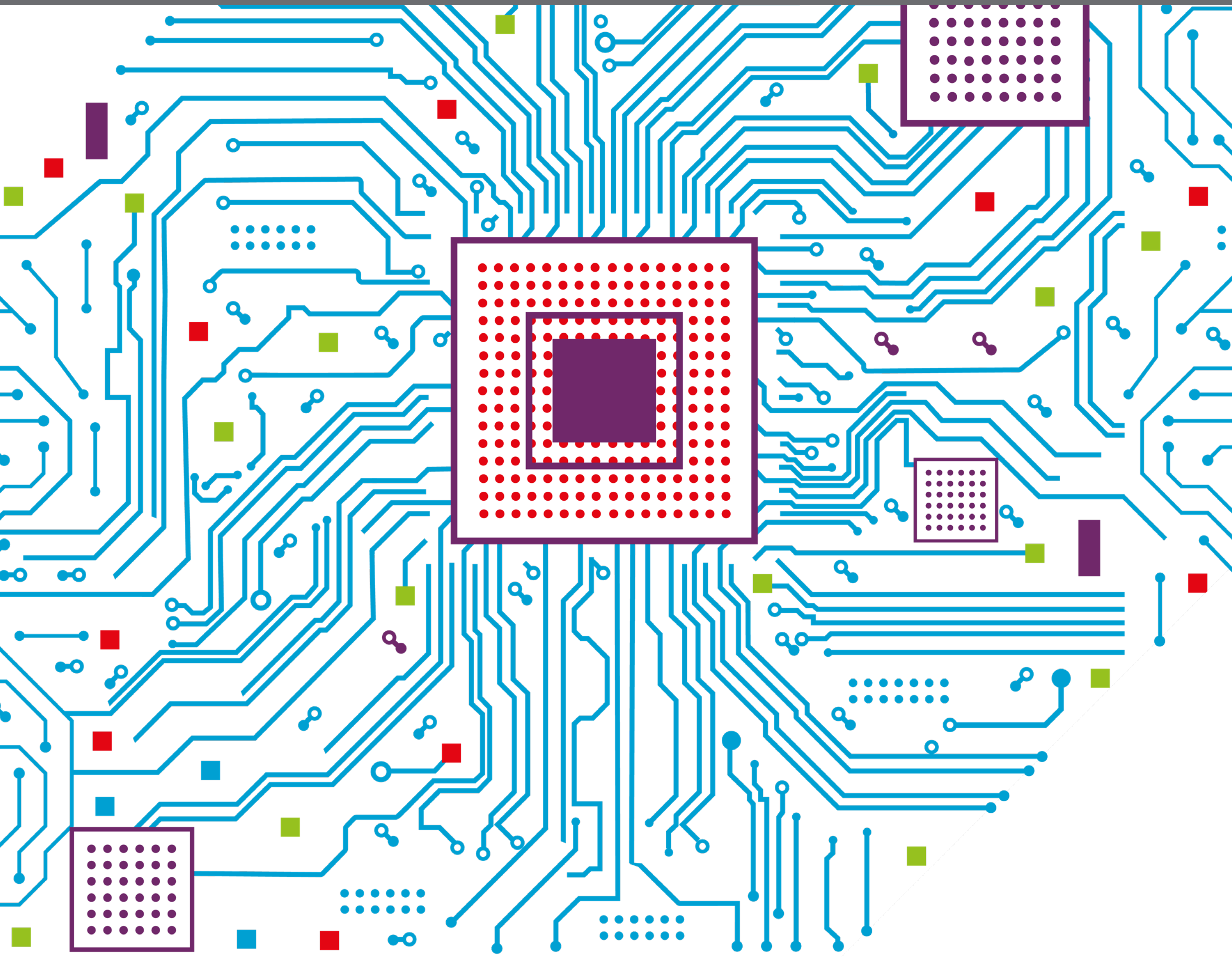
GAMES AND PLAY IN HCI

EDITED BY: Kathrin Gerling, Ioanna Iacovides, Marc Herrlich and Z. O. Toups
PUBLISHED IN: Frontiers in Computer Science and Frontiers in Psychology



