews conducted on your previous games. Co-creation workshops / game jams with players



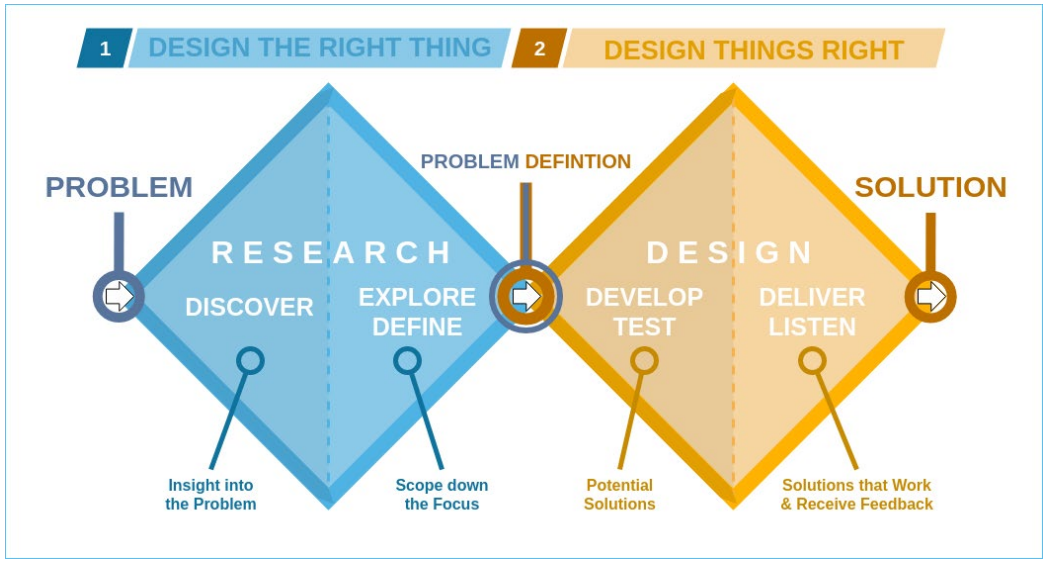
GUR and Double Diamond

- Explore & define: Critique and prune ideas.
- GUR data: user feedback
- Data sources: small-scale user tests on prototypes or sketches, surveys
- Also: heuristic evaluation



GUR and Double Diamond

- Develop & test: Create games & prototypes, often branching into multiple options
- Deliver & listen: Prune others than the best option
- Often intertwined – every sprint may include both
- GUR data: player behavior & feedback (voice, gameplay & facial video, interviews, questionnaires, game analytics/event logs)
- Data sources: game in alpha/beta/soft launch (A/B testing), online playtesting, traditional playtesting



Academic and industry GUR

- Industry: identifying issues, comparing multiple design alternatives, understanding who your players are and what they enjoy
- Research: advancing the understanding of player experience, evaluating novel game mechanics or other innovations
- Getting your degree: if your thesis is about developing a game, a novel game mechanic, a game development tool etc., you typically need some GUR methods to evaluate and report what you created and how players/users experience it

GUR methods

Vocabulary

- Sample size (N): How many user study participants or players
- Qualitative & quantitative research: two major categories of research methods

Qualitative GUR

- Small sample, typically 3 to 15
- Exploring and discovering research questions and/or hypotheses: “The user feedback suggests that A may be more motivating than B”
- Rich & and descriptive data “a few participants emphasized healing qualities of the game narrative, such as *‘kept quotes from the game in my everyday life [...] (which) I remind myself when my mental health begins to fade.’ ”* (Bopp et al. 2021)

”My Soul Got a Little Bit Cleaner”: Art Experience in Videogames

JULIA A. BOPP*, JAN B. VORNHAGEN*, and ELISA D. MEKLER, Aalto University

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/.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; • **Applied computing** → *Psychology*; *Arts and humanities.*

Additional Key Words and Phrases: player experience, emotion, art experience, videogames, empirical aesthetics

ACM Reference Format:
Julia A. Bopp, Jan B. Vornhagen, and Elisa D. Mekler. 2021. ”My Soul Got a Little Bit Cleaner”: Art Experience in Videogames. *Proc. ACM Hum.-Comput. Interact.* 5, CHI PLAY, Article 237 (September 2021), 19 pages. https://doi.org/10.1145/3474664

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 [58, 59, 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

*Both authors contributed equally to this research and are alphabetically listed as co-first authors

Authors’ address: Julia A. Bopp; Jan B. Vornhagen, jan.vornhagen@aalto.fi; Elisa D. Mekler, Aalto University, P.O. Box 1212, Aalto, 0076.

2573-0142/2021/9-ART237. https://doi.org/10.1145/3474664

Quantitative GUR

- Large sample, typically 30+
- Numerical descriptions of data
  - 30% of survey respondents selected X, with Y% margin of error
- Confirming hypotheses:
  - On average, players used more money / had a higher pass rate / reported higher mastery motivation with game version A compared to B, p<0.01, with 80% of power
- More reliable / generalizable results from tested players to other players from the same population (age, gender, gaming history)

3PP-R: Enabling Natural Movement in 3rd Person Virtual Reality

Inan Evin
inan.evin@aalto.fi
Aalto University

Toni Pesola
toni.pesola@aalto.fi
Aalto University

Maximus D. Kaos
maximus.kaos@aalto.fi
Aalto University

Tuukka M. Takala
tuukka.takala@aalto.fi
Aalto University & Waseda University

Perttu Hämäläinen
perttu.hamalainen@aalto.fi
Aalto University

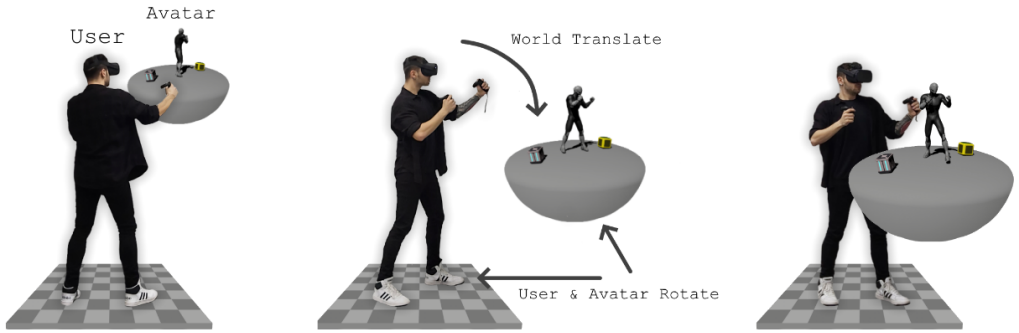


Figure 1: 3PP-R allows the user to interact in 3rd-Person Perspective (3PP) using natural movements, including body rotation without losing sight of the avatar. A virtual display such as a 3D miniature world model hovers in air in front of the user, showing a 3rd-person avatar. When the user turns, the display orbits around the user but does not rotate except for the avatar. From the user's perspective, the display appears fixed in the field of vision, and the world rotates around the avatar.

ABSTRACT

We propose 3PP-R, a novel Virtual Reality display and interaction technique that allows natural movement in 3rd-person perspective (3PP), including body rotation without losing sight of the avatar. A virtual display such as a World-in-Miniature model orbits around the user when the user turns, but does not rotate except for the user's avatar. From the user's perspective, the display appears fixed in the field of vision, while the world rotates around the avatar. 3PP-R combines the strengths of 3PP and 1st-person perspective (1PP): Similar to 1PP, it allows interacting with rich natural movements, while also reaping the benefits of 3PP, i.e., superior spatial awareness and animating the avatar without nauseating viewpoint movement, e.g., for joystick-controlled locomotion. We test 3PP-R in a maze navigation study, which indicates considerably less cybersickness in 3PP-R than in 1PP. We also demonstrate 3PP-R in dynamic game interaction including running, jumping, swinging on bars, and martial arts.

CCS CONCEPTS

• **Computing methodologies** → **Virtual reality**; • **Human-centered computing** → **Interaction techniques**.

KEYWORDS

Virtual reality display; 3rd person perspective; Virtual camera design

ACM Reference Format:
Inan Evin, Toni Pesola, Maximus D. Kaos, Tuukka M. Takala, and Perttu Hämäläinen. 2020. 3PP-R: Enabling Natural Movement in 3rd Person Virtual Reality. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '20), November 2–4, 2020, Virtual Event, Canada.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3410404.3414239

CHI PLAY '20, November 2–4, 2020, Virtual Event, Canada

ACM ISBN 978-1-4503-8074-4/20/11…$15.00
https://doi.org/10.1145/3410404.3414239

1 INTRODUCTION

A fundamental problem of Virtual Reality (VR) is how to seamlessly combine three features: 1) interaction using natural body movements, 2) the ability to navigate large virtual spaces, and 3) keeping cybersickness to a minimum. Using a 1st person perspective (1PP) and one-to-one viewpoint tracking minimizes cybersickness and allows natural interaction, but at the same time, it prevents navigating virtual spaces larger than the real interaction space. To solve the

Mixed methods: Combining both “what” and “why”

Figure 11. Adjectives that participants used to describe how the exaggerated movement felt

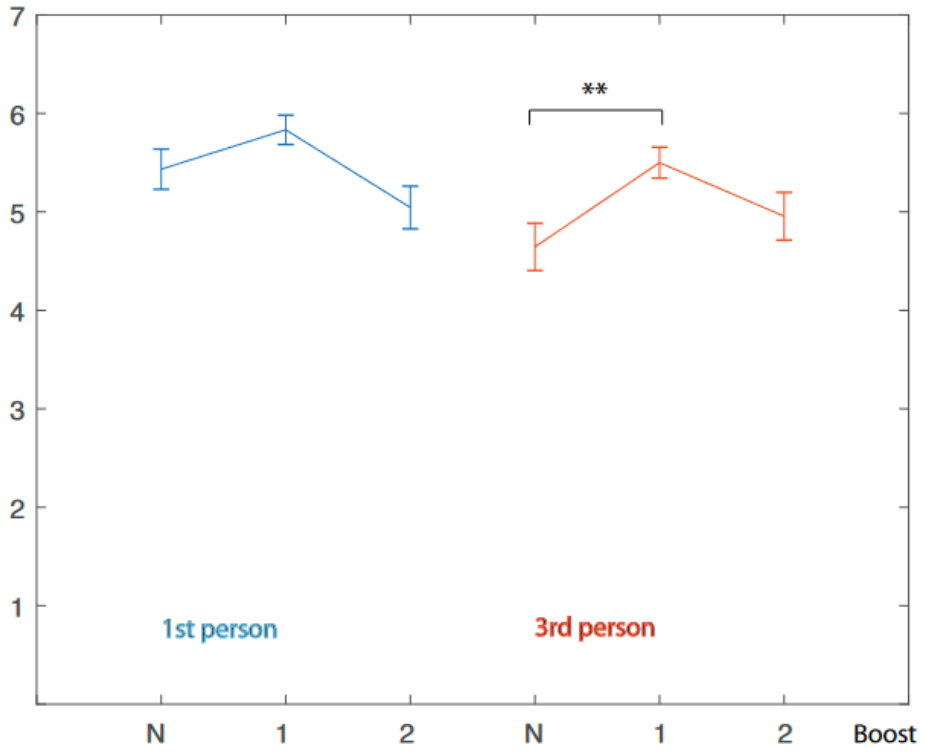


Figure 6. Naturalness of motion.

"I didn’t notice there was boosted movement, but I remember being very surprised that I was able to hit the high targets, initially it was kind of bizarre, like ’Wow am I really that good?’"

Quantitative research primer

Three Statistical Tests Every Game Developer Should Know

GDC
469K subscribers

Subscribe

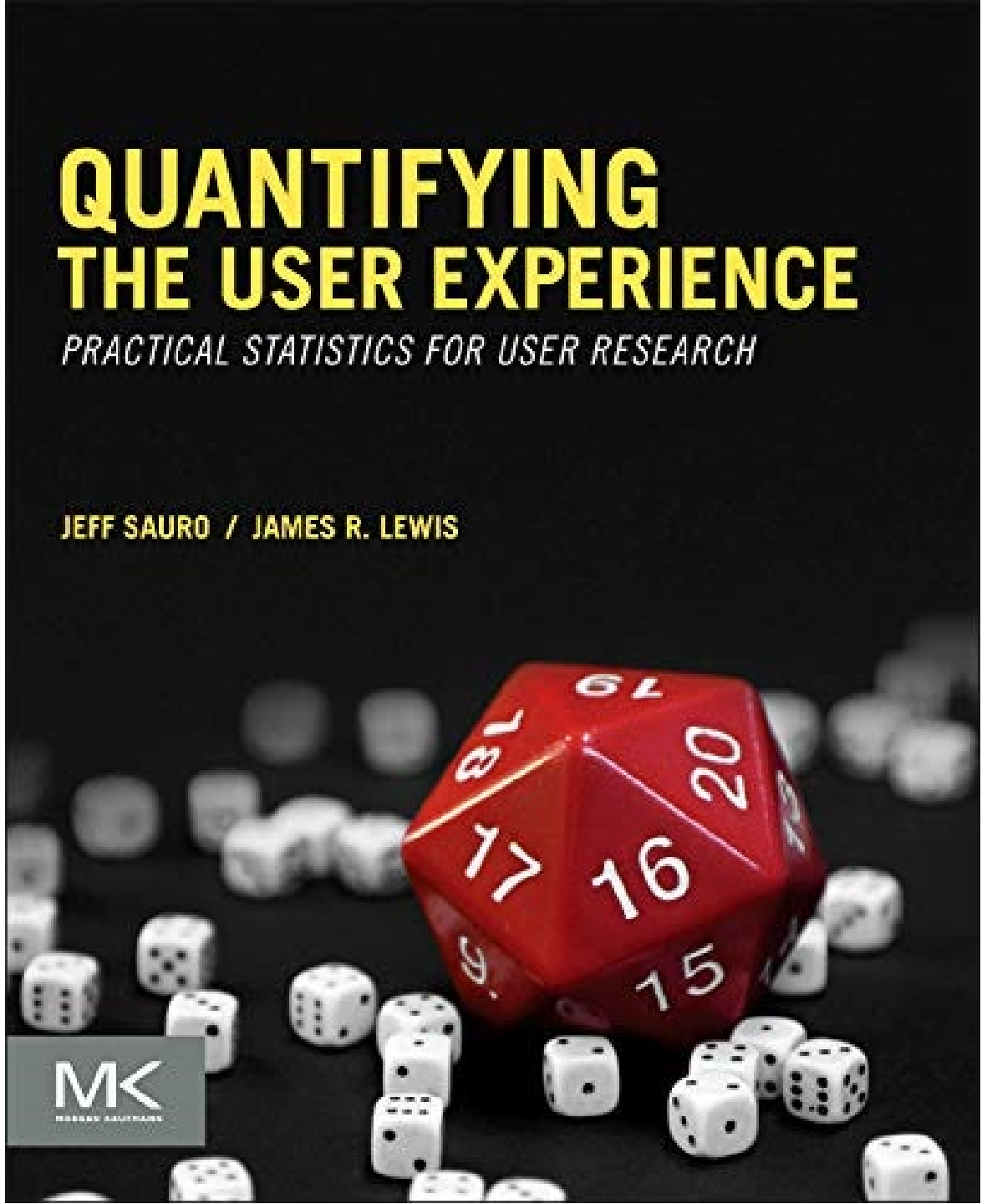
9K | Share | Clip | Save | ...

The video uses Excel, but that's becoming outdated

- Python provides easy to use and more extensive helper functions
- Colab notebooks allow you to easily load data from your Google drive and run everything in a browser, without installing any software
- https://www.w3schools.com/python/scipy/scipy_statistical_significance_tests.php
- https://colab.research.google.com/github/gureckis/nyu_lab_in_cognition/blob/master/book/tips/ultimate-guide-ttest-python.ipynb
- https://colab.research.google.com/github/gureckis/nyu_lab_in_cognition/blob/master/book/chapters/10/00-ttest.ipynb

A good introductory book

- Most common statistical testing methods and study designs
- Useful decision trees that you can follow to identify what you need



The video uses Excel, but that's becoming outdated

- Python provides easy to use and more extensive helper functions
- Colab notebooks allow you to easily load data from your Google drive and run everything in a browser, without installing any software
- https://www.w3schools.com/python/scipy/scipy_statistical_significance_tests.php
- https://colab.research.google.com/github/gureckis/nyu_lab_in_cognition/blob/master/book/tips/ultimate-guide-ttest-python.ipynb
- https://colab.research.google.com/github/gureckis/nyu_lab_in_cognition/blob/master/book/chapters/10/00-ttest.ipynb

Good quantitative questionnaires?

- PXI: a validated player experience questionnaire
- Enjoyment & intrinsic motivation
- Short

But, before using any questionnaire, read the items and ask yourself: Do the items really make sense for my research questions?

ELSEVIER

International Journal of Human-Computer Studies

Volume 135, March 2020, 102370

Development and validation of the player experience inventory: A scale to measure player experiences at the level of functional and psychosocial consequences

Vero Vanden Abeele [a], Katta Spiel [a b], Lennart Nacke [c], Daniel Johnson [d], Kathrin Gerling [a]

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.ijhcs.2019.102370

Get rights and content

Highlights

- Scale to measure player experience at functional and psychosocial level concurrently.
- Development and validation carried out with 64 games experts and 529 players.
- Results support discriminant and convergent validity, plus configural invariance.
- Economical instrument, measuring 10 constructs with 3 items each.
- Suited to measure PX across a variety of game genres and gamified applications.

Measuring challenge

- A questionnaire for measuring 4 aspects of game challenge: cognitive, performative, emotional, and decision-making

ELSEVIER

International Journal of Human-Computer Studies

Volume 137, May 2020, 102383

Measuring perceived challenge in digital games: Development & validation of the challenge originating from recent gameplay interaction scale (CORGIS)

Alena Denisova [a], Paul Cairns [b], Christian Guckelsberger [c], David Zendle [b]

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.ijhcs.2019.102383

Get rights and content

Highlights

- Scale measuring perceived challenge in video games.
- Four sub-scales (30 items) measuring four types of perceived challenge in video games: cognitive, performative, emotional, and decision-making challenge.
- Development and validation are carried out over three studies including 1390 players with diverse backgrounds playing video games from a range of genres.
- The questionnaire is a systematic, extensive, reliable, and valid tool to measure perceived challenge in video games.

Within & between subjects study designs

- Comparative quantitative experiments are split into *experimental conditions,* e.g., game versions A, B, C
  - *Independent variables*: The variables manipulated to create the conditions, e.g., choice of game mechanic (a categorical independent variable) or jump height multiplier (a real-valued independent variable)
  - *Dependent variables*: measured "outcome" variables of interest, e.g., a game level's pass rate or players' self-reported enjoyment
- Between subjects: each players only experiences one experimental condition
- Within subjects: all players experience all experimental conditions
- A/B testing: an industry term for between subjects quantitative comparisons implemented by different players experiencing a different game version for some period of time

Within subjects is a good default for school projects and Master's theses

- Pro: the effect of *individual differences* (age, gaming background etc) on the dependent variable can be averaged out in the statistical analyses.
  - This allows reliable results with a smaller sample size
- Con: the player learns during all conditions and the independent variables might become obvious, which may bias the data.
- Learning effects can be mitigated by *counterbalancing*
  - For N experimental conditions, there are N! permutations of condition orders. Every permutation should be experienced equally many times.
  - For examples, 3 conditions gives 3*2*1=6 permutations. Sample size should be a multiple of this
  - If there are more permutations than you have users, partial counterbalancing using the Latin Squares method is recommended (Google it)

What if you only have 1 condition?

- Usually, studies include at least one baseline condition for comparing
  - E.g., is the proposed new design better than the old design?
- If no baseline, must otherwise determine whether the results are good or not:
  - Compare against results reported by others
  - Compare against goals defined through some other means, e.g., the management has decided that completing a level should take less than 5 minute, on average
  - Use a *standardized* questionnaire such as System Usability Scale (SUS), which allows converting the score to a percentile rank, e.g., “better than 90%”

https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html

How many users do you need?