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88971-144-4
DOI 10.3389/978-2-88971-144-4

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

GAMES AND PLAY IN HCI

Topic Editors:
Kathrin Gerling, University of Lincoln, United Kingdom
Ioanna Iacovides, University of York, United Kingdom
Marc Herrlich, University of Kaiserslautern, Germany
Z. O. Toups, New Mexico State University, United States

Citation: Gerling, K., Iacovides, I., Herrlich, M., Toups, Z. O., eds. (2021). Games and Play in HCI. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-144-4

Table of Contents

04 ***Guidelines for the Development of Immersive Virtual Reality Software for Cognitive Neuroscience and Neuropsychology: The Development of Virtual Reality Everyday Assessment Lab (VR-EAL), a Neuropsychological Test Battery in Immersive Virtual Reality***
Panagiotis Kourtesis, Danai Korre, Simona Collina, Leonidas A. A. Doumas and Sarah E. MacPherson

28 ***External Assistance Techniques That Target Core Game Tasks for Balancing Game Difficulty***
Jawad Jandali Refai, Scott Bateman and Michael W. Fleming

44 ***What Is It Like to Be a Game?—Object Oriented Inquiry for Games Research, Design, and Evaluation***
Katta Spiel and Lennart E. Nacke

58 ***Moderators of Social Facilitation Effect in Virtual Reality: Co-presence and Realism of Virtual Agents***
Paweł M. Strojny, Natalia Dużmańska-Misiarczyk, Natalia Lipp and Agnieszka Strojny

70 ***Motivational Profiling of League of Legends Players***
Florian Brühlmann, Philipp Baumgartner, Günter Wallner, Simone Kriglstein and Elisa D. Mekler

88 ***Validation of User Preferences and Effects of Personalized Gamification on Task Performance***
Gustavo F. Tondello and Lennart E. Nacke

111 ***A Taxonomy of Coping Strategies and Discriminatory Stressors in Digital Gaming***
Cale J. Passmore and Regan L. Mandryk

131 ***Technology Facilitates Physical Activity Through Gamification: A Thematic Analysis of an 8-Week Study***
Dennis L. Kappen, Pejman Mirza-Babaei and Lennart E. Nacke

151 ***"HIIT" the ExerCube: Comparing the Effectiveness of Functional High-Intensity Interval Training in Conventional vs. Exergame-Based Training***
Anna Lisa Martin-Niedecken, Andrea Mahrer, Katja Rogers, Eling D. de Bruin and Alexandra Schättin

166 ***Me, Myself, and Not-I: Self-Discrepancy Type Predicts Avatar Creation Style***
Mitchell G. H. Loewen, Christopher T. Burris and Lennart E. Nacke

173 ***How the Visual Design of Video Game Antagonists Affects Perception of Morality***
Reyhan Pradantyo, Max V. Birk and Scott Bateman

frontiers in Computer Science

ORIGINAL RESEARCH
published: 14 January 2020
doi: 10.3389/fcomp.2019.00012

Guidelines for the Development of Immersive Virtual Reality Software for Cognitive Neuroscience and Neuropsychology: The Development of Virtual Reality Everyday Assessment Lab (VR-EAL), a Neuropsychological Test Battery in Immersive Virtual Reality

Panagiotis Kourtesis[1,2,3,4]*, *Danai Korre*[5], *Simona Collina*[3,4], *Leonidas A. A. Doumas*[2] and *Sarah E. MacPherson*[1,2]

[1] Human Cognitive Neuroscience, Department of Psychology, University of Edinburgh, Edinburgh, United Kingdom, [2] Department of Psychology, University of Edinburgh, Edinburgh, United Kingdom, [3] Lab of Experimental Psychology, Suor Orsola Benincasa University of Naples, Naples, Italy, [4] Interdepartmental Centre for Planning and Research "Scienza Nuova", Suor Orsola Benincasa University of Naples, Naples, Italy, [5] Centre for Intelligent Systems and Their Applications, School of Informatics, University of Edinburgh, Edinburgh, United Kingdom