- 6 or so when trying to discover problems with your game (qualitative)
- 10+ when trying to understand and define players (qualitative)
- 30+ for quantitative comparisons
- But, easily 100+ if
  - Noisy data, e.g., Likert-scale self-report measures of presence, immersion, player experience
  - The difference between compared game versions is small (in the measured variables)
  - Between-subjects experiment instead of within subjects

Determining the sample size

- Use G*Power: https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower
- For within subjects, select “difference between dependent means”
  - “paired samples” and “dependent means” are other common terms for within subjects data and analyses
- “Minimum viable” defaults: alpha=0.05 (the statistical significance threshold for the p-value) and power = 0.8 (80%)
- Effect size: look at pilot data, previous studies etc. to make an educated guess. Use 0.5 (“moderate”) if you have no idea.
  - Effect size can mean many things (https://en.wikipedia.org/wiki/Effect_size)
  - Here, we refer to the difference of means for *t*-test, where effect size refers to how many standard deviations of difference there is between the means
  - Also consider: what is an effect size that is not just statistically but also practically significant
- For more reliable results: alpha=0.01, and also power=0.95, if you are interested in minimizing false negative errors

Pitfalls

- Running quant. analyses is like rolling dice
- More rolls => higher chance of “success”
- For reliable results, one should decide on all details (data cleaning and preprocessing, tests used) before looking at the data.
- One also should not change any details after seeing the results
- Further: all statistical tests have assumptions that must hold (e.g., normally distributed data)

Psychological Science
Volume 22, Issue 11, November 2011, Pages 1359-1366

https://doi.org/10.1177/0956797611417632

General Article

False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant

Joseph P. Simmons[1], Leif D. Nelson[2], and Uri Simonsohn[1]

Abstract

In this article, we accomplish two things. First, we show that despite empirical psychologists’ nominal endorsement of a low rate of false-positive findings (≤ .05), flexibility in data collection, analysis, and reporting dramatically increases actual false-positive rates. In many cases, a researcher is more likely to falsely find evidence that an effect exists than to correctly find evidence that it does not. We present computer simulations and a pair of actual experiments that demonstrate how unacceptably easy it is to accumulate (and report) statistically significant evidence for a false hypothesis. Second, we suggest a simple, low-cost, and straightforwardly effective disclosure-based solution to this problem. The solution involves six concrete requirements for authors and four guidelines for reviewers, all of which impose a minimal burden on the publication process.

Keywords

methodology, motivated reasoning, publication, disclosure

No more “massaging the data until it tells a story”

- Replication crisis: many classic psychological results can’t be replicated by other researchers => probably false positives
- Professors and research groups engaging in questionable research practices (QRP:s) have lost their credibility, funding, and/or jobs
  - Forced smile (pen in mouth) doesn’t make you happier
  - Power posing does not boost your testosterone
  - Cautionary tale: Brian Wansink admitted to and promoted QRPs in his blog, tried to backpedal but the Internet does not forget https://web.archive.org/web/20170312041524/http:/www.brianwansink.com/phd-advice/the-grad-student-who-never-said-no

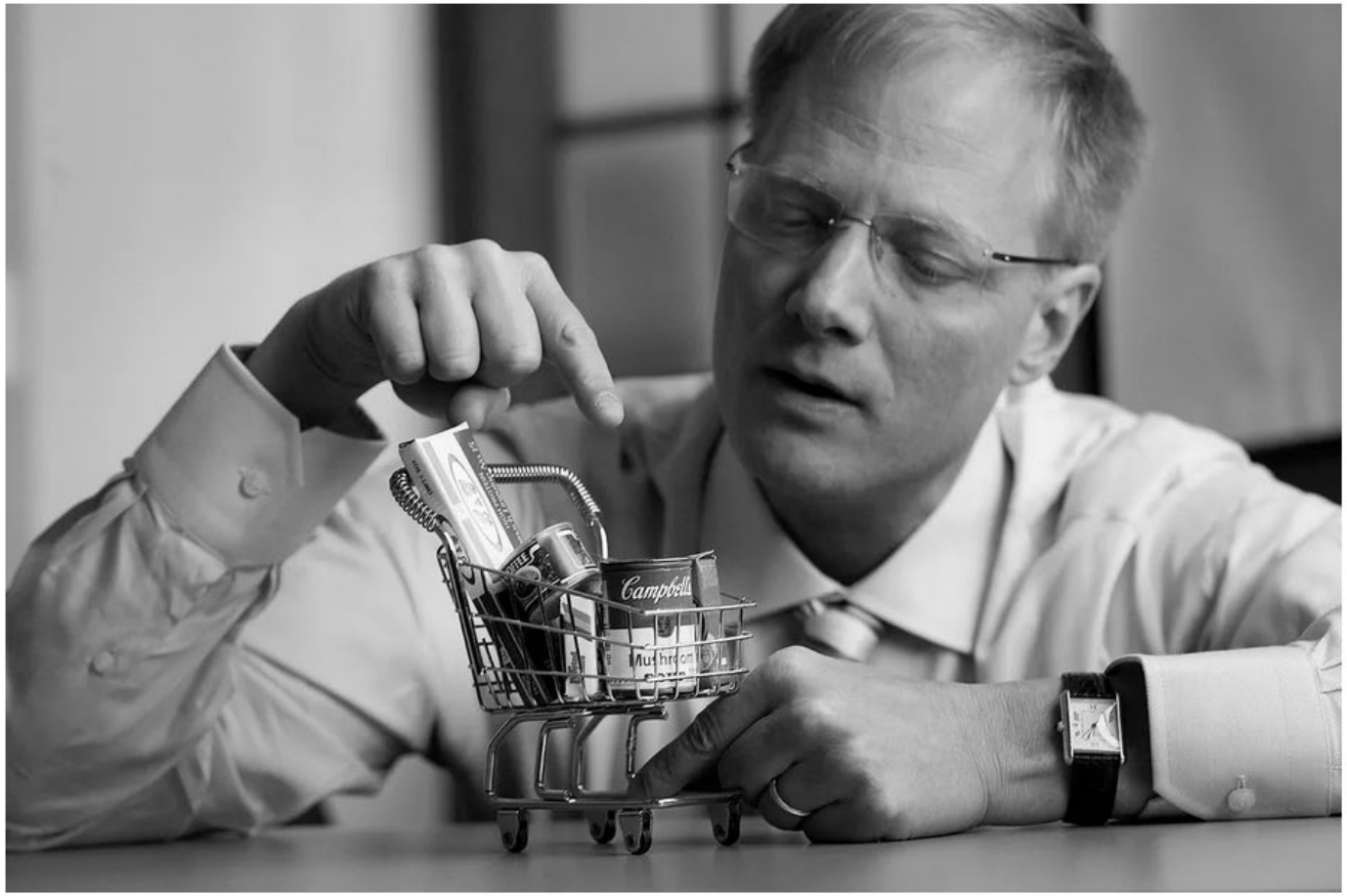
A top Cornell food researcher has had 15 studies retracted. That's a lot.

https://www.vox.com/science-and-health/2018/9/19/17879102/brian-wansink-cornell-food-brand-lab-retractions-jama

Brian Wansink is a cautionary tale in bad incentives in science.

By Brian Resnick and Julia Belluz | Updated Oct 24, 2018, 2:25pm EDT

SHARE



Brian Wansink just had six papers retracted from top journals. | Jason Koski

Most Read

1 Why is there so much secrecy in philanthropy?

2 Yes, the Ohio train wreck is an environmental disaster. No, it's not Chernobyl.

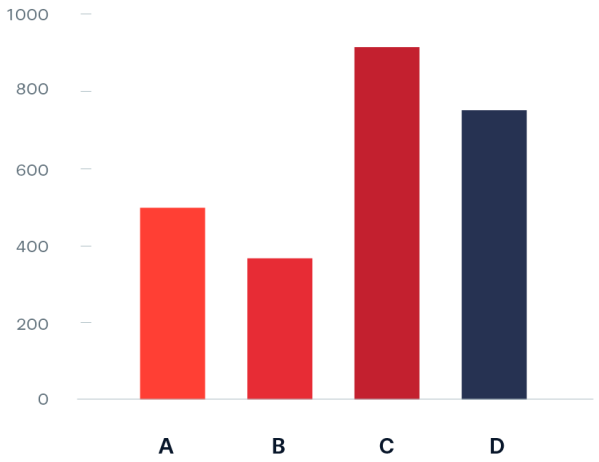
3 Can the Republican establishment finally stop Trump this time?

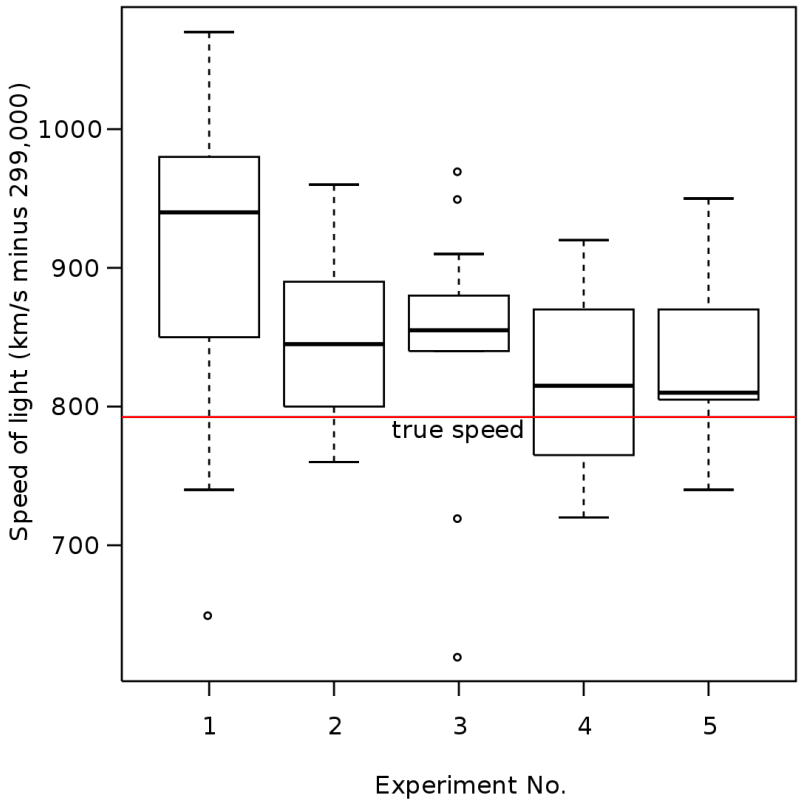
4 A juicy new legal filing reveals who really controls Fox News

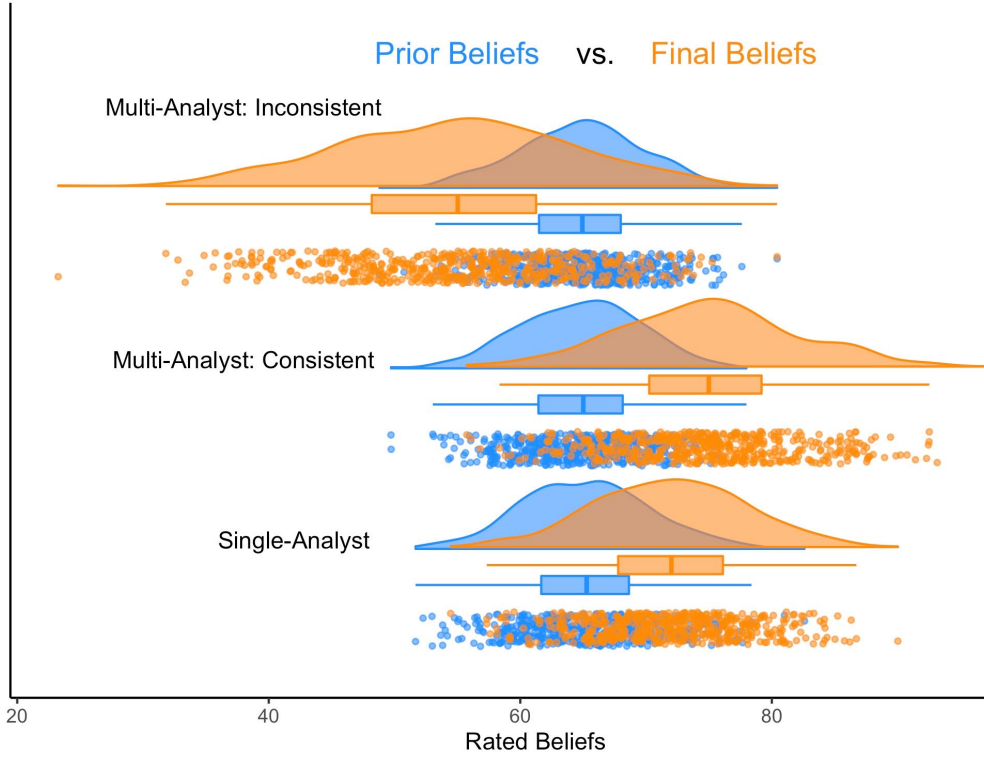
5 A top Senate Democrat has an extraordinarily radical plan to deal with Trump's worst judge

Reliability and visualization

- Bar charts < box plot < raincloud plot







Two-interval forced choice (2IFC)

- Each participant repetitively experiences the same two experimental conditions, in randomly ordered pairs without explicit labels about which condition is which, and has to select the one in each pair that satisfies some criterion:
  - The wine that tastes better
  - The video with less compression artefacts
  - The game with better game feel
- Benefit: This removes noise and can allow reliably determining even small differences, with a small number of participants
- Drawback: Time required per participant
  - Important to make each individual trial very short

In general: “Which was better?” is hard to answer after a delay or long experiments

- Participants might not remember or pay attention to the right things
- Comparison may become more reliable if:
  - Using spatial rather than temporal display of stimuli
  - Experiencing with minimum spatiotemporal distance

Spatially close => clear difference

Spatially distant => less clear

Which stimulus is brighter?

Stimulus 1

A good rule of thumb

- Do quantitative research only if
  - Your game can be deployed to a large number of players online (client code allows backend to customize things for each player, or you have both tested versions playable in a browser and recruit participants through an online service such as Prolific)

  or
  - You test IRL but can assume at least a moderate effect size and have enough time to recruit and manage 30+ participants

  or
  - Your experiment can be implemented as 2IFC
- For school projects and Master's theses, qualitative research is often a safer choice

Qualitative research primer

Thematic analysis

- A common qualitative GUR method
- Interview & open-ended questionnaire data
- Braun & Clarke is the most cited tutorial paper
- Procedure (roughly, see the paper for details):
  - Read through your data
  - Perform initial coding
  - Review and revise codes
  - Collate codes into themes
  - Write a report with illustrative quotes

www.QualResearchPsych.com *Qualitative Research in Psychology* 2006; **3:** 77–101

Using thematic analysis in psychology

Virginia Braun[1] and Victoria Clarke[2]

[1]University of Auckland and [2]University of the West of England

Thematic analysis is a poorly demarcated, rarely acknowledged, yet widely used qualitative analytic method within psychology. In this paper, we argue that it offers an accessible and theoretically flexible approach to analysing qualitative data. We outline what thematic analysis is, locating it in relation to other qualitative analytic methods that search for themes or patterns, and in relation to different epistemological and ontological positions. We then provide clear guidelines to those wanting to start thematic analysis, or conduct it in a more deliberate and rigorous way, and consider potential pitfalls in conducting thematic analysis. Finally, we outline the disadvantages and advantages of thematic analysis. We conclude by advocating thematic analysis as a useful and flexible method for qualitative research in and beyond psychology. *Qualitative Research in Psychology* 2006; **3:** 77–101

Key words: epistemology; flexibility; patterns; qualitative psychology; thematic analysis

Thematic analysis is a poorly demarcated and rarely acknowledged, yet widely used qualitative analytic method (Boyatzis, 1998; Roulston, 2001) within and beyond psychology. In this paper, we aim to fill what we, as researchers and teachers in qualitative psychology, have experienced as a current gap – the absence of a paper which adequately outlines the theory, application and evaluation of thematic analysis, and one which does so in a way accessible to students and those not particularly familiar with qualitative research.[1] That is, we aim to write a paper that will be useful as both a teaching and research tool in qualitative psychology. Therefore, in this paper we discuss theory and method for thematic analysis, and clarify

Correspondence: Virginia Braun, Department of Psychology, University of Auckland, Private Bag 92019, Auckland, New Zealand.
E-mail: v.braun@auckland.ac.nz

How coding works in practice

- Many possible coding styles:
  - Compact vs. verbose
  - Semantic vs. latent (semantic = as is, latent = reading between the lines)
- Practical tools:
  - Google Sheets (if your data can be formatted as one coded text per row)
  - Atlas.ti (free-form selection of coded texts in longer text documents)
- Coding in Atlas.ti: https://www.youtube.com/watch?v=SivD3Uwf6Ng

Content analysis

- Another common approach
- Also starts with reading and coding
- Codes/themes are less interpretative
- Inductive/open coding followed by deductive coding
- More focus on quantification of results, e.g., calculating frequencies of codes/themes

Review Article

Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study

Mojtaba Vaismoradi, PhD, MScN, BScN,[1,3] Hannele Turunen, PhD, RN[2] and Terese Bondas, PhD, RN[2,3]
[1]College of Human and Health Sciences, Swansea University, Swansea, UK, [2]Department of Nursing Science, Kuopio Campus, University of Eastern Finland, Kuopio, Finland and [3]Faculty of Professional Studies, University of Nordland, Bodø, Norway

Abstract Qualitative content analysis and thematic analysis are two commonly used approaches in data analysis of nursing research, but boundaries between the two have not been clearly specified. In other words, they are being used interchangeably and it seems difficult for the researcher to choose between them. In this respect, this paper describes and discusses the boundaries between qualitative content analysis and thematic analysis and presents implications to improve the consistency between the purpose of related studies and the method of data analyses. This is a discussion paper, comprising an analytical overview and discussion of the definitions, aims, philosophical background, data gathering, and analysis of content analysis and thematic analysis, and addressing their methodological subtleties. It is concluded that in spite of many similarities between the approaches, including cutting across data and searching for patterns and themes, their main difference lies in the opportunity for quantification of data. It means that measuring the frequency of different categories and themes is possible in content analysis with caution as a proxy for significance.

Key words content analysis, nursing, qualitative descriptive research, thematic analysis.

INTRODUCTION

In health care, qualitative methodologies aim to explore complex phenomena encountered by nurses, other providers, policy makers, and patients (Denzin & Lincoln, 2000; Sandelowski & Barroso, 2003a; Tong *et al*., 2007). The philosophy and the basic principles of methodologies, study aims and questions, and designs and data gathering criteria provide key differences between qualitative and quantitative methodologies (Ayres, 2007a). A belief in multiple realities, a commitment to identifying an approach to in-depth understanding of the phenomena, a commitment to participants' viewpoints, conducting inquiries with the minimum disruption to the natural context of the phenomenon, and reporting findings in a literary style rich in participant commentaries are the main characteristics of qualitative methodologies (Streubert Speziale & Carpenter, 2007).

Qualitative methodologies consist of the philosophical perspectives, assumptions, postulates, and approaches that researchers employ to render their work open to analysis, critique, replication, repetition, and/or adaptation and to choose research methods. In this respect, qualitative methodologies refer to research approaches as the tools with which researchers design their studies, and collect and analyse their data (Given, 2008). Qualitative methodologies are not a single research approach, but different epistemological perspectives and pluralism have created a range of "approaches" such as grounded theory, phenomenology, ethnography, action research, narrative analysis, and discourse analysis.

Qualitative research in the field of health has, at times, been undertaken without identification of the specific methodology used. The term "approach" is used in this article to differentiate it from the narrower term "methods." This indicates a coherent epistemological viewpoint about the nature of enquiry, the kind of knowledge discovered or produced, and the kind of strategies that are consistent with this (Giorgi, 1970; Holloway & Todres, 2005).

Qualitative approaches share a similar goal in that they seek to arrive at an understanding of a particular phenomenon from the perspective of those experiencing it. Therefore, the researcher needs to determine which research approach can answer their research questions (Streubert Speziale & Carpenter, 2007). There is a considerable overlap among available qualitative approaches in terms of methods, procedures, and techniques. Such an overlap of epistemological, aesthetic, ethical, and procedural concerns can encourage a generic view of qualitative research, considering it a "family" approach in which the similarities are more important than the differences, and where the notion of flexibility becomes an important value and quest. However, there is another point of view, concerned with how such flexibility can lead to inconsistency and a lack of coherence (Holloway &

Correspondence address: Hannele Turunen, Department of Nursing Science, Kuopio Campus, University of Eastern Finland, Kuopio, Finland. PO Box 1627, 70211 Kuopio. Email: hannele.turunen@uef.fi
Conflict of interest: None.
Received 20 March 2012; revision received 30 December 2012; accepted 28 January 2013.

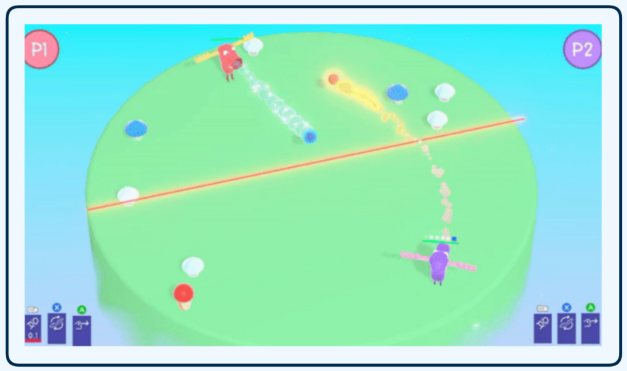
doi: 10.1111/nhs.12048

Exercise: Let’s get a feel of coding GUR data

- Make a copy of this data to your google drive and code in Google Sheets: https://urly.fi/32bU
- Coding instructions: *Please conduct a qualitative coding of the messages of a Discord support channel of Hypehype, a new game development tool. The codes provide compact summaries of the topics that users need help with, using one or just a few words. Multiple codes per message are separated by semicolons. If a message is not a help request, we format the code as Other (explanation). Common examples of those are Other (support response) and Other (support asks for clarification).*
- If some messages are particularly hard to code, enter “1” in the hard to code column

Collating codes to themes

- In the Hypehype data, what higher-level topic categories can you identify? (E.g., for grouping tutorials under main headers)
- Building a code hierarchy in Atlas.ti: https://www.youtube.com/watch?v=fEJD3SNtmBU



AhmadErfani.com/moosio

Issue

Suggestion

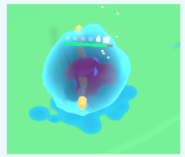
Codes

Visuals

- Map
  - Shape — We will have different maps
  - Transformation
  - Borders
- Customization — Character customization menu will be added
- Cool Down — can't notice

Add them to player space UI

- Water Effect — Misunderstanding



Make it faster

Mechanics

- Buttons
  - Spin
    - Hard to remember — Having training level?
    - Make them collectable? — Not for now
  - Dodge — intuitive enough
  - Aim
    - good for party games
    - Hard to rotate stationary — Fix it
- Bullet — More Bullet types
- Power Up — Having crazy rare collectable powerups — Cool Idea
- Obstacles — Destructible

Dynamic

- Very fun when both player have low health
  - Higher chance of getting good mushrooms when you are weak
  - Increase weak player bullets damage
  - Have e.g 3 rounds at each game

A note on interview technique

- Semi-structured interviews are the most common: You have a template of questions, but you can also improvise and ask follow-up questions, or skip a question if it was already answered
- Participants want to please you – “Is this good?” will produce biased answers.
  - Better to explicitly ask for positives, negatives, and suggestions for improvements

1st person methods

- Qualitative research with N=1
- Autobiographical design: Iterative design with the designer as the user
- Autoethnography: Researching and understanding something through personal participation, observation, and note taking

Introduction to the Special Issue on First-Person Methods in HCI

AUDREY DESJARDINS, School of Art + Art History + Design, University of Washington
OSCAR TOMICO, ELISAVA and Eindhoven University of Technology
ANDRÉS LUCERO, Aalto University
MARTA E. CECCHINATO, Northumbria University
CARMAN NEUSTAEDTER, School of Interactive Arts + Technology, Simon Fraser University

In this introduction to the special issue on First-Person Methods in (Human-Computer Interaction) HCI, we present a brief overview of first-person methods, their origin, and their use in Human-Computer Interaction. We also detail the difference between first-person methods, second-person, and third-person methods, as a way to guide the reader when engaging the special issue articles. We articulate our motivation for putting together this special issue: we wanted a collection of works that would allow HCI researchers to develop further, define, and outline practices, techniques and implications of first-person methods. We trace links between the articles in this special issue and conclude with questions and directions for future work in this methodological space: working with boundaries, risk, and accountability.

CCS Concepts: • **Human-centered computing** → **HCI design and evaluation methods**;

Additional Key Words and Phrases: First-person research, first-person methods, autobiographical design, autoethnography, somaesthetics, design research, HCI research

ACM Reference format:
Audrey Desjardins, Oscar Tomico, Andrés Lucero, Marta E. Cecchinato, and Carman Neustaedter. 2021. Introduction to the Special Issue on First-Person Methods in HCI. *ACM Trans. Comput.-Hum. Interact.* 28, 6, Article 37 (December 2021), 12 pages.
https://doi.org/10.1145/3492342

1 MOTIVATION FOR THIS SPECIAL ISSUE

Alongside the array of Human-Computer Interaction (HCI) methodological tools, first-person research methods offer a chance for researchers to not only investigate the mundane, ongoing, and ubiquitous presence of technology in everyday life, but also to acknowledge their own positionality in research and design, and to rely on first-hand experience as a mode of knowing. Through past and current usage of first-person methods, we notice how this shift in epistemological commitments has the potential of yielding rich, honest, and authentic reflections and insights about our ongoing lives with technology. We have seen, for example, the benefits for HCI and design researchers of using methods such as autoethnography (e.g., [8, 49, 67, 69]),

Authors' addresses: A. Desjardins, School of Art + Art History + Design, University of Washington, Box 353440, Seattle, WA 98195-3440 USA; email: adesjard@uw.edu; O. Tomico, ELISAVA and Eindhoven University of Technology (double affiliation), La Rambla 32, Barcelona, 08002, Spain; email: otomico@elisava.net; A. Lucero, Aalto University, Otaniementie 14, Espoo, 02150, Finland; email: lucero@acm.org; M. E. Cecchinato, Northumbria University, 2 Ellison Place, Newcastle upon Tyne, NE1 8ST, UK; email: marta.cecchinato@northumbria.ac.uk; C. Neustaedter, Simon Fraser University, 8888 University Dr., Burnaby, B.C. Canada, V5A 1S6; email: carman@sfu.ca.

1073-0516/2021/12-ART37
https://doi.org/10.1145/3492342

Transformative game experiences: an autoethnography

Title: Transformative game experiences: an autoethnography
Transformatiiviset pelikokemukset: autoetnografia
Author(s): Väkevä, Jaakko
Date: 2022-08-23
Language: en
Pages: 7 + 115
Major/Subject: Information Networks
Degree programme: Master's Programme in Information Networks
Supervising professor(s): Mekler, Elisa
Thesis advisor(s): Mekler, Elisa
Keywords: video games, art, transformative experience, self-reflection, autoethnography, empirical aesthetics
Location: Archive
OEV Access to electronic archive copy

Abstract:

Art is often hailed for its capacity to produce deeply profound experiences that can transform us as persons and alter our perspectives on life. This thesis sets out to explore, in the form of an autoethnography, how such a transformative power could manifest itself in the evolving digital medium of video games. For the study, I played through five critically acclaimed game titles in an attempt to detect and record transformative experiences by collecting both introspective and retrospective personal data of my game experiences in the form of field notes, reflective journals, and memos; the five games included in the study were God of War, Doki Doki Literature Club!, The Stanley Parable, Bloodborne, and The Beginner's Guide. I perceived my emotionally profound experiences of Doki Doki Literature Club! and The Beginner's Guide as transformative; both game experiences greatly widened my understanding of games as a medium, but also provided me with newfound perspectives on topics such as depression and creativity, as well as increased my capacity for empathy and self-understanding. Closer analysis and interpretation of the personal data resulted in identifying four overarching themes distinguishing my transformative game experiences from non-transformative ones; these themes can be characterized as follows: (1) a heightened state "cognitive and emotional involvement", (2) a conception of "playing as myself", (3) an impression of "unconventional game design", and (4) post-play "re-engagement with the game experience". I then examined these themes in a wider context, discussing their relations to existing literature and their possible broader implications regarding transformative game experiences. It seems that several aspects of my personal transformative game experience could potentially correspond to concepts that in the field of empirical aesthetics have been used to model transformative art perception outcomes, but further research is required.

Autobiographical Design in HCI Research: Designing and Learning through Use-It-Yourself

Carman Neustaedter
School of Interactive Arts + Technology
Simon Fraser University
Surrey, BC, Canada
carman_neustaedter@sfu.ca

Phoebe Sengers
Information Science
Cornell University
Ithaca, NY, USA
sengers@cs.cornell.edu

ABSTRACT
Designing a system with yourself as a target user and evaluating the design through your own self-usage is commonly considered a questionable approach in HCI research. Perhaps for this reason, HCI research including extensive self-usage of a design is underdocumented. Yet such self-usage *does* happen and many researchers have found great value in the lessons learned from it. Our goal in this paper is to bring these hidden practices to light and offer guidelines for how HCI researchers can usefully engage in what we term 'autobiographical design'—design research drawing on extensive, genuine usage by those creating or building a system. Through interviews with HCI experts who have engaged in variations of autobiographical design, we draw out the possibilities and limitations of autobiographical design methods and lay out best practices for its use as an HCI research method.

Author Keywords
Autobiographical design; design research methods

ACM Classification Keywords
H.5.2 [Information interfaces and presentation]: User Interfaces – Evaluation/methodology;

INTRODUCTION

User-centered design is a cornerstone of the field of HCI, since incorporating users in the design process can lead to better systems [35]. There are many methods for incorporating end users into design research, ranging from participatory design to lab or field evaluations [35]. There are also design methods that focus on individual, idiosyncratic experiences as a basis for design. Cooper, for example, emphasizes "designing for one" through creating personas, or fictional characterizations of specific users [11]. Others argue that designers should focus on *non*-typical users such as extreme characters [14], atypical user groups [23], or lead users [42] as a source of innovation.

DIS 2012, June 11-15, 2012, Newcastle, UK.

But what about when those individual users are the researchers or designers themselves? Nobody would contest that designers, developers, and researchers frequently use their own systems during design in order to test concepts, learn through actual usage, or find and fix software bugs. In fact, many would argue this is an important step *before* putting the design before other end users. Though never truly formalized as a research method, such "eat your own dog food" methods have been applied in industry for years [10,15]. Beyond this, some systems are used throughout their design based on designers' or researchers' real needs. For example, early research from CSCW reported many lessons from media spaces based on technologists' use of their own systems or that of their colleagues, based on intensive and authentic self-usage (e.g., [7,16,33]).

Despite these examples of learning and designing through self-usage, we know of few accounts of this as a reported HCI research method (the notable exception is [17]). In fact, HCI research methodologies focusing on researchers' own experiences as users of a system may seem heretical to both the ethical goal of user-centered design and the epistemological goal of basing research in HCI on objective, third-party knowledge. Yet self-usage by researchers *does* happen, whether it is reported or not. Thus, as a community, we feel it is important that we better understand this method and how it should be used.

Our goal, then, is to shed light on the research practice that we term **autobiographical design:** *design research drawing on extensive, genuine usage by those creating or building the system*. (By 'genuine' usage we mean it is based on true needs of the researchers, rather than them pretending to have needs expected of targeted users). We chose the term 'autobiographical design' as we felt a core attribute of the method is that a researcher or designer's own experiences are embodied in the design of a system and its exploration. That is, as the researcher(s) build the system, they use it themselves, learn about the design space, and evaluate and iterate the design based on their own experiences. Within this definition, there are certainly nuances, and our paper explores these.

Our interest in this topic began with our own projects. Neustaedter designed two systems for himself and his family, using and studying each system over the course of a year [27,34]. Sengers used a similar approach for three

Revealing Tensions in Autobiographical Design in HCI

Audrey Desjardins and Aubree Ball
School of Art + Art History + Design
University of Washington
Seattle, WA USA
adesjard@uw.edu, aubreeb@uw.edu

ABSTRACT
While self-usage has long been regarded as a questionable approach in human-computer interaction (HCI) research, recent projects have shown the successful use of autobiographical design as a method to investigate long-term and intimate relations between people and technologies in everyday life. In an effort to continue the development of methodological best practices, we need to acknowledge with more nuance the tensions that arise in use. In this paper, we articulate such tensions by examining two first-hand accounts of using autobiographical design and four autobiographical design projects of other HCI researchers. Our findings address: genuine needs, design participation, intimacy, reflexivity, and authorial voice. Our contribution is constituted of critical insights into the complexities of using autobiographical design and recommendations for researchers interested in using this method.

Author Keywords
Autobiographical design; autoethnography; first person research; design research.

ACM Classification Keywords
H.5.m Miscellaneous

INTRODUCTION

As computing continues to enter many aspects of our daily lives, HCI researchers have expanded their research methods to better investigate the complex and multifaceted relationships between humans and computers. In the context of domestic technologies, for instance, HCI researchers have developed and adapted methods such as ethnography and ethnomethodology inspired inquiries (e.g., [3,10,40,44]), field studies and prototype deployments (e.g., [16,28]), cultural and technology probes (e.g., [19,26]), design research and research through design (e.g., [21,29,33]), and participatory design (e.g., [36]).

DIS '18, June 9–13, 2018, Hong Kong

ACM ISBN 978-1-4503-5198-0/18/06…$15.00
https://doi.org/10.1145/3196709.3196781

In addition to these methods, first-person research perspectives have started to become more prominent in HCI research, particularly as research foci move from public/work spheres to private spheres of life [6]—where situations are intimate and require long-term investigation, for example in areas like the home [13], self-tracking [43], and wearables [5]. Including autoethnography [11,17] and autobiographical design [32], these methods grant a central place to the researcher's experience in the design and/or long-term usage of a system. By acknowledging the role, the perspective, and the experience of the researcher, this methodological standpoint provides a provocative and radically different approach to the more commonly used third-person perspectives in HCI research.

In this paper, we bring our attention specifically to autobiographical design research, defined by Neustaedter and Sengers as "*design research drawing on extensive, genuine usage by those creating or building the system*" [32:514]. At its core, the designer or maker is simultaneously the user of the system created. This allows for a deep understanding of the user needs as well as the constraints in the design and construction of the system. The method holds benefits such as allowing long-term studies, responding to a genuine need, supporting fast iterative tinkering, and opening access to intimate and personal spaces and situations [32]. In our paper, we focus on autobiographical design because, as design researchers, we are interested in understanding not only how interactive systems are lived with but also how they are designed, built, repaired, and iteratively appropriated.

In the last decade, autobiographical design accounts have become more frequent and common in HCI research (e.g., [7,8,20,23,24,41]). However, while reporting on the findings of their studies, researchers rarely share the details about how they conducted their autobiographical design, the challenges they encountered, and the subtle ways they might have adjusted the method to their specific context. As a result, the complexities of conducting autobiographical design research are still widely underexplored. In fact, autobiographical design is often portrayed as a novel method, where novelty eclipses deeper issues around the tensions that may arise when using the method. If HCI researchers are increasingly considering using autobiographical design research as a way to study intimate, long-term, and personal relations between computers and humans, it is crucial that we articulate the complexities of using this method with more precision and

Other methods

Concurrent think aloud

- Common qualitative data collection method
- Players asked to narrate their experience: What they perceive, think & feel, and what decisions they make and why
- In practice, players have difficulty remembering specific instructions. If you need something specific, remember to run a pilot study before larger-scale data collection

Video-assisted stimulated recall

- Stimulated recall = using material like photos to assist recall.
- Watching and discussing game test video is a good way to let the users narrate their experience and highlight important parts that they forgot to mention, or didn’t mention because of focusing on the game

Cognitive & pluralistic walkthroughs

Initial usability & learnability evaluation based on game or UI sketches or paper prototypes

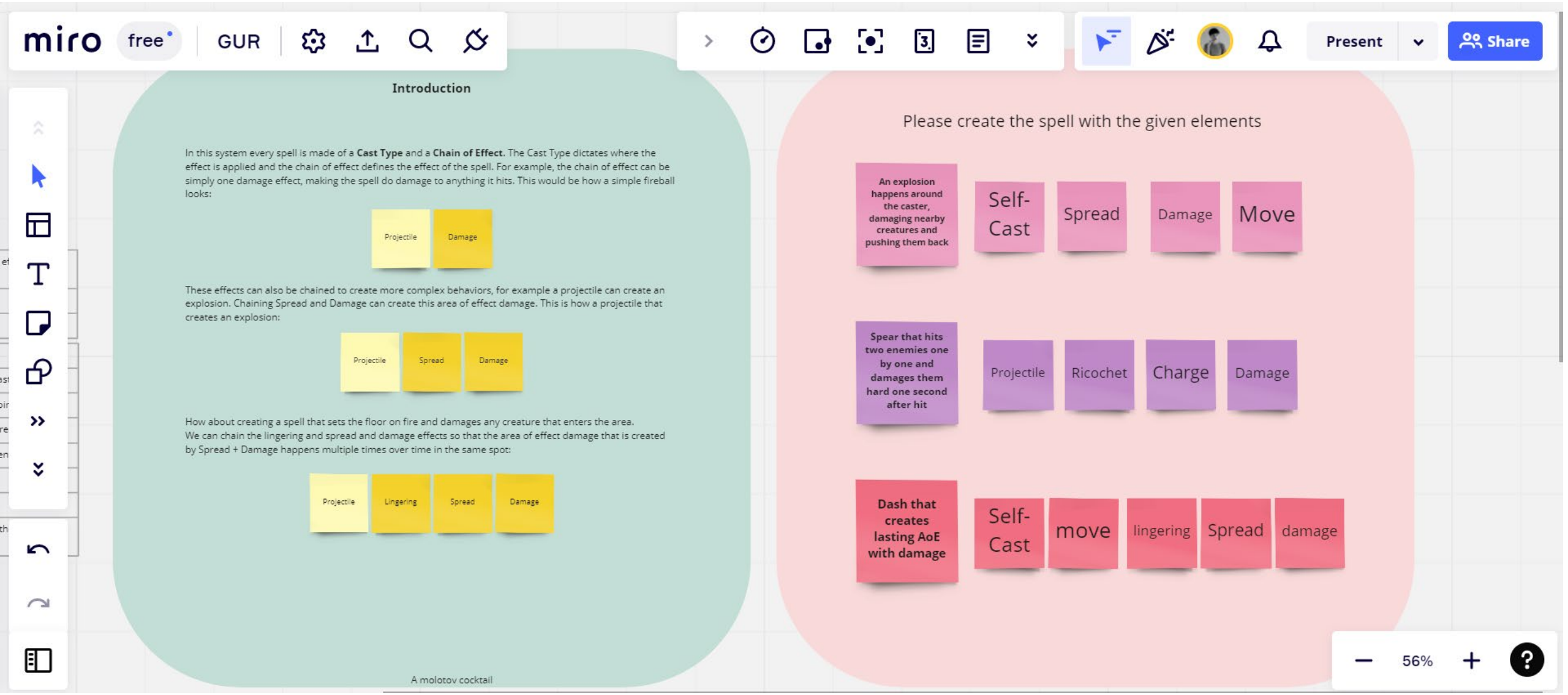
For instance, participants can be shown an UI sketch and asked to:

- Explain how they would perform a task
- Explain what UI elements mean and how they work

https://www.nngroup.com/articles/cognitive-walkthrough-workshop/

https://en.wikipedia.org/wiki/Pluralistic_walkthrough

Paper prototyping with Miro boards



Heuristic evaluation

- Expert evaluation based on a checklist

Heuristic Evaluation for Games: Usability Principles for Video Game Design

David Pinelle
University of Nevada, Las Vegas
4505 Maryland Parkway
Las Vegas, NV 89154-4019
pinelle@cs.unlv.edu

Nelson Wong
University of Saskatchewan
110 Science Place
Saskatoon, SK, Canada S7N 5C9
nelson.wong@usask.ca

Tadeusz Stach
Queen's University
25 Union Street, Goodwin Hall
Kingston, ON, Canada K7L 3N6
tstach@cs.queensu.ca

ABSTRACT
Most video games require constant interaction, so game designers must pay careful attention to usability issues. However, there are few formal methods for evaluating the usability of game interfaces. In this paper, we introduce a new set of heuristics that can be used to carry out usability inspections of video games. The heuristics were developed to help identify usability problems in both early and functional game prototypes. We developed the heuristics by analyzing PC game reviews from a popular gaming website, and the review set covered 108 different games and included 18 from each of 6 major game genres. We analyzed the reviews and identified twelve common classes of usability problems seen in games. We developed ten usability heuristics based on the problem categories, and they describe how common game usability problems can be avoided. A preliminary evaluation of the heuristics suggests that they help identify game-specific usability problems that can easily be overlooked otherwise.