OPEN ACCESS

Edited by:
Ioanna Iacovides,
University of York, United Kingdom

Reviewed by:
Manuela Chessa,
University of Genoa, Italy
Magdalena Mendez-Lopez,
University of Zaragoza, Spain
Lukas Gehrke,
Technische Universität Berlin, Germany

***Correspondence:**
Panagiotis Kourtesis
pkourtes@exseed.ed.ac.uk

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Computer Science

Received: 11 September 2019
Accepted: 16 December 2019
Published: 14 January 2020

Citation:
Kourtesis P, Korre D, Collina S, Doumas LAA and MacPherson SE (2020) Guidelines for the Development of Immersive Virtual Reality Software for Cognitive Neuroscience and Neuropsychology: The Development of Virtual Reality Everyday Assessment Lab (VR-EAL), a Neuropsychological Test Battery in Immersive Virtual Reality. Front. Comput. Sci. 1:12. doi: 10.3389/fcomp.2019.00012

Virtual reality (VR) head-mounted displays (HMD) appear to be effective research tools, which may address the problem of ecological validity in neuropsychological testing. However, their widespread implementation is hindered by VR induced symptoms and effects (VRISE) and the lack of skills in VR software development. This study offers guidelines for the development of VR software in cognitive neuroscience and neuropsychology, by describing and discussing the stages of the development of Virtual Reality Everyday Assessment Lab (VR-EAL), the first neuropsychological battery in immersive VR. Techniques for evaluating cognitive functions within a realistic storyline are discussed. The utility of various assets in Unity, software development kits, and other software are described so that cognitive scientists can overcome challenges pertinent to VRISE and the quality of the VR software. In addition, this pilot study attempts to evaluate VR-EAL in accordance with the necessary criteria for VR software for research purposes. The VR neuroscience questionnaire (VRNQ; Kourtesis et al., 2019b) was implemented to appraise the quality of the three versions of VR-EAL in terms of user experience, game mechanics, in-game assistance, and VRISE. Twenty-five participants aged between 20 and 45 years with 12–16 years of full-time education evaluated various versions of VR-EAL. The final version of VR-EAL achieved high scores in every sub-score of the VRNQ and exceeded its parsimonious cut-offs. It also appeared to have better in-game assistance and game mechanics, while its improved graphics substantially increased the quality of the user experience and almost eradicated VRISE. The results substantially support the feasibility of the development of effective VR research and clinical software without the presence of VRISE during a 60-min VR session.

Keywords: virtual reality, prospective memory, episodic memory, cybersickness, executive function, neuropsychology, everyday functioning, attention

INTRODUCTION

In cognitive neuroscience and neuropsychology, the collection of cognitive and behavioral data is predominantly achieved by implementing psychometric tools (i.e., cognitive screening and testing). The psychometric tools are principally limited to paper-and-pencil and computerized (i.e., 2D and 3D applications) forms. Psychometric tools in clinics and/or laboratories display several limitations and discrepancies between the observed performance in the laboratory/clinic and the actual performance of individuals in everyday life (Rizzo et al., 2004; Bohil et al., 2011; Parsons, 2015). The functional and predictive association between an individual's performance on a set of neuropsychological tests and the individual's performance in various everyday life settings is called ecological validity. Ecological validity is considered an important issue that cannot be resolved by the currently available assessment tools (Rizzo et al., 2004; Bohil et al., 2011; Parsons, 2015).