Author Keywords
Usability evaluation, heuristic evaluation, video games

ACM Classification Keywords
H5.2 [User Interfaces]: Evaluation/methodology

INTRODUCTION
One of the main goals in video game design is to entertain and engage the user. This can involve several aspects of design, including game story, pacing, challenge level, and game mechanics [7]. However, since most games require constant interaction, game designers must also pay careful attention to usability issues. Failure to design usable game interfaces can interfere with the larger goal of creating a compelling experience for users, and can have a negative effect on the overall quality and success of a game.

CHI 2008, April 5–10, 2008, Florence, Italy.

In this paper, we define game usability as *the degree to which a player is able to learn, control, and understand a game.* Our definition is based on an early informal survey of usability problems cited in critical game reviews and on playability heuristics described by Federoff [12] and Desurvire et al. [7]. Game usability does not address issues of entertainment, engagement, and storyline, which are strongly tied to both artistic issues (e.g. voice acting, writing, music, and artwork) and technical issues (graphic and audio quality, performance issues).

Some usability issues seen in games are similar to those seen in other application areas, such as the need to design for visual consistency and readability. However, games also have usability considerations that are not seen in other areas. For example, user errors are usually undesirable in other domains, but are expected in many games since they are designed to challenge users and to force them to develop new skills so that they can achieve in-game objectives.

Game designers need methods for identifying usability problems both in early designs and in more mature prototypes [7]. Playtesting is one of the most common ways to uncover design problems [13], yet this method needs a playable prototype that only exists in the later stages of the development process. Formal methods do not exist to allow designers to carry out less expensive usability inspections of games, and to evaluate early, non-functional prototypes. Further, many common usability inspection techniques are not appropriate for games since they either rely on formal specifications of task sequences [3,4,19,29] or are oriented around user interface concepts used in desktop applications [24].

One technique that has the potential to be useful in allowing evaluations of game prototypes is heuristic evaluation. Heuristic evaluation is an inspection technique where evaluators explore an interface using a set of usability principles, called heuristics [24]. Heuristic evaluation does not make assumptions about task structure, and it is flexible enough to be adapted to specialized domains [25,10].

We believe that a custom set of design principles are needed so that heuristic evaluation can be used to find usability problems in video games. Several researchers have

Problem category	Key issues	Example	Total
1. Unpredictable / inconsistent response to user's actions	poor hit detection, poor in-game physics, inconsistent response to input	"You'll often find yourself swinging away and watching your blade go right through your foes with no effect whatsoever." [21]	42
2. Does not allow enough customization	does not allow user to change video and audio settings, difficulty, or game speed	"Worst of all, there is no way to adjust the game's speed. You're stuck constantly veering between way-too-fast combat and way-too-slow travel to engage enemies and explore maps." [34]	11
3. Artificial intelligence problems	problems with pathfinding, problems with computer controlled teammates	"Your own teammates are a mixed bag in terms of intelligence, as they're sometimes good about getting behind cover and laying down covering fire, but they also do crazy things, like rushing forward into the open or constantly getting in your line of fire." [27]	22
4. Mismatch between camera/view and action	bad camera angle, view is obstructed, view does not adjust to user's action quickly enough	"If you leave the camera alone you'll either be running around blind or the camera will get stuck on something and just twitch randomly for a while." [22]	23
5. Does not let user skip non-playable content	cannot skip video and audio clips, frequently repeated sequences	"Making matters even worse, you can't skip over any of this blah-blah-blah wordiness and often have to exhaust all of the dialogue options to open up new discussion topics or the ability to solve a problem." [35]	6
6. Clumsy input scheme	bad input mappings, limited device support, limited control customization	"You can't type your name in—you must instead select letters with your mouse. And you can't use the keyboard to navigate." [14]	21
7. Difficult to control actions in the game	oversensitive controls, unnatural controls, unresponsive controls	"Chopper controls run the gamut from very touchy to absurdly touchy. For instance, it's almost impossible to keep an Osprey's rotors pointed toward the sky." [36]	31

Heuristic Evaluation for Gameful Design

Gustavo F. Tondello
HCI Games Group
University of Waterloo
Waterloo, ON, Canada
gustavo@tondello.com

Marim Ganaba
HCI Games Group
University of Waterloo
Waterloo, ON, Canada
mariganaba@gmail.com

Dennis L. Kappen
HCI Games Group
University of Ontario Institute of Technology
Oshawa, ON, Canada
dennis.kappen@humber.ca

Lennart E. Nacke
HCI Games Group
University of Waterloo
Waterloo, ON, Canada
lennart.nacke@acm.org

Elisa D. Mekler
Center for Cognitive Psychology & Methodology
University of Basel
Basel, Switzerland
HCI Games Group
University of Waterloo
elisamekler@gmail.com

CHI PLAY'16 Extended Abstracts, October 16-19, 2016, Austin, TX, USA
ACM 978-1-4503-4458-6/16/10.
http://dx.doi.org/10.1145/2968120.2987729

Abstract
Despite the emergence of many gameful design methods in the literature, there is a lack of evaluation methods specific to gameful design. To address this gap, we present a new set of guidelines for heuristic evaluation of gameful design in interactive systems. First, we review several gameful design methods to identify the dimensions of motivational affordances most often employed. Then, we present a set of 28 gamification heuristics aimed at enabling experts to rapidly evaluate a gameful system. The resulting heuristics are a new method to evaluate user experience in gameful interactive systems.

Author Keywords
Heuristic Evaluation; Gamification; Gameful Design.

ACM Classification Keywords
H.5.2. User Interfaces: Evaluation methodology; H.5.3. Group Interfaces: Evaluation methodology.

Introduction
Gamification, using game design elements in non-game contexts [9], complements user experience (UX) methods to increase user motivation and engagement in many different fields, such as education, health, social networks, and business [14,33,38]. Some gameful design frameworks and methods have been suggested

More methods

- https://en.wikipedia.org/wiki/Comparison_of_usability_evaluation_methods

Example of a brief qualitative GUR study

Research goal

Are the game rules understandable and intuitive?

Research questions:

1. Which game rules player can figure out on their own?
2. Which game rules are considered difficult/unclear/unintuitive?

Process

Playtest

1. Start voice recording
2. Explain the basic rules
3. Playtest
    - Until the player is stuck
    - Let player share ideas/feedback
    - Ask for more concrete answers as useful

Analysis

1. Generate transcript with Microsoft Word Transcribe feature
2. Assign codes for player comments with Atlasti
3. Analyze the results based on frequency of topics

RESULTS

7 playtesters

23 codes

Total codes (unique codes)

* Result underestimates the experience of researcher

Results

1.	**Controls**		**8 (7)**
	1.	**Camera controls**	**4 (3)**
	2.	Keyboard controls	2
	3.	Planting controls	2
2.	**Planting**		**10 (7)**
	1.	Collecting seeds 1 up	2
	2.	**Planting**	**8 (5)**
3.	**Movement**		**8 (4)**
	1.	Movement 1 up	2
	2.	Movement 2 down	2 (1)
	3.	**Getting on top of a tree**	**4 (3)**
4.	**Water**		**6 (5)**
	1.	**Lowering water level**	**5 (4)**
	2.	Rising water	1
5.	**Unclear visuals**		**8 (7)**
	1.	"Connected" trees	1
	2.	**Collecting seeds**	**3**
	3.	Distinguishing height	3 (2)
	4.	Starting with a seed	1
	5.	**Bug: planted tiles color**	**2***
6.	**Other**		**10 (7)**
	1.	Afraid of wrong moves	1
	2.	**Difficulty curve**	**5 (2)***
	3.	Suggestion: show path	1

The Unspoken rules

Cannot move more than 1 step upwards	0
Cannot move not into flooded space	0
Can move 2+ downwards	1
Tree must have seeds to pick seeds	3
Character must be on the tree level to pick seeds	2
Tree grows only after next move action	**5**
Movement/Collect -> Water Level increases	1
Plant -> Water Level decreases	**4**

Further analysis

Camera controls **(3)**

- Difficult to control with mouse (either too fast or slow)
- Suggestion to use arrow keys / 90 degree angles

Planting **(5)**

- Tree grows only after player move away -> feedback of planting not clear
- Might also relate to a visual bug and unintuitive control scheme
- Improvement: single UI button instead of mouse click

Getting on top of a tree **(3)**

- 3 testers figured out without instructions (”This is amazing. You should make puzzles with this.”)
- Improvement: tutorial

Lowering water level **(4)**

- ” I think it probably relates to the trees.”, ”I wasn't sure how much that's planting lower the sea level”, ”Wasn't sure how much planting lowers the sea level.”, ”That’s maybe a bit unintuitive.”

Collecting seeds **(3)**

- Visually unclear which trees have seeds -> improve visuality

Difficulty Curve **(2*)**

- But only 1 playtester solved the second level.

DISCUSSION

Potential issues

- Only used recording to save playtesting/interview
  - Poor recording quality -> machine made transcript was barely understandable
- No prewritten question list
  - Playtesters got slightly different set of questions depending on what happened during the gameplay
- Difficulty curve & poor on boarding
  - 4 test levels trying to include main concepts of the game -> difficulty increases too fast
  - The experience of the playtesters with the current version does not match the real version

Conclusions

- Basic rules of the game like movement were understood by everyone
- More game spesific rules were more difficult to understand, especially planting
- Some testers also had confusion about suprising game rules like the ability to move one step upwards.
- Improving the visual feedback and controls is necessary
- Slow onboarding, introducing all game rules separately
- Potentially redesigning planting mechanism

AI-assisted data analysis & generation

Overview

- GPT-3 and other Large Language Models can role-play participants and answer to virtual interview questions
- GPT-3 can also do automatic qualitative coding, if given a few examples
- Models like OpenAI Whisper can automatically transcribe think aloud and interviews.
- Deep Neural Networks can analyze facial expressions and the tone of player voice
- Deep Reinforcement Learning agents can learn to play games and use interfaces in simulation, and simulation data has at least some predictive power about real user behavior

AI-assisted data generation

AI playtesting

- Data from AI players can predict some aspects of human data
- For example, AI pass rate of levels can be proportional to human pass rate, e.g., AI=*a**human + *b*
- If you measure *a* and *b* on some players and levels, can generalize to other players and levels

Predicting Game Difficulty and Engagement Using AI Players

SHAGHAYEGH ROOHI, Aalto University, Finland
CHRISTIAN GUCKELSBERGER, Aalto University, Finland
ASKO RELAS, Rovio Entertainment, Finland
HENRI HEISKANEN, Rovio Entertainment, Finland
JARI TAKATALO, Rovio Entertainment, Finland
PERTTU HÄMÄLÄINEN, Aalto University, Finland

This paper presents a novel approach to automated playtesting for the prediction of human player behavior and experience. We have previously demonstrated that Deep Reinforcement Learning (DRL) game-playing agents can predict both game difficulty and player engagement, operationalized as average pass and churn rates. We improve this approach by enhancing DRL with Monte Carlo Tree Search (MCTS). We also motivate an enhanced selection strategy for predictor features, based on the observation that an AI agent's best-case performance can yield stronger correlations with human data than the agent's average performance. Both additions consistently improve the prediction accuracy, and the DRL-enhanced MCTS outperforms both DRL and vanilla MCTS in the hardest levels. We conclude that player modelling via automated playtesting can benefit from combining DRL and MCTS. Moreover, it can be worthwhile to investigate a subset of repeated best AI agent runs, if AI gameplay does not yield good predictions on average.

CCS Concepts: • **Human-centered computing** → **User models**; • **Computing methodologies** → *Modeling and simulation.*

Additional Key Words and Phrases: Player Modelling, AI Playtesting, Game AI, Difficulty, Player Engagement, Pass Rate Prediction, Churn Prediction, Feature Selection

ACM Reference Format:
Shaghayegh Roohi, Christian Guckelsberger, Asko Relas, Henri Heiskanen, Jari Takatalo, and Perttu Hämäläinen. 2021. Predicting Game Difficulty and Engagement Using AI Players. *Proc. ACM Hum.-Comput. Interact.* 5, CHI PLAY, Article 231 (September 2021), 17 pages. https://doi.org/10.1145/3474658

1 INTRODUCTION

The development of a game typically involves many rounds of playtesting to analyze players' behaviors and experiences, allowing to shape the final product so that it conveys the design intentions and appeals to the target audience. The tasks involved in human playtesting are repetitive and tedious, come with high costs, and can slow down the design and development process substantially. Automated playtesting aims to alleviate these drawbacks by reducing the need for

Authors' addresses: Shaghayegh Roohi, Aalto University, Espoo, Finland, shaghayegh.roohi@aalto.fi; Christian Guckelsberger, Aalto University, Espoo, Finland, christian.guckelsberger@aalto.fi; Asko Relas, Rovio Entertainment, Espoo, Finland; Henri Heiskanen, Rovio Entertainment, Espoo, Finland; Jari Takatalo, Rovio Entertainment, Espoo, Finland; Perttu Hämäläinen, Aalto University, Espoo, Finland, perttu.hamalainen@aalto.fi.

2573-0142/2021/9-ART231 $15.00
https://doi.org/10.1145/3474658

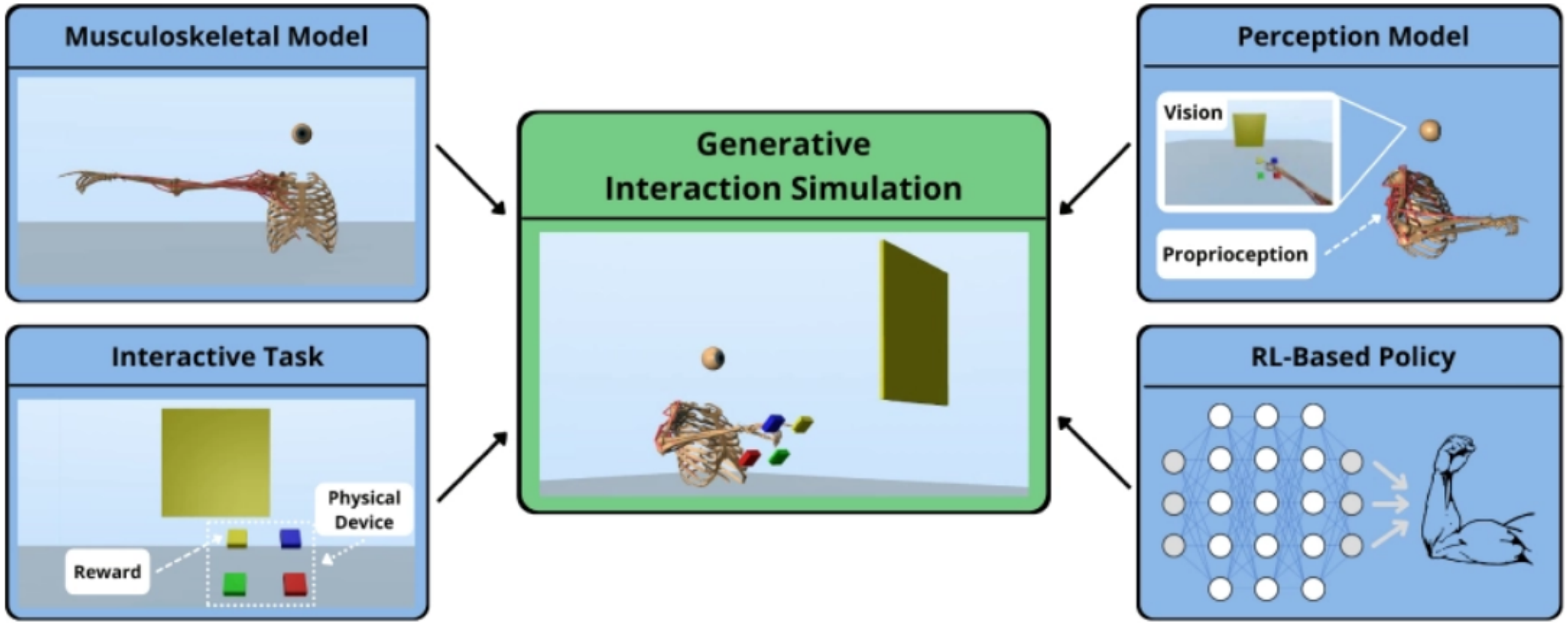
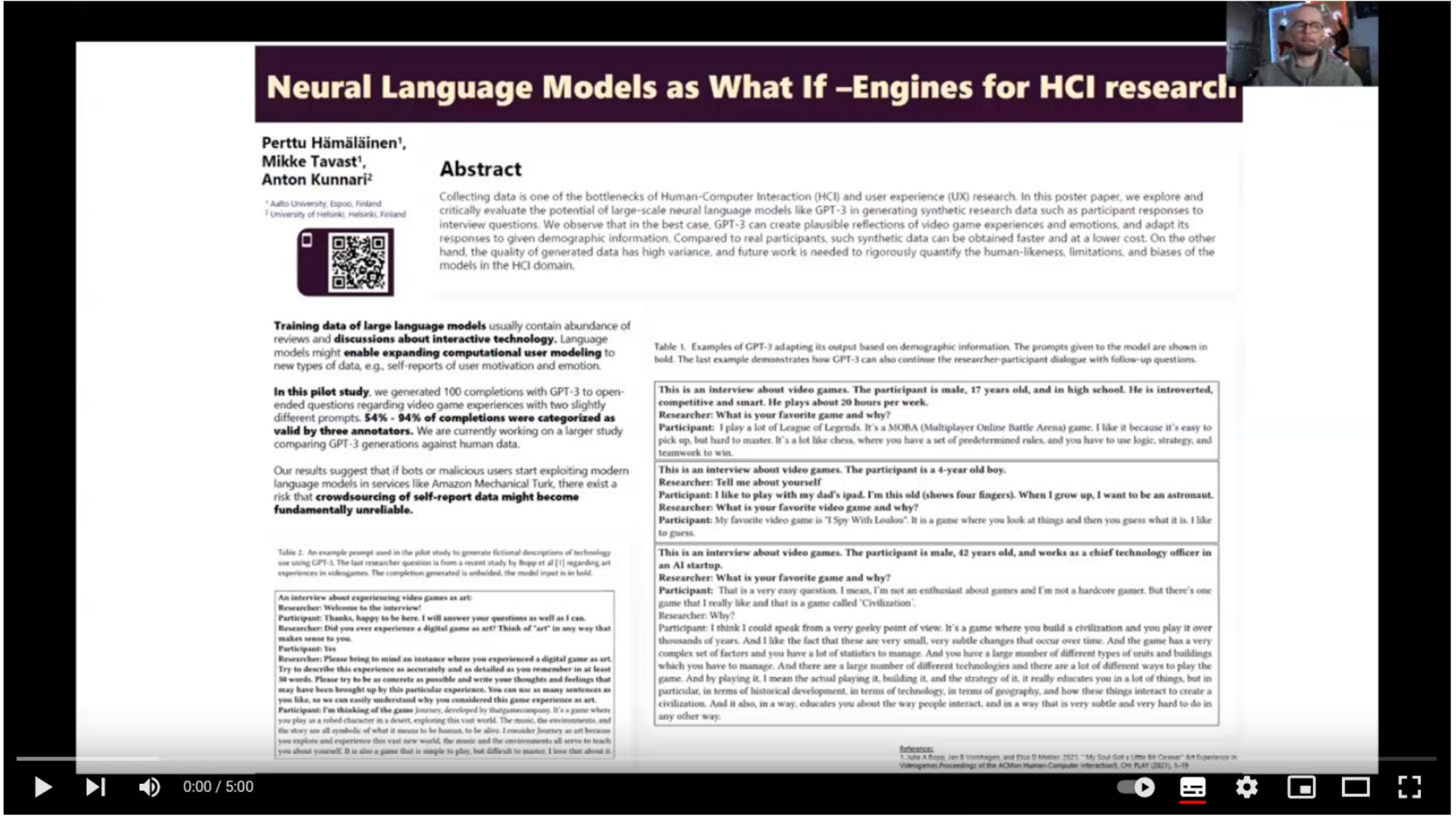