Ecological validity is especially important in the assessment of certain cognitive functions, which are crucial for performance in everyday life (Chaytor and Schmitter-Edgecombe, 2003). In particular, executive functioning (e.g., multitasking, planning ability, and mental flexibility) has been found to predict occupational and academic success (Burgess et al., 1998). Similarly, the ecologically valid measurement of memory (e.g., episodic memory) and attentional processes (e.g., selective, divided, and sustained attention) have been seen as predictors of overall performance in everyday life (Higginson et al., 2000). Lastly, prospective memory (i.e., the ability to remember to carry out intended actions at the correct point in the future; McDaniel and Einstein, 2007) plays an important role in everyday life and the assessment of prospective memory abilities requires ecologically valid tasks (Phillips et al., 2008).

Current ecologically valid tests are not thought to encompass the complexity of real-life situations (Rizzo et al., 2004; Bohil et al., 2011; Parsons, 2015). Assessments which take place in real-world settings (e.g., performing errands in a shopping center) are time consuming and expensive to set up, lack experimental control over the external situation (e.g., Elkind et al., 2001), cannot be standardized for use in other labs, and are not feasible for certain populations (e.g., individuals with psychiatric conditions or motor difficulties; Rizzo et al., 2004; Parsons, 2015). The traditional approaches in cognitive sciences encompass the employment of static and simple stimuli, which lack ecological validity. Instead, immersive virtual reality (VR) technology enables cognitive scientists to accumulate advanced cognitive and behavioral data through the employment of dynamic stimuli and interactions with a high degree of control within an ecologically valid environment (Rizzo et al., 2004; Bohil et al., 2011; Parsons, 2015). Furthermore, VR can be combined with non-invasive imaging techniques (Makeig et al., 2009; Bohil et al., 2011; Parsons, 2015), wearable mobile brain/body imaging (Makeig et al., 2009), and can be used for rehabilitation and treatment purposes (Rizzo et al., 2004; Bohil et al., 2011; Parsons, 2015).

VR has great potential as an effective telemedicine tool that may resolve the current methodological problem of ecological validity (Rizzo et al., 2004; Bohil et al., 2011; Parsons, 2015; Parsons et al., 2018). However, the appropriateness of VR, especially for head-mounted display (HMD) systems, is still controversial (Bohil et al., 2011; de França and Soares, 2017; Palmisano et al., 2017). The principal concern is the adverse symptomatology (i.e., nausea, dizziness, disorientation, fatigue, and instability) which stems from the implementation of VR systems (Bohil et al., 2011; de França and Soares, 2017; Palmisano et al., 2017). These adverse VR induced symptoms and effects (VRISE) endanger the health and safety of the users (Parsons et al., 2018), decrease reaction times and overall cognitive performance (Nalivaiko et al., 2015), while increasing body temperature and heart rates (Nalivaiko et al., 2015), cerebral blood flow and oxyhemoglobin concentration (Gavgani et al., 2018), brain activity (Arafat et al., 2018), and the connectivity between brain regions (Toschi et al., 2017). Hence, VRISE may compromise the reliability of cognitive, physiological, and neuroimaging data (Kourtesis et al., 2019a).

However, VRISE predominantly stem from hardware and software inadequacies, which more contemporary commercial VR hardware and software do not share (Kourtesis et al., 2019a,b). The employment of modern VR HMDs analogous to or more cutting-edge than the HTC Vive and/or Oculus Rift, in combination with ergonomic VR software, appear to significantly mitigate the presence of VRISE (Kourtesis et al., 2019a,b). However, the selection of suitable VR hardware and/or software demands acceptable technological competence (Kourtesis et al., 2019a). Minimum hardware and software features have been suggested to appraise the suitability of VR hardware and software (Kourtesis et al., 2019a). The technical specifications of the computer and VR HMD are adequate to assess their quality (Kourtesis et al., 2019a), while the virtual reality neuroscience questionnaire (VRNQ) facilitates the quantitative evaluation of software attributes and the intensity of VRISE (Kourtesis et al., 2019b).

Another limitation is that the implementation of VR technology may necessitate high financial costs, which hinders its widespread adoption by cognitive scientists. In the 90s, the cost of a VR lab with basic features cost between $20,000 and 50,000, where nowadays the cost has decreased considerably (Slater, 2018). At present, the cost of a VR lab with basic features (e.g., a HMD, external hardware, and laptop) is between

$2,000 and 2,500. However, the development of VR software is predominantly dependent on third parties (e.g., freelancers or companies) with programming and software development skills (Slater, 2018). A solution that will promote the adoption of immersive VR as a research and clinical tool might be the in-house development of VR research/clinical software by computer science literate cognitive scientists or research software engineers.

The current study endeavors to offer guidelines on the development of VR software by presenting the development of the Virtual Reality Everyday Assessment Lab (VR-EAL). Since the assessment of prospective memory, episodic memory, executive functions, and attention are likely to benefit from ecologically valid approaches to assessment, VR-EAL attempts to be one of the first neuropsychological batteries to apply immersive VR to assess these cognitive functions. However, the ecologically valid assessment of these cognitive functions demands the development of a realistic scenario with several scenes and complex interactions while avoiding intense VRISE factors.

The VR-EAL development process is presented systematically, aligned with the steps that cognitive scientists should follow to achieve their aim of designing VR studies. Firstly, the preparation stages are described and discussed. Secondly, the structure of the application (e.g., order of the scenes) is presented and discussed in terms of offering comprehensive tutorials, delivering a realistic storyline, and incorporating a scoring system. Thirdly, a pilot study is conducted to evaluate the suitability of the different versions of VR-EAL (i.e., alpha, beta, final) for implementation in terms of user experience, game mechanics, in-game assistance, and VRISE.

DEVELOPMENT OF VR-EAL

Rationale and Preparation

Prospective memory encompasses the ability to remember to initiate an action in the future (Anderson et al., 2017). The prospective memory action may be related to a specific event (e.g., when you see this person, give him a particular object) or time (e.g., at 5 p.m. perform a particular task). Attentional control processes, executive functioning, the difficulty of the filler/distractor tasks, the length of the delay between encoding the intention to perform a task and the presentation of the stimulus-cue, as well as the length of the ongoing task, all affect prospective memory ability (Anderson et al., 2017). Therefore, the VR-EAL scenarios need to incorporate both types of prospective memory actions and consider the length and difficulty of the distractor tasks and delays, as well as attentional and executive functioning.

The main theoretical frameworks of prospective memory are the preparatory attentional and memory (PAM) and the multiprocess (MP) theories (Anderson et al., 2017). The PAM theory suggests that performing prospective memory tasks efficiently requires a constant top-down monitoring for environmental and internal cues in order to recall the intended action and perform it (Smith, 2003; Smith et al., 2007). For example, an individual wants to buy a pint of milk after work. On her way home, she is vigilant (i.e., monitoring) about recognizing prompts (e.g., the sign of a supermarket) that will remind her of her intention to buy a pint of milk. In addition to PAM's top-down monitoring, MP theory suggests that bottom-up spontaneous retrieval also enables effective performance on prospective memory tasks (McDaniel and Einstein, 2000, 2007). Going back to the previous example, when the individual is not being vigilant (i.e., passive), she sees an advert pertaining to dairy products, which triggers the retrieval of her intention to buy a pint of milk. VR-EAL is required to incorporate both predominant retrieval strategies in line with these main theoretical frameworks of prospective memory (i.e., PAM and MP). This may be achieved by including scenes where the user should be vigilant (i.e., PAM) so they recognize a stimulus associated with the prospective memory task (e.g., notice a medicine on the kitchen's table in order to take it after having breakfast), as well as scenes where the user passively (i.e., MP) will attend to an obvious stimulus related to the prospective task (e.g., while being in front of the library, the user needs to remember to return a book).