Figure 1: We present an approach for generative simulation of interaction with perceptually controlled biomechanical models interacting with physical devices. The users are modelled with a combination of muscle-actuated biomechanical models and perception models, and we use deep reinforcement learning to learn control policies by maximizing task-specific rewards. As a showcase, we apply a state-of-the-art upper body model to four HCI tasks of increasing difficulty: pointing, tracking, choice reaction, and parking a remote control car via joystick.

https://dl.acm.org/doi/fullHtml/10.1145/3526113.3545689



Neural Language Models as What If -Engines for HCI Research (IUI 2022 poster presentation)

Perttu Hämäläinen
248 subscribers

Subscribe

https://www.youtube.com/watch?v=SuZu4Grfb_I

0 Share Clip Save

Language Models Can Generate Human-Like Self-Reports of Emotion

Mikke Tavast
mikke.tavast@aalto.fi
Aalto University
Espoo, Finland

Anton Kunnari
anton.kunnari@helsinki.fi
University of Helsinki
Helsinki, Finland

Perttu Hämäläinen
perttu.hamalainen@aalto.fi
Aalto University
Espoo, Finland

ABSTRACT

Computational interaction and user modeling is presently limited in the domain of emotions. We investigate a potential new approach to computational modeling of emotional response behavior, by using modern neural language models to generate synthetic self-report data, and evaluating the human-likeness of the results. More specifically, we generate responses to the PANAS questionnaire with four different variants of the recent GPT-3 model. Based on both data visualizations and multiple quantitative metrics, the human-likeness of the responses increases with model size, with the largest Davinci model variant generating the most human-like data.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**.

KEYWORDS

Language models, GPT-3, PANAS, emotion, affect

ACM Reference Format:
Mikke Tavast, Anton Kunnari, and Perttu Hämäläinen. 2022. Language Models Can Generate Human-Like Self-Reports of Emotion. In *27th International Conference on Intelligent User Interfaces (IUI '22 Companion), March 22–25, 2022, Helsinki, Finland.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3490100.3516464

1 INTRODUCTION

Computational user modeling is advancing rapidly and can produce human-like predictions of user behavior and experience [7, 12, 15]. However, this is presently limited in the domain of emotions. Here, our aim is to investigate a potential new approach by using modern neural language models to generate synthetic self-report data about affect. We employ the recent GPT-3 model to generate responses to Positive and Negative Affect Schedule (PANAS), a widely used questionnaire designed to measure positive and negative affect [21]. GPT-3 is a large neural language model trained to predict the next word in a sequence [5]. The trained model takes as its input a piece of text—a prompt—and generates a continuation of desired length.

PANAS consists of 10 positive affect and 10 negative affect items: emotional words such as excited, proud, guilty, and upset. The task is to rate how much one has felt these states during a specified time period. In the original validation study [21] each of the 20 items was shown to strongly load on only one of the two largely uncorrelated factors, establishing the 10 item positive and negative scales. In HCI research, PANAS has been used to, for example, operationalize positive user experience [16, 19].

We are not aware of previous studies trying to directly predict psychological scale responses using language models. Other natural language processing (NLP) methods have been used, for example, to predict scale responses [3] and affective ratings of words [20], and a recent preprint compared transformer models to human data in a benchmark NLP task [13]. Other work has, for example, used transformer model representations to predict brain imaging data [6, 10, 18].

2 METHODS

We generated 150 completions to the 20 items of the PANAS scale with four GPT-3 models of increasing size: Ada, Babbage, Curie, and Davinci[1]. The responses to the PANAS items were generated with a prompt that described a research interview (see Table 1). To increase the variability in the prompts, each of the 150 interviews had a unique "participant" (varying age, gender, job, and hobby), a short description of "who" is being interviewed.

After the participant information, the prompt contained three example responses to questions answered in a Likert scale. To minimize bias in the training examples, the "participant" gave the answers 1, 3, and 5 once in each interview. The same three examples were used throughout the experiment, but their order was randomized for every interview.

The 20 PANAS items were queried one-by-one, in random order. Once the model generated a completion to a item, everything strating from the first appearance of the string "Researcher:" or the first newline character was cut from the completion. If the completion generated something outside desired responses (1,2,3,4 or 5), the whole interview for the participant was run again. The number of errors in proportion to all of the completions are presented in Section 2.1. After trimming the completion, the item-response pair was saved and included in the prompt for the next item. The order of the previous item-response pairs in the prompt was randomized for every new completion. For response generation, we used a maximum response length of 64 tokens and the default OpenAI parameters (temperature=0.7, top_p=1.0,frequency_penalty=0,presence_penalty=0, best_of=1).

We downloaded a human PANAS reference dataset from Open Science Framework (osf.io). The dataset was originally collected for a study by Anvari and Lakens concerning methods of determining the smallest effect size of interest [2]. Here we use the datapoints designated as T1 in the dataset (https://osf.io/3a5up/, [1]).

[1]Data and analysis code: https://github.com/mtavast/gpt-panas

IUI '22 Companion, March 22–25, 2022, Helsinki, Finland

ACM ISBN 978-1-4503-9145-0/22/03.
https://doi.org/10.1145/3490100.3516464

Evaluating Large Language Models in Generating Synthetic HCI Research Data: a Case Study

Perttu Hämäläinen*
perttu.hamalainen@aalto.fi
Aalto University
Espoo, Finland

Mikke Tavast*
mikke.tavast@aalto.fi
Aalto University
Espoo, Finland

Anton Kunnari
anton.kunnari@helsinki.fi
University of Helsinki
Helsinki, Finland

ABSTRACT

Collecting data is one of the bottlenecks of Human-Computer Interaction (HCI) research. Motivated by this, we explore the potential of large language models (LLMs) in generating synthetic user research data. We use OpenAI's GPT-3 model to generate open-ended questionnaire responses about experiencing video games as art, a topic not tractable with traditional computational user models. We test whether synthetic responses can be distinguished from real responses, analyze errors of synthetic data, and investigate content similarities between synthetic and real data. We conclude that GPT-3 can, in this context, yield believable accounts of HCI experiences. Given the low cost and high speed of LLM data generation, synthetic data should be useful in ideating and piloting new experiments, although any findings must obviously always be validated with real data. The results also raise concerns: if employed by malicious users of crowdsourcing services, LLMs may make crowdsourcing of self-report data fundamentally unreliable.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**.

KEYWORDS

User experience, User models, Language models, GPT-3

ACM Reference Format:
Perttu Hämäläinen, Mikke Tavast, and Anton Kunnari. 2023. Evaluating Large Language Models in Generating Synthetic HCI Research Data: a Case Study. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3544548.3580688

1 INTRODUCTION

HCI Researcher: Dear AI, do you think you could role-play a user study participant for me?

AI: *Sure, I can try.*

Researcher: Let's imagine you're a scientist and the user study is about academic writing tools. What is your favorite writing tool and why?

AI: *My favorite writing tool would have to be Microsoft Word. I find it very user-friendly and versatile, and it has everything I need to get the job done.*

Researcher: Why do you prefer Word over LaTeX?

AI: *LaTeX is a great tool for typesetting, but I find it to be much more difficult to use than Word. With Word, I can easily format my document the way I want it to look, and I don't have to worry about code.*

The dialogue above was generated using OpenAI Playground[1], a tool that allows one to input a piece of text—*a prompt*—and ask the GPT-3 large language model (LLM) [10] to generate a plausible continuation. We wrote the boldface parts and let GPT-3 generate the italicized continuations. The result is characteristic of the phenomenon we investigate in this paper: Through learning to model and predict various kinds of human-produced texts ranging from technical documentation to online discussions and poetry, LLMs like GPT-3 give the appearance of "understanding" human experiences such as interactive product use. Of course, the internal operation of the models differs from the internal psychological and neurophysiological processes of humans—LLMs simply learn to predict the next symbol (or impute missing symbols) in a sequence. Nevertheless, on a purely behavioral level, the results can be very human-like.

Much of HCI research is conducted using verbal data such as interviews and questionnaires (e.g., [3, 61, 72]), but collecting such data can be slow and expensive. Therefore, the above suggests that *LLMs might be useful in generating synthetic/hypothetical data for HCI research*, a notion we explore empirically in this paper. LLMs are typically trained on enormous Internet datasets such as Common Crawl [67]), including an abundance of online discussions about interactive technology and products such as phones, computers, and games. Therefore, it seems plausible that LLMs could generate, e.g., realistic 1st-person accounts of technology use, and answer natural language questions about user experiences, motivations, and emotions. We emphasize that we do not claim that such synthetic LLM data could ever be a replacement for data from real human participants. We simply consider that synthetic based data might be useful in some contexts, for example, when piloting ideas or designing an interview paradigm.

In effect, we view LLMs as a new kind of search engine into the information, opinions, and experiences described in their Internet-scale training data. Unlike traditional search engines, LLMs can be queried in the form of a narrative such as a fictional interview. Furthermore, LLMs exhibit at least some generalization capability to new tasks and data (e.g., [45, 71, 81]). This presents an untapped opportunity for counterfactual *What if?* exploration, e.g., allowing

*Perttu Hämäläinen and Mikke Tavast contributed equally to this work.

CHI '23, April 23–28, 2023, Hamburg, Germany

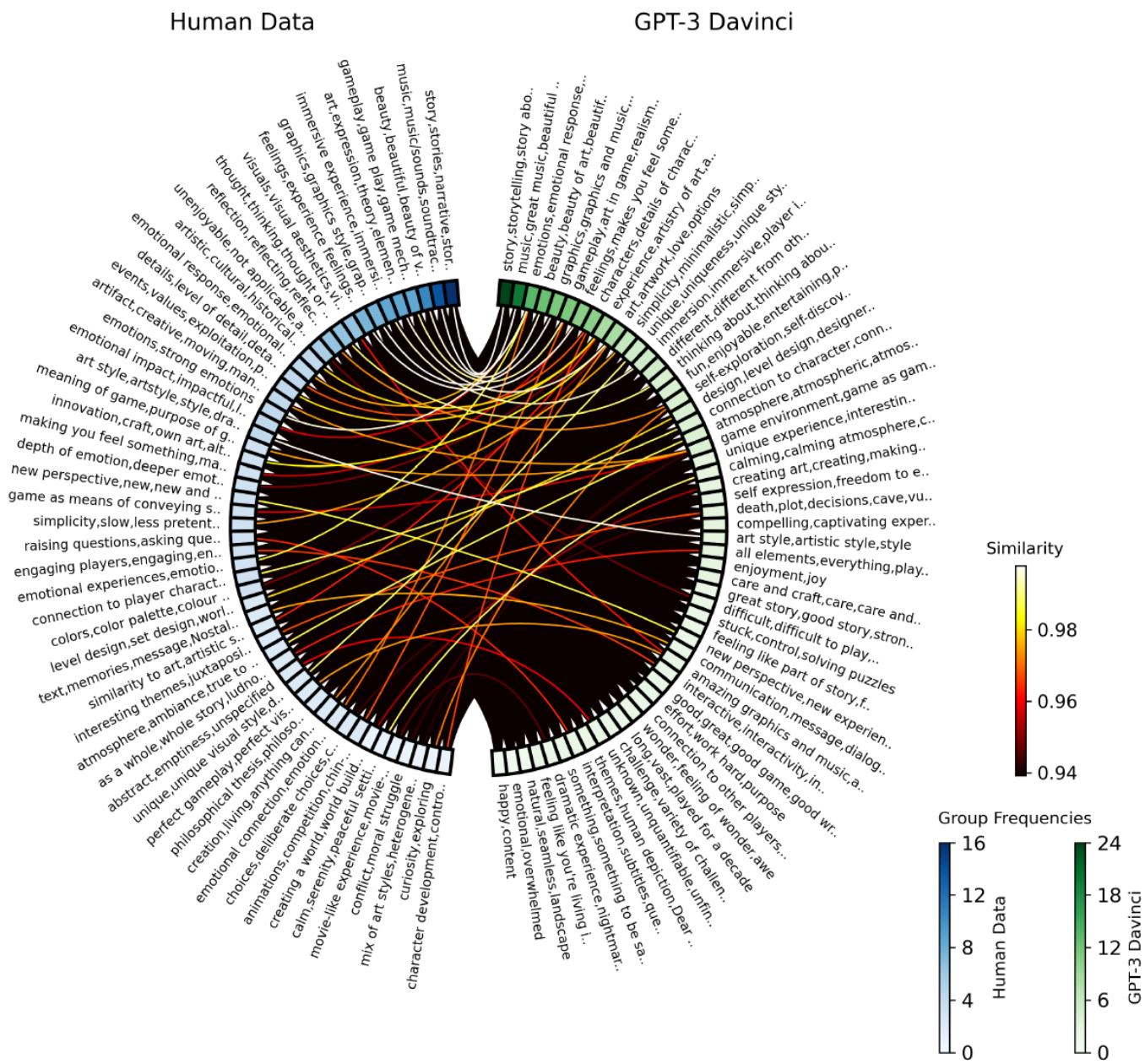


Figure 7: A circular graph presenting the human and GPT-3 davinci data resulting from the automatic qualitative coding. Each human data group is connected with a line to the most similar group in the GPT-3 data. The lines are color coded based on cosine similarity. The color coding and sorting of the group nodes is based in how frequent the groups were in the two datasets (groups with highest frequency on top). Here, group frequencies are reported as percentages.

Rank	Human data		GPT-3 davinci		GPT-3 text-davinci-002	
1.	*The Legend of Zelda: BOTW*	**10**	*Journey*	**44**	*Journey*	**151**
2.	*Journey*	**7**	*The Last of Us*	**12**	*Flower*	**5**
3.	*Nier: Automata*	**7**	*Dear Esther*	**8**	*That Dragon, Cancer*	**3**
4.	*Red Dead Redemption 2*	**6**	*Portal*	**7**	*Braid*	**2**
5.	*The Last of Us Part II*	**6**	*Bioshock*	**6**	*Shadow of the Colossus*	**2**
6.	*Firewatch*	**5**	*Shadow of the Colossus*	**5**	*Dreams of Geisha*	**1**
7.	*Hollow Knight*	**5**	*The Path*	**5**	*Final Fantasy VII*	**1**
8.	*Disco Elysium*	**4**	*Limbo*	**3**	*Flow*	**1**
9.	*Life Is Strange*	**4**	*Mirror's Edge*	**3**	*Frog Fractions*	**1**
10.	*Bioshock*	**3**	*The Stanley Parable*	**3**	*Halo 5: Guardians*	**1**
11.	*Shadow of the Colossus*	**3**	*Final Fantasy IX*	**2**	*Kingdom Hearts*	**1**
12.	*The Witcher 3*	**3**	*Final Fantasy VII*	**2**	*The Legend of Zelda: BOTW*	**1**
13.	*Undertale*	**3**	*Flower*	**2**	*Nier: Automata*	**1**
14. –>	*... and 97 other games*	**113**	*... and 65 other games*	**69**	*... and 10 other games*	**10**

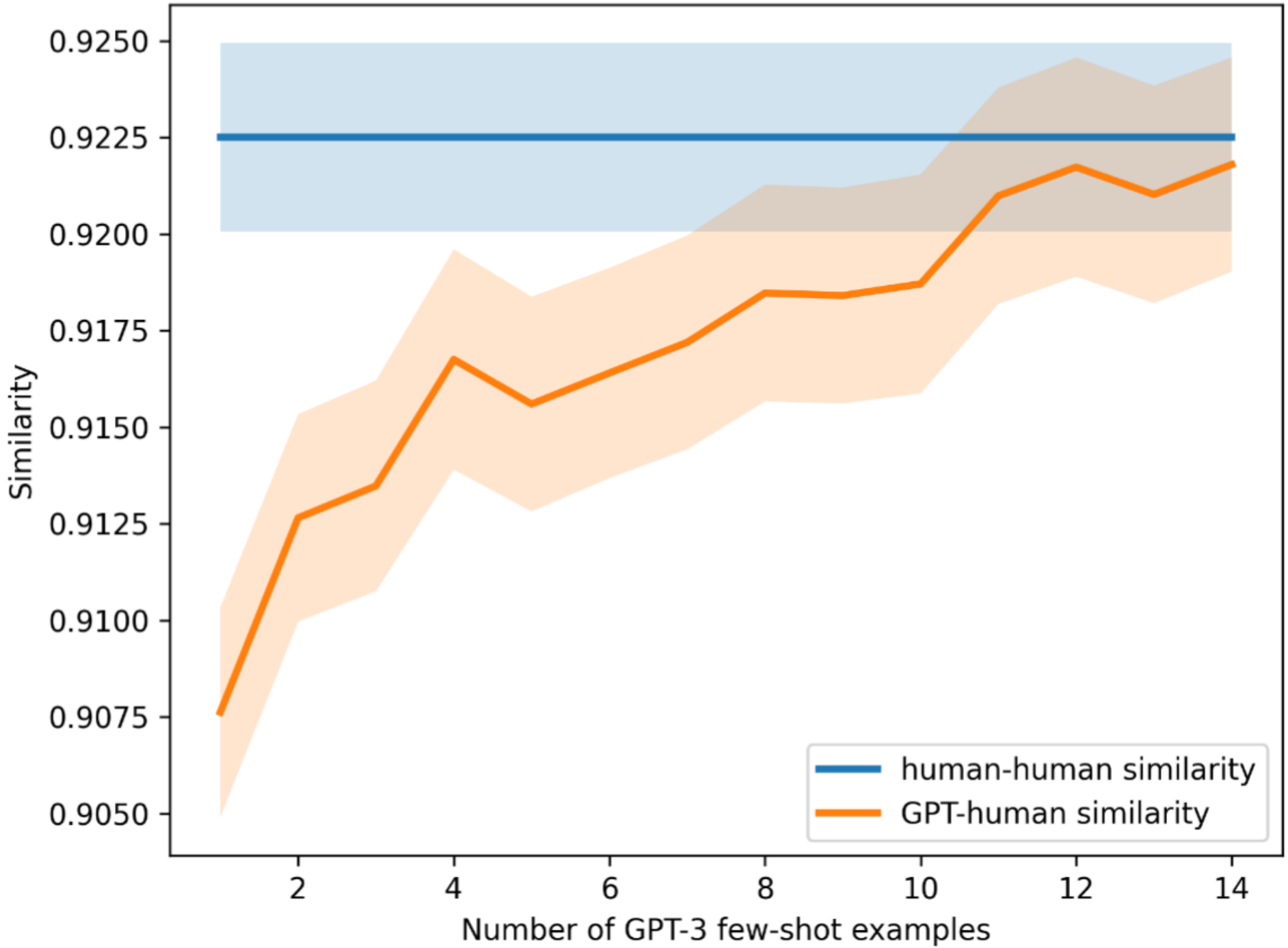
AI-assisted data analysis

AI-based speech recognition

- Very useful for transcribing interviews
- Word 360: Aalto account gives you some free transcription hours per month.
- OpenAI Whisper: a recent very good transcription model
  - Many options for using: OpenAI API, Colab notebooks, and you can also download and run the model on your own GPU.
  - WhisperX = Whisper + 3rd party speaker recognition (e.g., researcher & participant): https://github.com/m-bain/whisperX
- Word 360 is better at recognizing and labeling multiple speakers, but WhisperX makes less transcription errors.
- Whatever the AI tool, make sure you have as good audio quality as possible. Laptop microphone < phone < a good microphone

Table 1. An example of an prompt that frames coding as a text continuation task for GPT-3. The prompt comprises a short instruction text and manually coded few-shot examples separated by "###". Note that the prompt does not limit the model to the codes given as examples, and according to our experiments, it can generalize based on the examples and generate new codes in the same style when needed.