Notably, the ecologically valid assessment of executive (i.e., planning and multitasking), attentional (i.e., selective visual, visuospatial, and auditory attention), and episodic memory processes is an equally important aim of VR-EAL. The relevant literature postulates that the everyday functioning of humans is dependent on cognitive abilities, such as attention, episodic memory, prospective memory, and executive functions (Higginson et al., 2000; Chaytor and Schmitter-Edgecombe, 2003; Phillips et al., 2008; Rosenberg, 2015; Mlinac and Feng, 2016; Haines et al., 2019). However, the assessment of these cognitive functions requires an ecologically valid approach to indicate the quality of the everyday functioning of the individual in the real world (Higginson et al., 2000; Chaytor and Schmitter-Edgecombe, 2003; Phillips et al., 2008; Rosenberg, 2015; Mlinac and Feng, 2016; Haines et al., 2019). However, the assessment (i.e., tasks) of these cognitive functions in VR-EAL will also serve as distractor tasks for the prospective memory components of the paradigm. Hence, the VR-EAL distractor tasks are vital to the prospective memory tasks, but at the same time, they are adequately challenging within a continuous storyline (see **Table 1**).

Furthermore, ecologically valid tasks performed in VR environments demand various game mechanics and controls to facilitate ergonomic and naturalistic interactions, and these need to be learnt by users. The scenario should include tutorials that allow users to spend adequate time learning how to navigate, use and grab items, and how the VE reacts to his/her actions (Gromala et al., 2016; Jerald et al., 2017; Brade et al., 2018; see **Table 1**). Additionally, the scenario should consider the in-game instructions and prompts offered to users such as directional arrows, non-player characters (NPC), signs, labels, ambient sounds, audio, and videos that aid performance (Gromala et al., 2016; Jerald et al., 2017; Brade et al., 2018). Importantly, this user-centered approach appears to particularly favor non-gamers in terms of performing better and enjoying the VR experience (Zaidi et al., 2018). Thus, the development of VR-EAL should be aligned with these aforementioned suggestions.

The first step of the development process was to select the target platform. In VR's case, this is the VR HMD, which allows

TABLE 1 | VR-EAL Scenario.

Order	Type	Description
Scene 1	*Tutorial*	Basic interactions and navigation
Scene 2	*Tutorial*	Interactive boards (recognition and planning)
Scene 3	Storyline	List of prospective memory tasks, shopping list (immediate recognition), and itinerary (planning)
Scene 4	*Tutorial*	List of mechanics for the prospective memory tasks, prompts, and notes
Scene 5	*Tutorial*	Cooking
Scene 6	Storyline	Prepare breakfast (multi-tasking) and take medication (prospective memory, event-based, short delay)
Scene 7	*Tutorial*	Tutorial: collect items
Scene 8	Storyline	Collect items from the living-room (selective visuospatial attention) and take a chocolate pie out of the oven (prospective memory, event-based, short delay)
Scene 9	*Tutorial*	Interaction with 3D non-player characters
Scene 10	Storyline	Call Rose (prospective memory task, time-based, short delay)
Scene 11	*Tutorial*	Gaze interaction
Scene 12	Storyline	Detect posters on both sides of the road (selective visual attention)
Scene 13	*Tutorial*	Shopping, how to collect the items from the supermarket
Scene 14	Storyline	Collect the shopping list items from the supermarket (delayed recognition)
Scene 15	Storyline	Go to the bakery to collect the carrot cake (prospective memory task, time-based, medium delay)
Scene 16	Storyline	False prompt before going to the library (prospective memory task, event-based, medium delay)
Scene 17	Storyline	Return the red book to the library (prospective memory task, event-based, medium delay)
Scene 18	*Tutorial*	Auditory interaction
Scene 19	Storyline	Detect sounds from both sides of the road (selective auditory attention)
Scene 20	Storyline	False prompt before going back home (prospective memory task, time-based, long delay)
Scene 21	Storyline	When you return home, give the extra pair of keys to Alex (prospective memory task, event-based, long delay)
Scene 22	Storyline	Put away the shopping items and take the medication (prospective memory task, time-based, long delay)

various interactions to take place within a virtual environment (VE) during the neuropsychological assessment. In our previous work (Kourtesis et al., 2019a), we have highlighted a number of suggested minimum hardware and software features which appraise the suitability of VR hardware and software. Firstly, interactions with the VE should be ergonomic in order to elude or alleviate the presence of VRISE. Also, the utilization of 6 degrees of freedom (DoF) wands (i.e., controllers) facilitates ergonomic interactions and provides highly accurate motion tracking. Lastly, the two types of HMD that exceed the minimum standards and support 6DoF controllers are the HTC Vive and Oculus Rift; hence, the target HMD should have hardware characteristics equal to or greater than these high-end HMDs (Kourtesis et al., 2019a). VR-EAL is developed to be compatible with HTC Vive, HTC Vive Pro, Oculus Rift, and Oculus Rift-S.

The second step was to select which game engine (GE) should be used to develop the VR software. For the development of VR-EAL, the feasibility of acquiring the required programming and software development skills was an important criterion for the selection of the GE because the developer of VR-EAL (i.e., the corresponding author) is a cognitive scientist who did not have any background in programming or software development. The two main GEs are Unity and Unreal. Unity requires C# programming skills, while Unreal requires C++ programming skills. Learners of C#, either experienced or inexperienced programmers, appear to experience a greater learning curve than learners of C++ (Chandra, 2012). While Unity and Unreal are of equal quality (Dickson et al., 2017), Unity as a GE has been found to be more user-friendly, and easier to learn compared to Unreal (Dickson et al., 2017). Also, Unity has an extensive online community and online resources (e.g., 3D models, software development kits; SDK), and documentation (Dickson et al., 2017). For these reasons, Unity was preferred for the development of VR-EAL. However, either Unreal or Unity would have been a sensible choice since both GEs offer high quality tools and features for software development (Dickson et al., 2017).

The final step was the acquisition of skills and knowledge. A cognitive scientist with a background either in computer or psychological sciences should have knowledge of the cognitive functions to be studied, as well as, intermediate programming and software development skills pertinent to the GE. The acquisition of these skills enables the cognitive scientist to design the VR software in agreement with the capabilities of the GE and the research aims. In VR-EAL's case, its developer meticulously studied the established ecologically valid paper-and-pencil tests such as the Test of Everyday Attention (TEA; Robertson et al., 1994), the Rivermead Behavioral Memory Test—III (RBMT-III; Wilson et al., 2008), the Behavioral Assessment of the Dysexecutive Syndrome (BADS; Wilson et al., 1998), and the Cambridge Prospective Memory Test (CAMPROMPT; Wilson et al., 2005). In addition, other research and clinical software were considered. For example, the Virtual Reality Shopping Task (Canty et al., 2014), Virtual Reality Supermarket (Grewe et al., 2014), Virtual Multiple Errands Test (Rand et al., 2009), the Invisible Maze Task (Gehrke et al., 2018), and the Jansari

Assessment of Executive Function (Jansari et al., 2014) are non-immersive VR software which assess cognitive functions such as executive functions, attentional processes, spatial cognition, and prospective memory.

Finally, the developer of VR-EAL attained intermediate programming skills in C# and software development skills in Unity. This was predominantly achieved by attending online specializations and tutorials on websites such as Coursera, Udemy, CodeAcademy, SoloLearn, and EdX. Also, a developer may consider established textbooks such as the "The VR book: Human-centered design for virtual reality" (Jerald, 2015), "3D user interfaces: theory and practice" (LaViola et al., 2017), and "Understanding virtual reality: Interface, application, and design" (Sherman and Craig, 2018). To sum it up, the acquisition of these skills enabled progression to the next stage of the development of VR-EAL, which is the writing of the scenarios/scripts.

Tutorials and Mechanics

VR-EAL commences with two tutorial scenes. The first tutorial allows the user to learn how to navigate using teleportation, to hold and manipulate items (e.g., throwing them away), how to use items (e.g., pressing a button), as well as to familiarize themselves with the in-game assistance objects (e.g., a directional arrow or a sign; see **Figure 1**). The user is prompted to spend adequate time learning the basic interactions and navigation system because these game mechanics and in-game assistance methods are fundamental to most scenes in VR-EAL.