The following presents a qualitative coding of answers from a video game research study. The answers explain why a participant experienced a game as art. The codes summarize the given reasons as compactly as possible. If an answer lists multiple reasons, the corresponding codes are separated by semicolons.
###
Answer: The questions it raised and the highly emotional connection that emerged between me and the game, the experience.
Codes: raising questions; emotional connection
###
Answer: For a game experience to feel like a work of art to me, it would usually be an immersive experience that creates a real emotional response. Since games accomplish this through a combination of illustration, animation, sound, music, storytelling elements all together, I would consider these types of experiences art.
Codes: immersive experience; emotional response
###
Answer: The fact that each asset was hand drawn in such a unique style.
Codes: unique visual style
###
Answer: *<each new answer to code inserted here>*
Codes:


0.9250
0.9225
0.9200
0.9175
0.9150
0.9125
0.9100
0.9075
0.9050
Similarity
2
4
6
8
10
12
14
Number of GPT-3 few-shot examples
human-human similarity
GPT-human similarity

Topic modeling

- Natural Language Processing (NLP) problem related to thematic analysis: identify frequent topics in some mass of text
- Run the BERTopic algorithm in a Colab notebook: https://colab.research.google.com/drive/1FieRA9fLdkQEGDIMYl0I3MCjSUKVF8C-?usp=sharing

Neural Network Based Facial Expression Analysis of Game Events: A Cautionary Tale

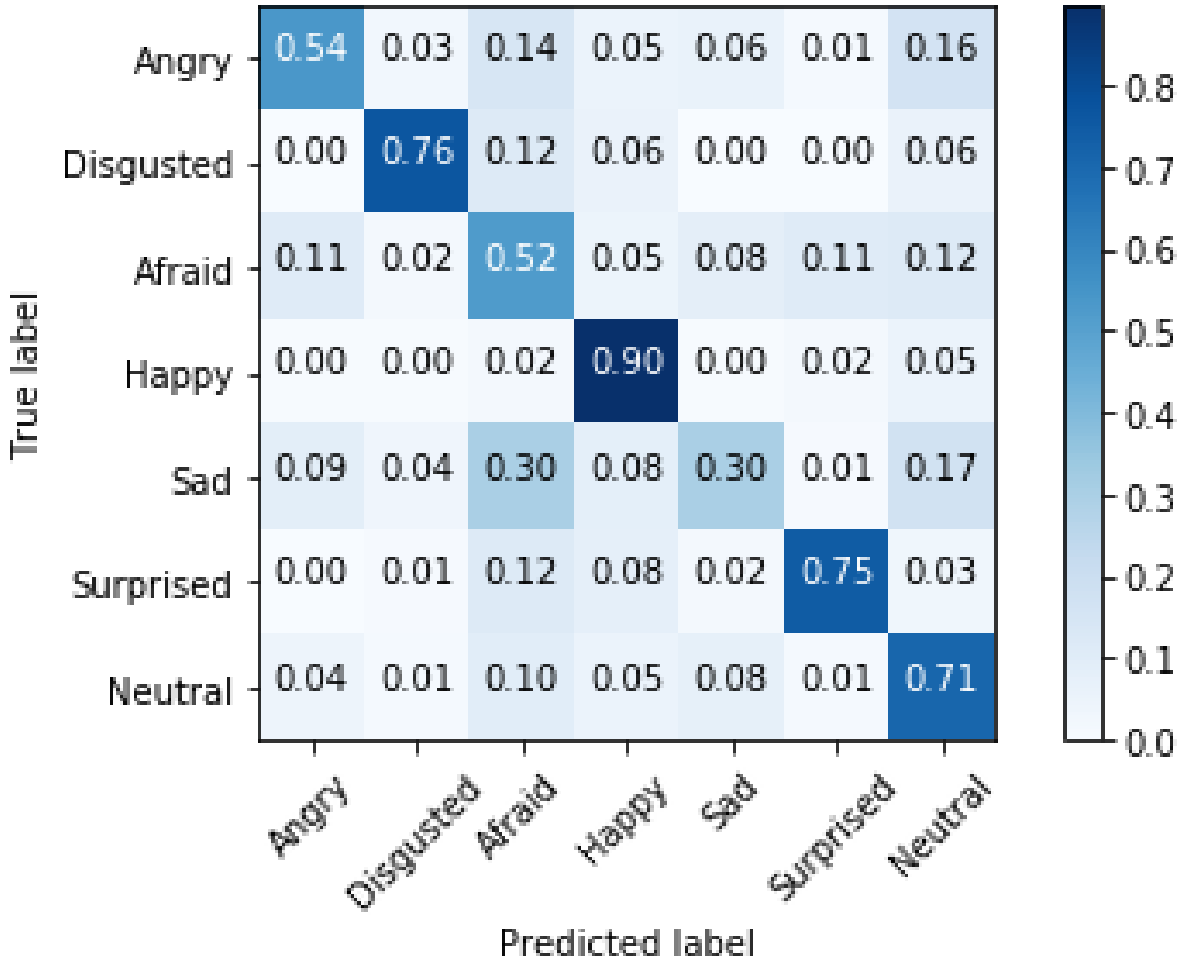
Shaghayegh Roohi, Jari Takatalo, J. Matias Kivikangas, Perttu Hämäläinen

Why focus on events?

- Every player may take a different path through a game
- Time-series signals of emotions cannot be directly compared or aggregated
- However, it’s easy to log out key events with time stamps: player deatch, player getting a reward etc.
- This paper: measure the affect gradient of events: Average change in emotional facial expression around specific events.

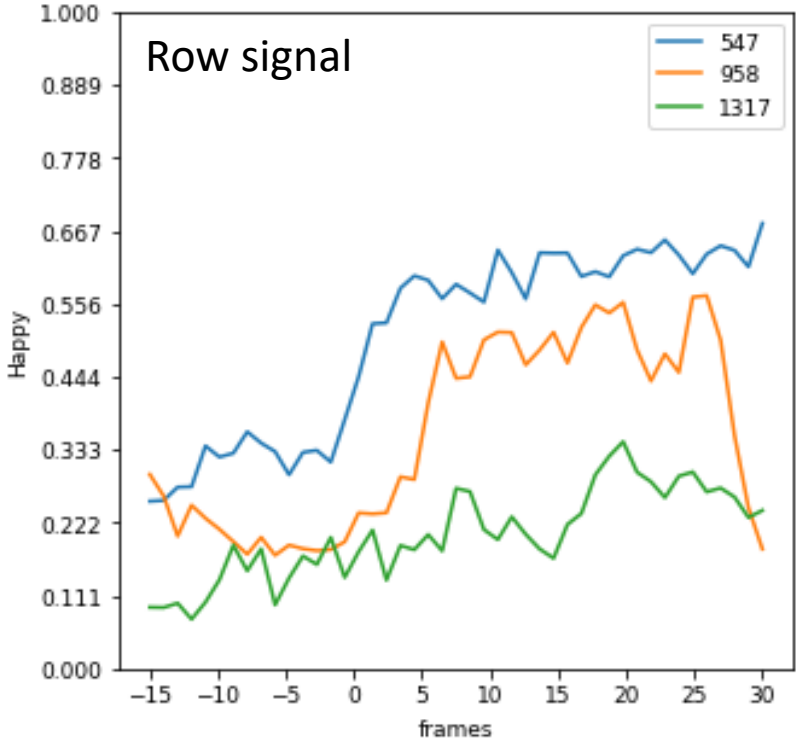
Approach

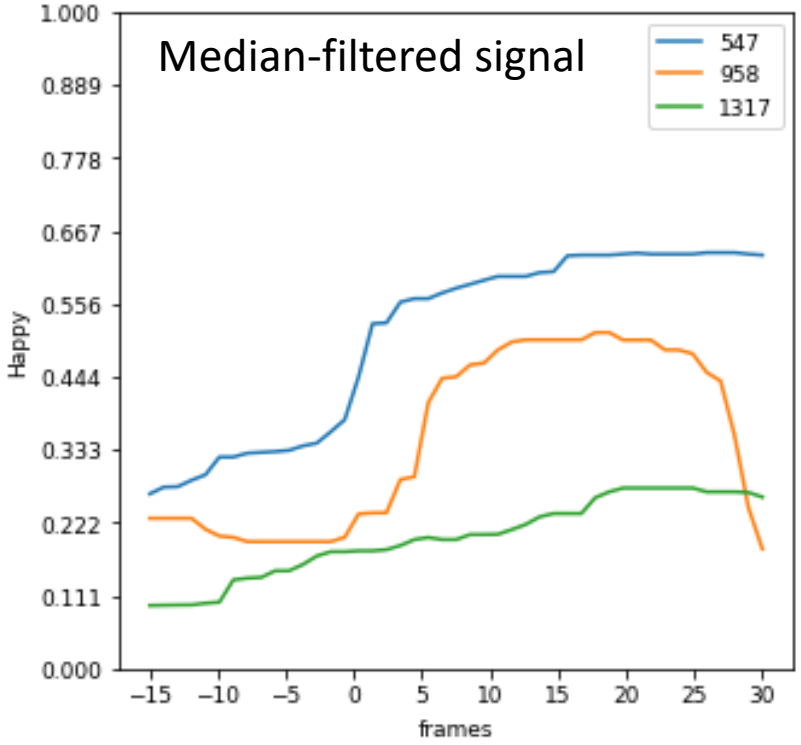
- Computer vision using facial video of the player
- Convolutional neural networks trained with free 7-emotion Kaggle dataset
- Focus on ”Happy”, because it’s most robust
- Analyze response to game events like dying or killing enemy



Affect gradient

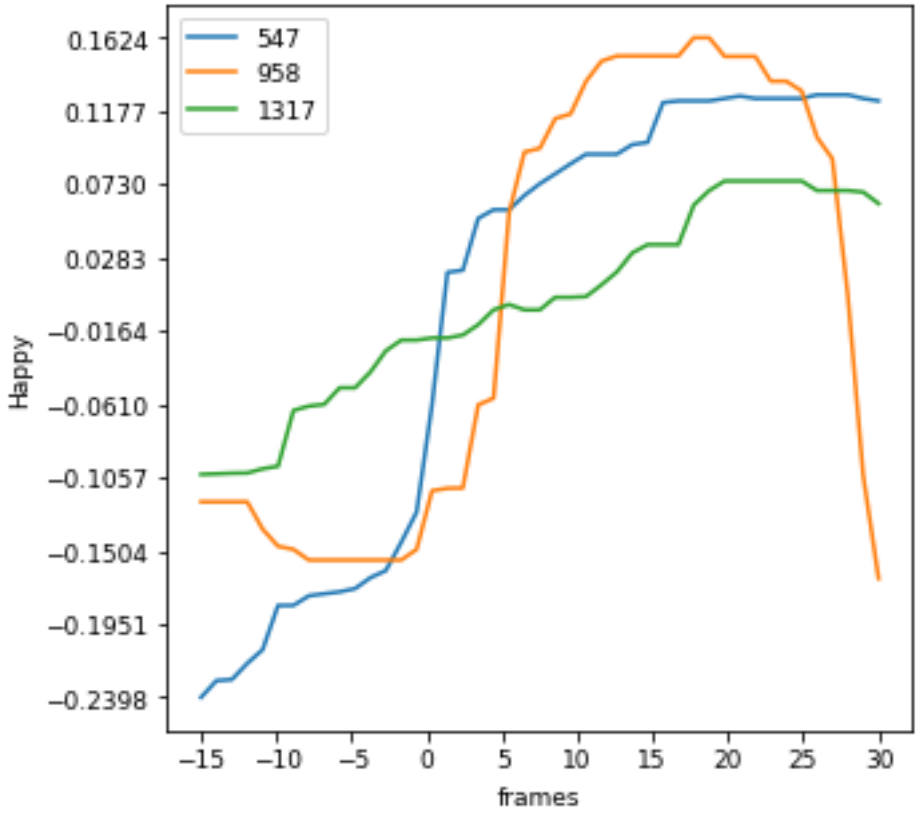
- Preprocess the data with median filtering
- Extract segments of fixed length around each logged game event





Affect gradient

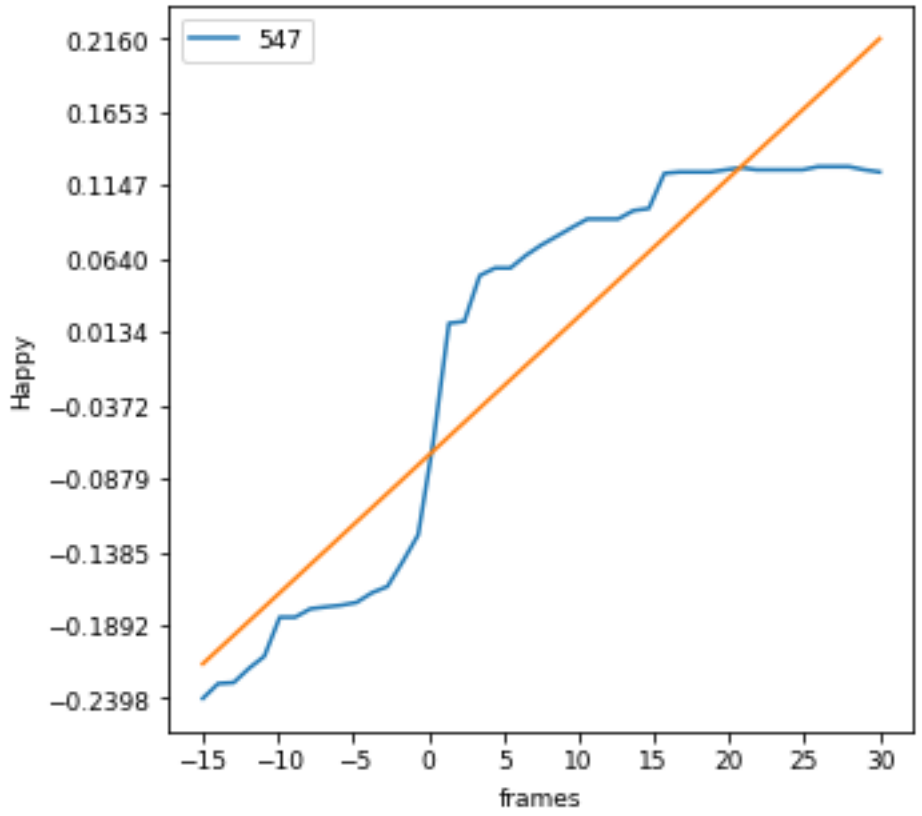
- Normalize the segments to have zero mean



Normalized signal

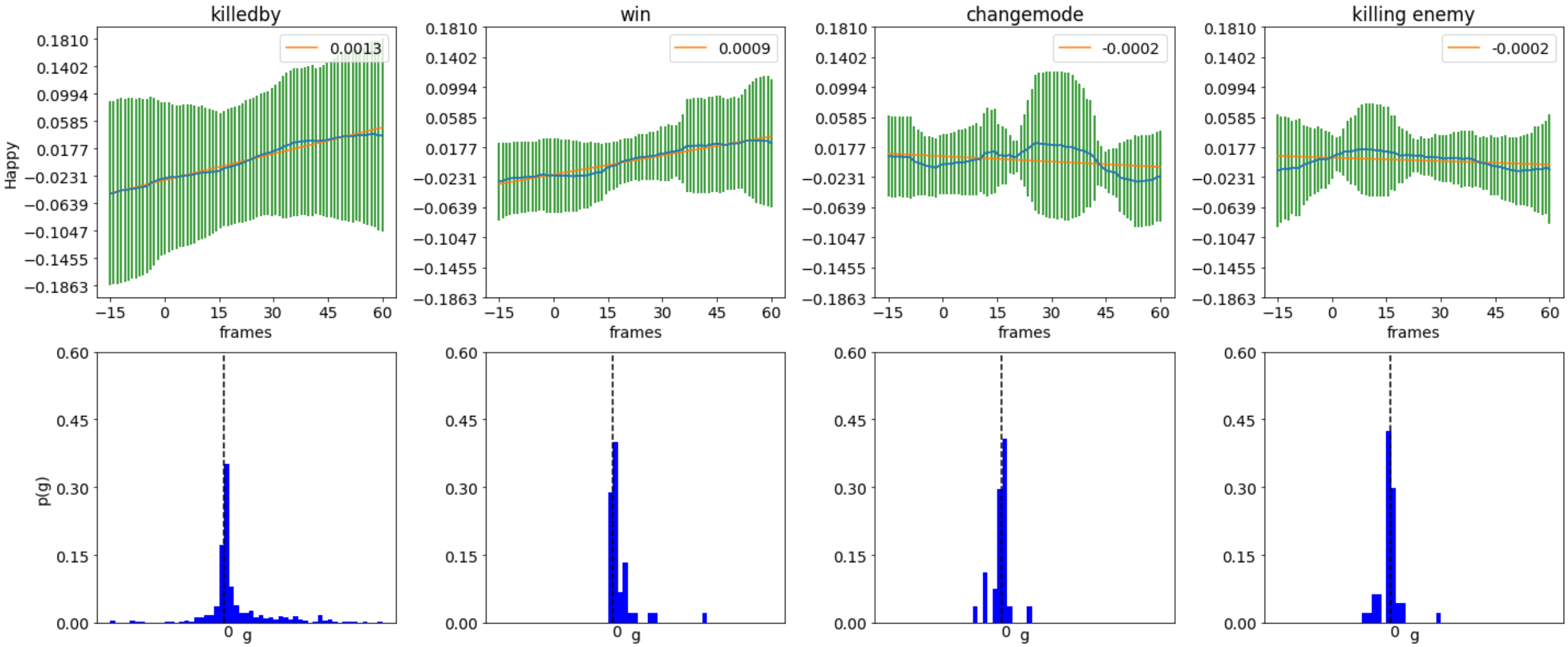
Affect gradient

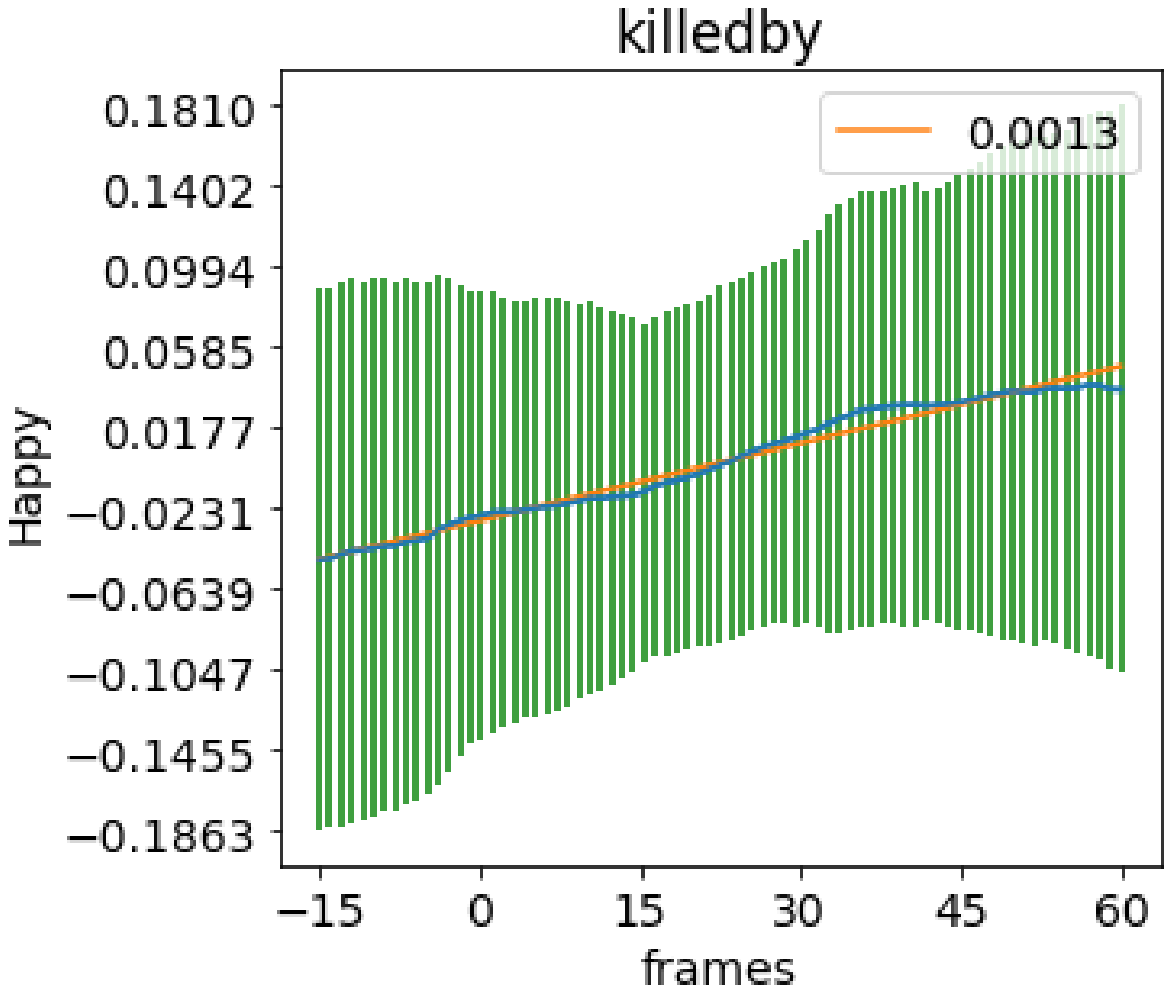
- Fit lines to the extracted segments
  - The slopes give the affect gradients of the events

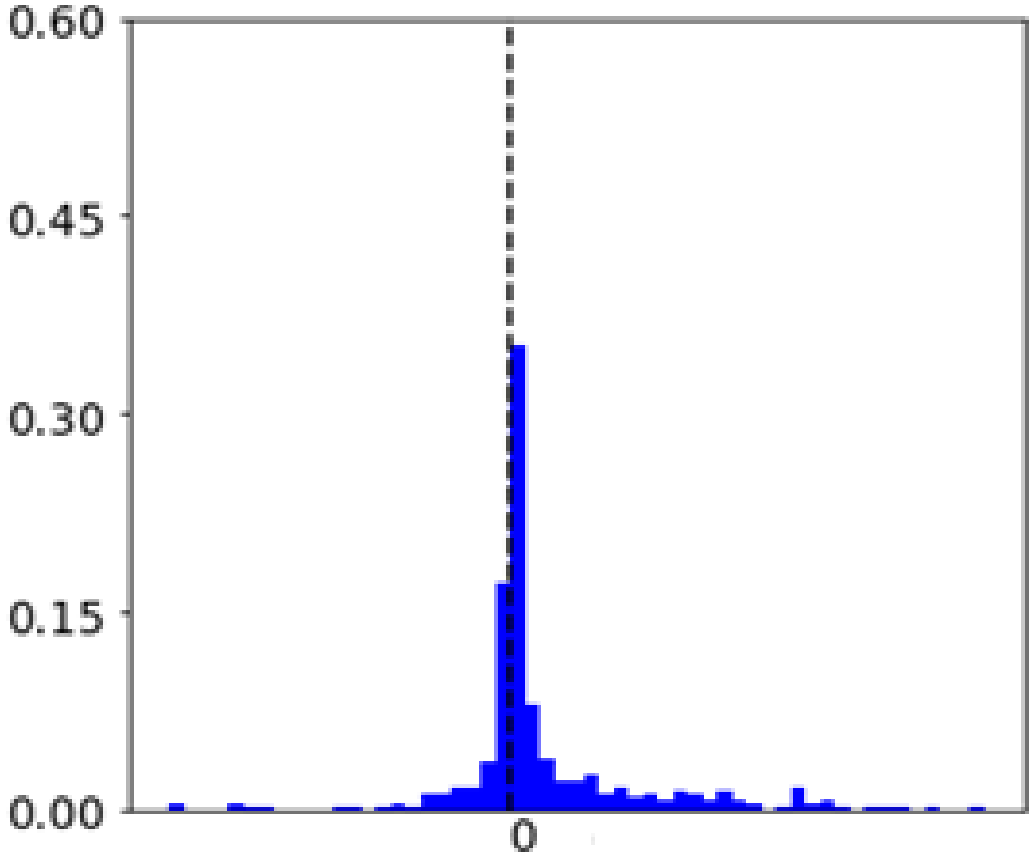


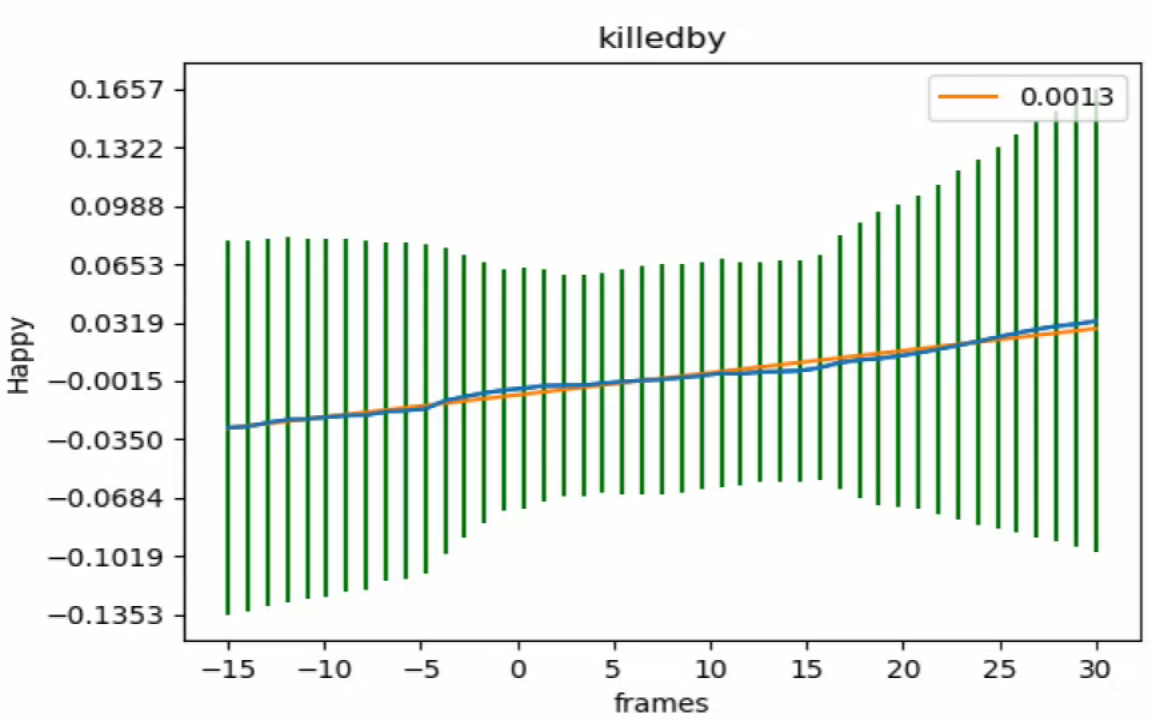
Affect gradient

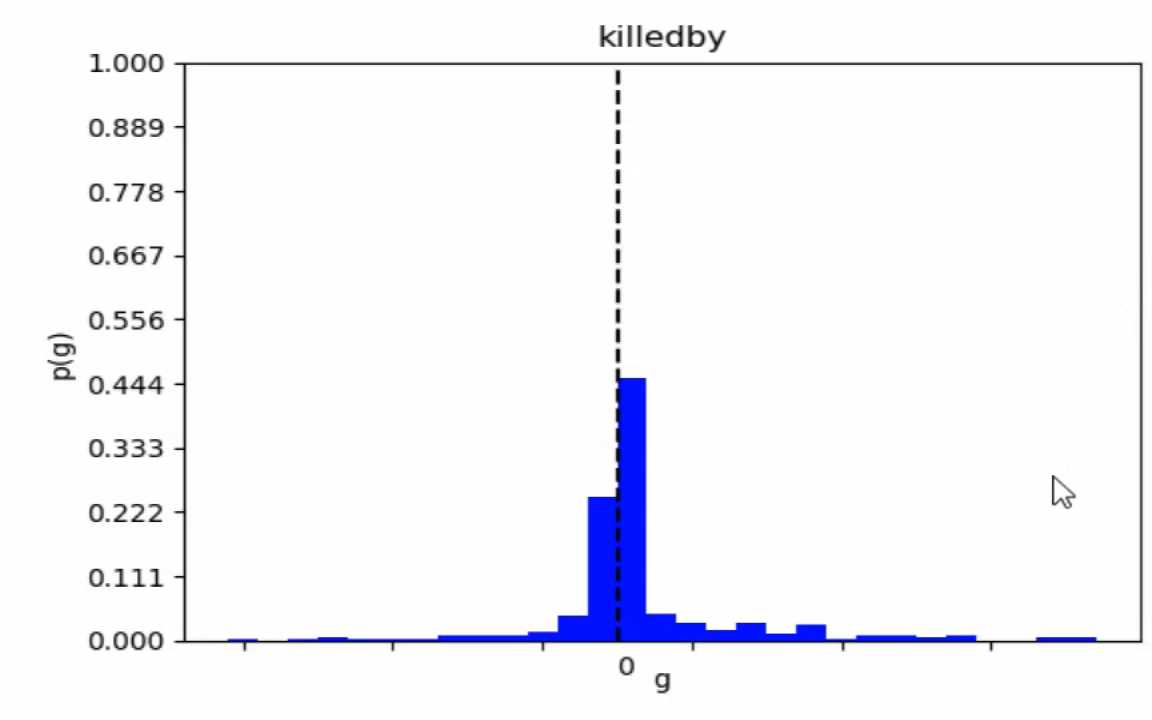
Affect gradient summary plots




killedby
0.0013
0.1810
0.1402
0.0994
0.0585
0.0177
−0.0231
−0.0639
−0.1047
−0.1455
−0.1863
Happy
−15
0
15
30
45
60
frames


0.60
0.45
0.30
0.15
0.00
0


killedby
0.0013
0.1657
0.1322
0.0988
0.0653
0.0319
−0.0015
−0.0350
−0.0684
−0.1019
−0.1353
Happy
−15
−10
−5
0
5
10
15
20
25
30
frames

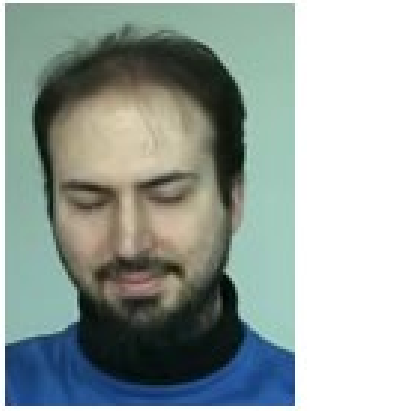

killedby
1.000
0.889
0.778
0.667
0.556
0.444
0.333
0.222
0.111
0.000
p(g)
0
g

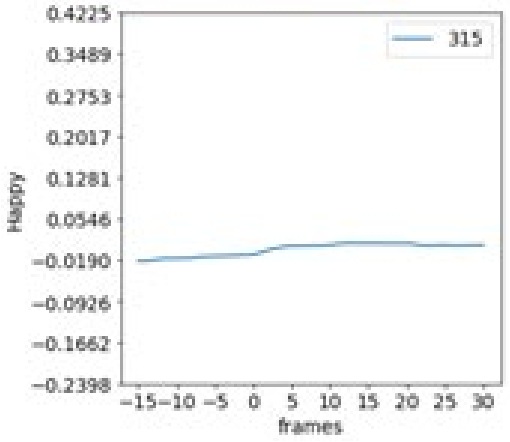
Getting Killed Makes Players Happy, According to a Neural Network

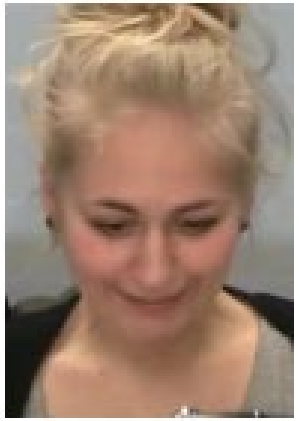
Shaghayegh Roohi, Jari Takatalo, J. Matias Kivikangas, Perttu Hämäläinen

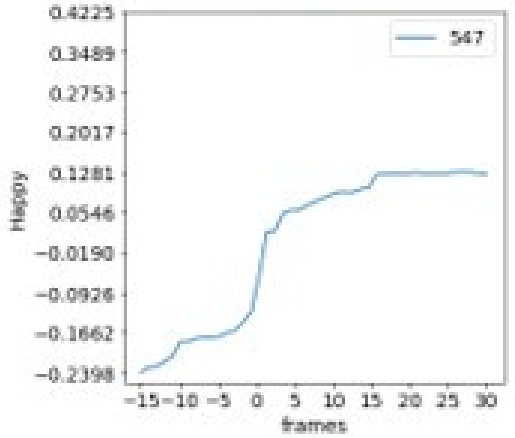
Results replicate previous psychophysiological studies

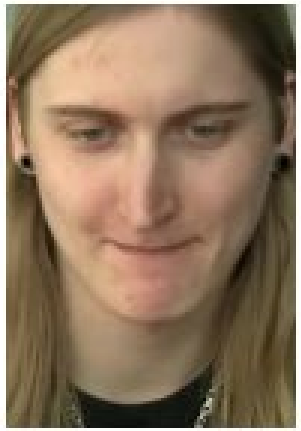
- Getting killed produces a smile

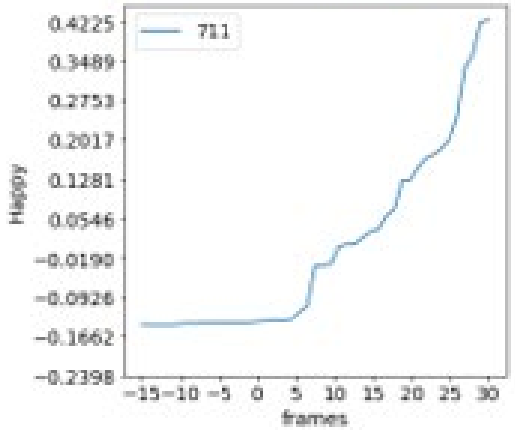




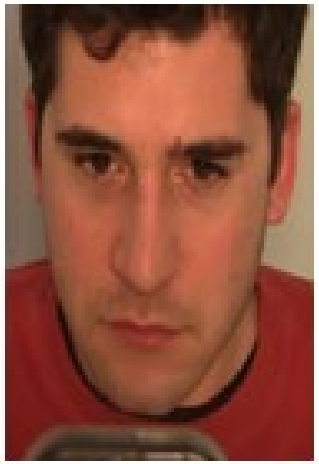


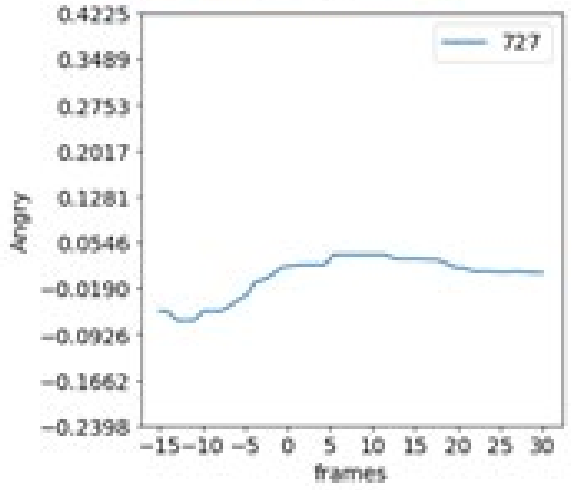


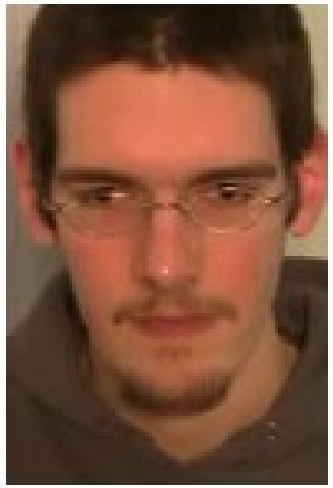


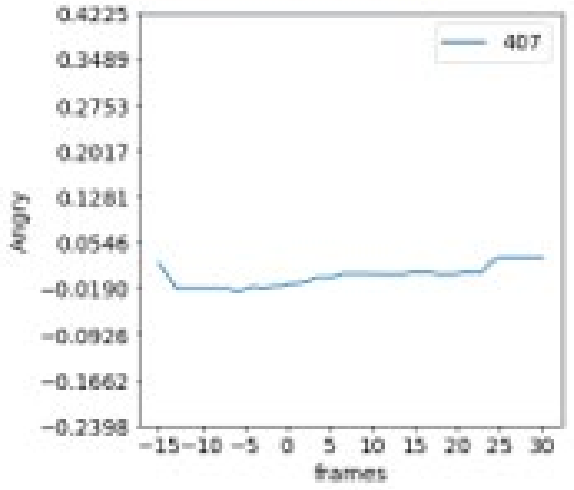


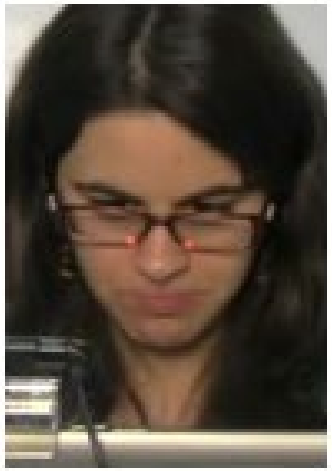
Killing enemies: concentrated frown interpreted as neutral or negative emotion

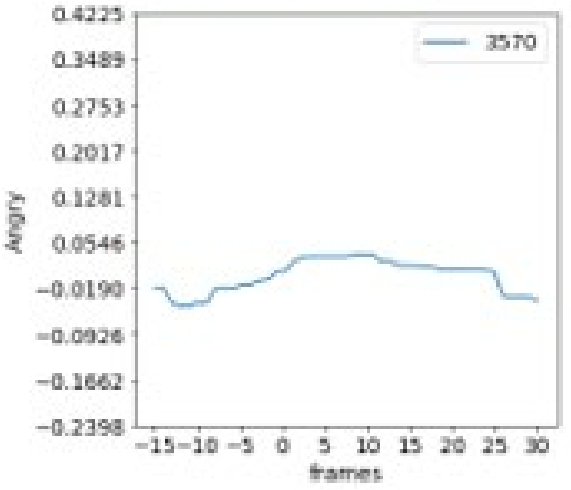












A Good Reason to Die: How Avatar Death and High Challenges Enable Positive Experiences

Serge Petralito[1], Florian Brühlmann[1], Glena Iten[1], Elisa D. Mekler[2] and Klaus Opwis[1]

[1]Center for Cognitive Psychology and Methodology, Department of Psychology, University of Basel
[2]HCI Games Group, Games Institute, University of Waterloo
{s.petralito, florian.bruehlmann, glena.iten, klaus.opwis}@unibas.ch, emekler@uwaterloo.ca

ABSTRACT

Appropriate challenges and challenge-skill balance are usually key to positive player experiences. However, some games such as the successful series *Dark Souls* are notorious for their excessive difficulty. Yet, there has been little empirical investigation of why players enjoy games they constantly struggle and fail with. We surveyed 95 participants right after the release of *Dark Souls III* about their experiences with the game, employing both open questions and different player experience measures. Players generally enjoyed challenging play sessions and mostly reported positive experiences, with achievement and learning moments strongly contributing to positive experiences. However, these factors themselves were enabled by negative events such as difficulties and avatar death. Our findings showcase that negative events bear a potential for forming positive and meaningful experiences, thus expanding previous knowledge about the role of challenge and failing in games. Moreover, the significance of hard-earned achievements extends present design conventions.

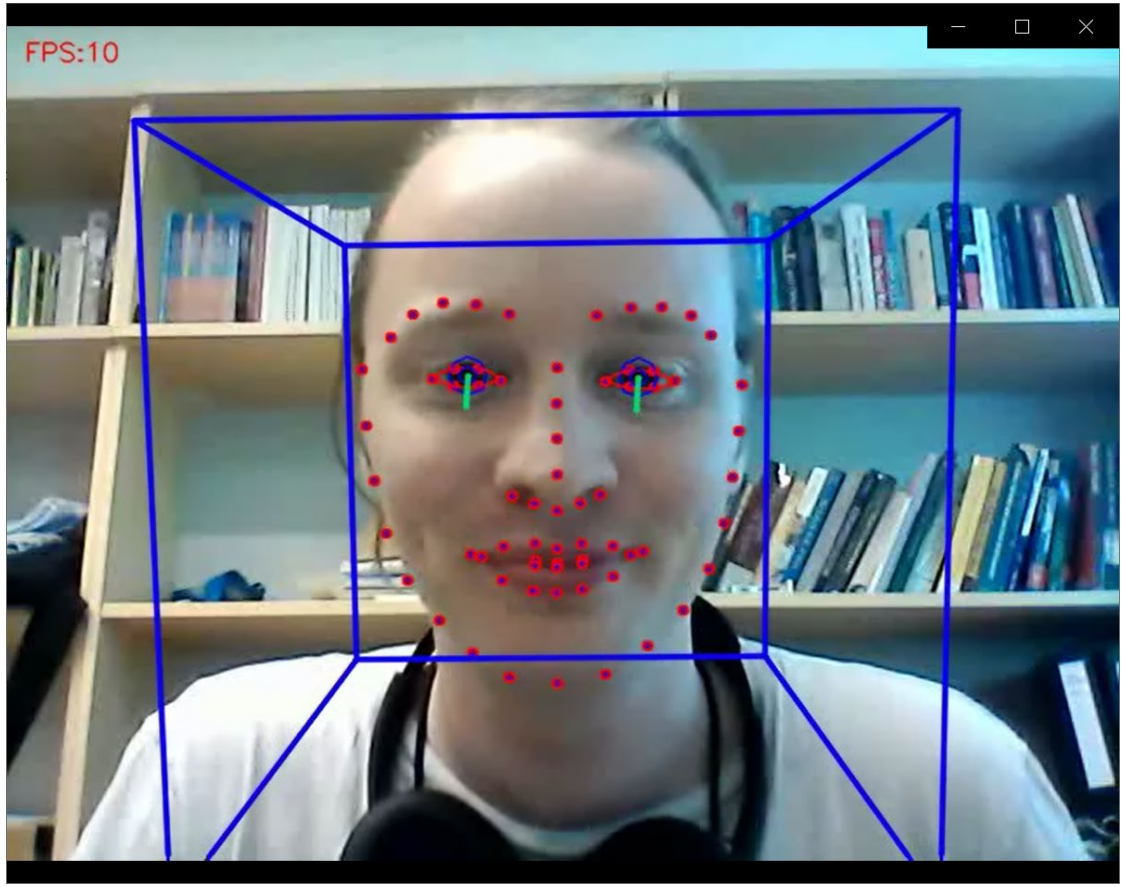
balance between challenge and skill. Hence, if challenge demands imposed by the game are too high or too low in regard to the player's skill level, playing the game leads to anxiety or boredom. The significance of an ideal challenge-skill balance is strongly emphasized in current research [3, 4, 13, 23, 33, 37], where adjustable and adaptive difficulty mechanics play an integral part in keeping this balance [8, 14, 35, 39]. Moreover, balance and accessibility represent two key notions of the *casual revolution*, a design trend towards making games more accessible by removing perceived barriers, penalties and frustrations and targeting much broader audiences than games used to over roughly a decade ago [20, 22]. In conclusion, challenge in current literature and modern game design has to a large extent been treated as a *Goldilocks factor*: The difficulty of a game should be neither too demanding nor too low in order to avoid negative experiences and frustrations.

In light of present design conventions, some exceptional games stand out, ignoring most of the conventional balancing efforts by implementing very high challenges and high consequential

OpenFace: a facial behavior analysis toolkit

https://github.com/TadasBaltrusaitis/OpenFace/wiki

- Openface is an easy to use opensource toolkit that detects facial landmarks, head pose, eye-gaze direction and facial action units



Action unit estimation

- Openface can estimate facial movements based on Facial Action Coding System (FACS)
- In FACS, facial movements are coded as different action units (AU): https://imotions.com/blog/facial-action-coding-system/
- Openface estimates the presence and the intensity (scale from 0 to 5) of different AUs in each frame of the video
- For the accuracy of AU estimates, see Baltrusaitis, Zadeh, Lim, & Morency (2018)

How to use, step by step (Windows Powershell)

- Openface is operated trough command line interface: e.g. with PowerShell (Windows) or xterm (Unix)
- 1. Install: https://github.com/TadasBaltrusaitis/OpenFace/wiki/Windows-Installation
- 2. Open Windows Powershell
- 3. Change the directory to the openface folder, for example:
  - cd C:
  - cd \...\...\...\OpenFace_2.0.5_win_x64\

Useful command line arguments

- 4: Execute a command, examples:
  - Extract features (CSV-file): .\FeatureExtraction.exe -f "C:\...\...\filename.avi”
  - Extract AU estimates (CSV-file): \FeatureExtraction.exe -aus –f "C:\...\...\...\filename.avi”
  - Visualize the data: .\FeatureExtraction.exe –verbose "C:\...\...\filename.avi”
- 5. Analyze the data ☺
- Also possible to use GUI with the argument ./OpenFaceOffline.exe
- List of all possible arguments: https://github.com/TadasBaltrusaitis/OpenFace/wiki/Command-line-arguments

Administrator: Windows PowerShell

```
PS C:\Openface\OpenFace_2.0.5_win_x64> .\FeatureExtraction.exe -f "C:\my videos\video.avi" -verbose
Reading the landmark detector/tracker from: model/main_ceclm_general.txt
Reading the landmark detector module from: model\cen_general.txt
Reading the PDM module from: model\pdms/In-the-wild_aligned_PDM_68.txt....Done
Reading the Triangulations module from: model\tris_68.txt....Done
Reading the intensity CEN patch experts from: model\patch_experts/cen_patches_0.25_of.dat....Done
Reading the intensity CEN patch experts from: model\patch_experts/cen_patches_0.35_of.dat....Done
Reading the intensity CEN patch experts from: model\patch_experts/cen_patches_0.50_of.dat....Done
Reading the intensity CEN patch experts from: model\patch_experts/cen_patches_1.00_of.dat....Done
Reading part based module....left_eye_28
Reading the landmark detector/tracker from: model\model_eye/main_clnf_synth_left.txt
Reading the landmark detector module from: model\model_eye\clnf_left_synth.txt
Reading the PDM module from: model\model_eye\pdms/pdm_28_l_eye_3D_closed.txt....Done
Reading the intensity CCNF patch experts from: model\model_eye\patch_experts/left_ccnf_patches_1.00_synth_lid_.txt....Done
Reading the intensity CCNF patch experts from: model\model_eye\patch_experts/left_ccnf_patches_1.50_synth_lid_.txt....Done
Done
Reading part based module....right_eye_28
Reading the landmark detector/tracker from: model\model_eye/main_clnf_synth_right.txt
Reading the landmark detector module from: model\model_eye\clnf_right_synth.txt
Reading the PDM module from: model\model_eye\pdms/pdm_28_eye_3D_closed.txt....Done
Reading the intensity CCNF patch experts from: model\model_eye\patch_experts/ccnf_patches_1.00_synth_lid_.txt....Done
Reading the intensity CCNF patch experts from: model\model_eye\patch_experts/ccnf_patches_1.50_synth_lid_.txt....Done
Done
Reading the landmark validation module....Done
Reading the AU analysis module from: AU_predictors/main_dynamic_svms.txt
Reading the AU predictors from: AU_predictors\AU_all_best.txt... Done
Reading the PDM from: AU_predictors\In-the-wild_aligned_PDM_68.txt... Done
Reading the triangulation from:AU_predictors\tris_68_full.txt... Done
Attempting to read from file: C:\my videos\video.avi
Device or file opened
Starting tracking
Reading the MTCNN face detector from: model/mtcnn_detector/MTCNN_detector.txt
Reading the PNet module from: model/mtcnn_detector\PNet.dat
Reading the RNet module from: model/mtcnn_detector\RNet.dat
Reading the ONet module from: model/mtcnn_detector\ONet.dat
0% 10% 20% 30% 40% 50% 60% 70% Closing output recorder
Closing input reader
Closed successfully
Postprocessing the Action Unit predictions
PS C:\Openface\OpenFace_2.0.5_win_x64> .\FeatureExtraction.exe -f "C:\my videos\video.avi" -verbose
```

Add-in Express

GPowerWin_3.1.9.2

18.42
11.6.2019

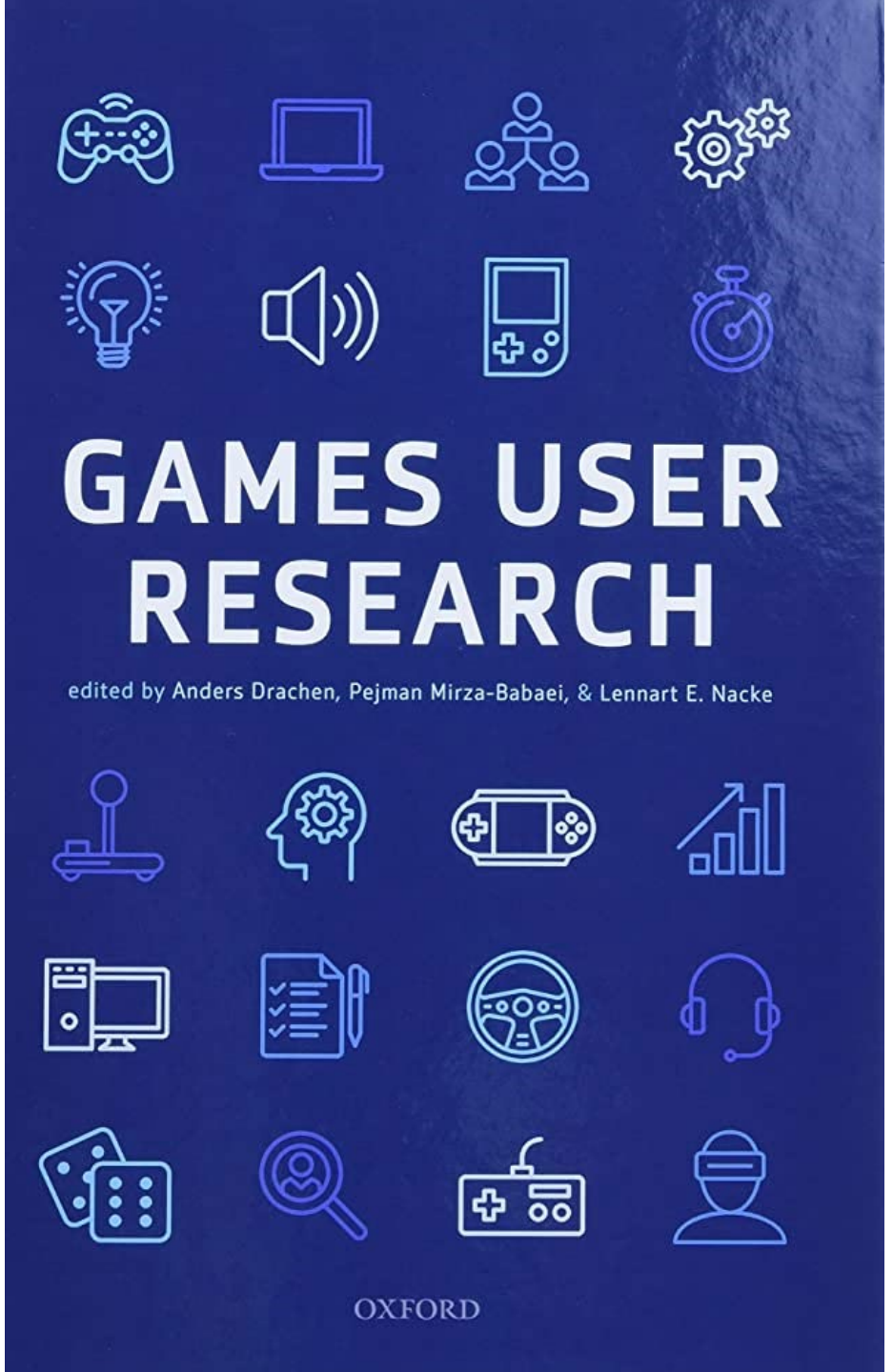
References

- Baltrusaitis, T., Zadeh, A., Lim, Y. C., & Morency, L. (2018). OpenFace 2.0: Facial Behavior Analysis Toolkit. *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, 59–66. https://doi.org/10.1109/FG.2018.00019

Resources

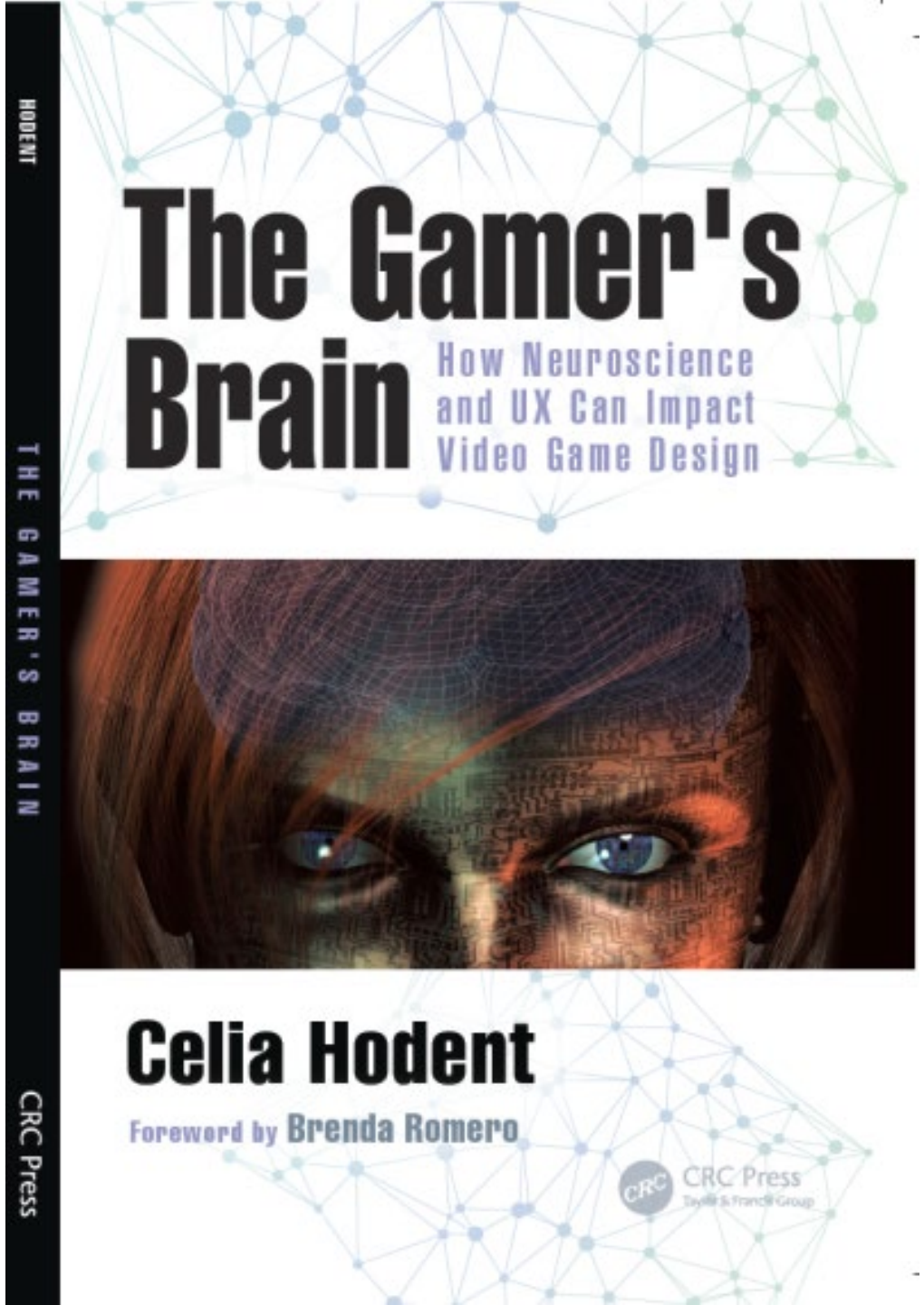
- GUR challenges and best practices in the industry
- Post-launch GUR
- Designing a GUR lab
- Biometric measures
- Game analytics
- GUR case studies of published games

…and more



Resources

- What is player experience?
- Can give ideas on what to measure
- Similar topics also covered on the Game Analysis course



Playtest Masterclass | Research Skills

How to run unbiased player interviews

Every stage of preparing, running and analysing your playtest interviews. Plus tips on how to reveal player's honest thoughts, and apply that to game design decisions.

Playtest Masterclass

Run better internal playtests

Run tighter, more useful internal playtests and make the most of your colleague's time.

Playtest Masterclass | Research Skills

Top remote playtest platforms for unmoderated testing

Pick the right online playtest platform for your needs

https://gamesuserresearch.com/articles/

Exercise: Plan your GUR experiment

- Time: 1h
- Use any method and data, e.g, your test your game or analyze discussions in r/games
- Describe your plan using 1-2 slides, answering:
  - What are your main research questions?
  - What data will you collect and how?
  - How will you analyze the data?
  - Who will utilize the results and how?

Add your slides to a shared Google Slides

Tuesday & Wednesday

- Tuesday: prepare your GUR experiment as planned
- Wednesday: run the experiment. Recruit others than students on this class as participants if you can, but be prepared to participate yourself.
  - 9-12: Group A tests, Group B as participants if needed
  - 13-16: Group B tests, Group A as participants if needed

Group A: Atte, Apogrowth, Moosio, Tom & Yan, VR heat

Group B: Puzzle, Marmortal, Tomorrow island, Spell Crafting, Patrols, Jarno

Thursday & Friday

- Thursday: AI-based game testing (Christian teaching)
- Friday 9.15-12: your GUR data analysis (Perttu will be there to help with AI-based analyses, Python etc)
- Friday 13-14: Two invited case study presentations
- Friday 14.15-16: presentations and discussion (5 min presentation + 5 min discussion per group)