The second tutorial instructs the user how to use interactive boards (i.e., use a map or select items from a list). This tutorial is specific to the tasks that the user should perform in the subsequent storyline scene. Similarly, the remaining tutorials are specific to their subsequent scene (i.e., the actual task) in which the user is assessed. This design enables the user to perform the tasks, without providing them with an overwhelming amount of information that may confuse the user. However, the tutorial in the fourth scene is specific to the prospective memory tasks that are performed in several scenes throughout the scenario. The instructions for all prospective memory tasks (i.e., what should be performed and when) are provided during the third scene (i.e., storyline-bedroom scene), but the first prospective memory task is not performed until the sixth scene (i.e., the cooking task; see **Table 1**).

In scene 4, the user learns how to use a VR digital watch, use prospective memory items and notes (toggle on/off the menu), and follow prospective memory prompts. These game mechanics are essential to successfully perform the prospective memory tasks. The VR digital watch is the main tool for checking the time in relation to the time-based prospective memory tasks, while the prospective memory notes are crucial reminders for the time- and event-based prospective memory tasks. Subsequently, in scene 5, the user completes a tutorial where s/he learns how to use the oven and the stove as well as the snap-drop-zones to perform the cooking task. The snap-drop-zones are game objects, which are containers that the user may attach other game objects too. In scene 7, the user learns how to collect items using the snap-drop-zones attached to the left controller (see **Figure 2**).

In scene 9, the user learns how to interact with the 3D non-player characters (NPC). The user is required to talk to the NPC to initiate a conversation (i.e., detection of a sound through the mic input), and use the interactive boards to select a response, which either presents a dichotomous choice (i.e., "yes" or "no") or a list of items (see **Figure 2**). These interactions with the NPC are central to the assessment of prospective memory, and the user should effectively interact with the NPC in six scenes to successfully perform an equal number of time- and event-based prospective memory tasks.

In scene 11, the user learns how to use gaze interactions. There is a circular crosshair, which indicates the collision point of a ray that is emitted from the center of the user's visual field. The user is required to direct the circular crosshair over the targets and avoid the distractors (see **Figure 2**). The user needs to effectively perform a practice trial to proceed to the next scene. The practice trial requires the user to spot the three targets and avoid all the distractors while moving. If the user is unsuccessful, then the practice trial is re-attempted. This procedure is repeated until the user effectively completes the practice trial.

Scene 13 is a short tutorial where the user is reminded how to collect items using the snap-drop-zones attached to the left controller and remove an item from the snap-drop-zone in cases where an item is erroneously picked up. In scene 18, the user learns how to detect target sounds (i.e., a bell) and avoid distractors (i.e., a high-pitched and a low-pitched bell). The user looks straight ahead and presses the trigger button on the right controller when a target sound is heard on the right side. Likewise, the user presses the trigger button on the left controller when a target sound is heard on the left side (see **Figure 2**). The sounds are activated by trigger-zones, which are placed within the itinerary of the user. This tutorial is conducted in a similar way to the scene 11 tutorial (i.e., gaze interaction). The user, while being on the move, needs to detect three target sounds and avoid the distractors to proceed to the next scene.

The time spent on each tutorial is recorded to provide the learning time for the various interaction systems (i.e., game mechanics). However, in the scene 11 and 18 tutorials, the practice trial times are also recorded. The collected data (i.e., time spent on tutorials and the attempts to complete the practice trials) for each tutorial are added to a text file that contains the user's data (i.e., performance scores on every task).

Storyline and Scoring

The required times to complete scenes and tasks are recorded. However, the task times are measured independently from the total scene times. Additionally, in the scenes where the user should perform prospective memory tasks, the number of times and the duration that the prospective memory notes appeared are also measured. These variables indicate how many times the user relies on the prospective memory notes, and how long they read them for.

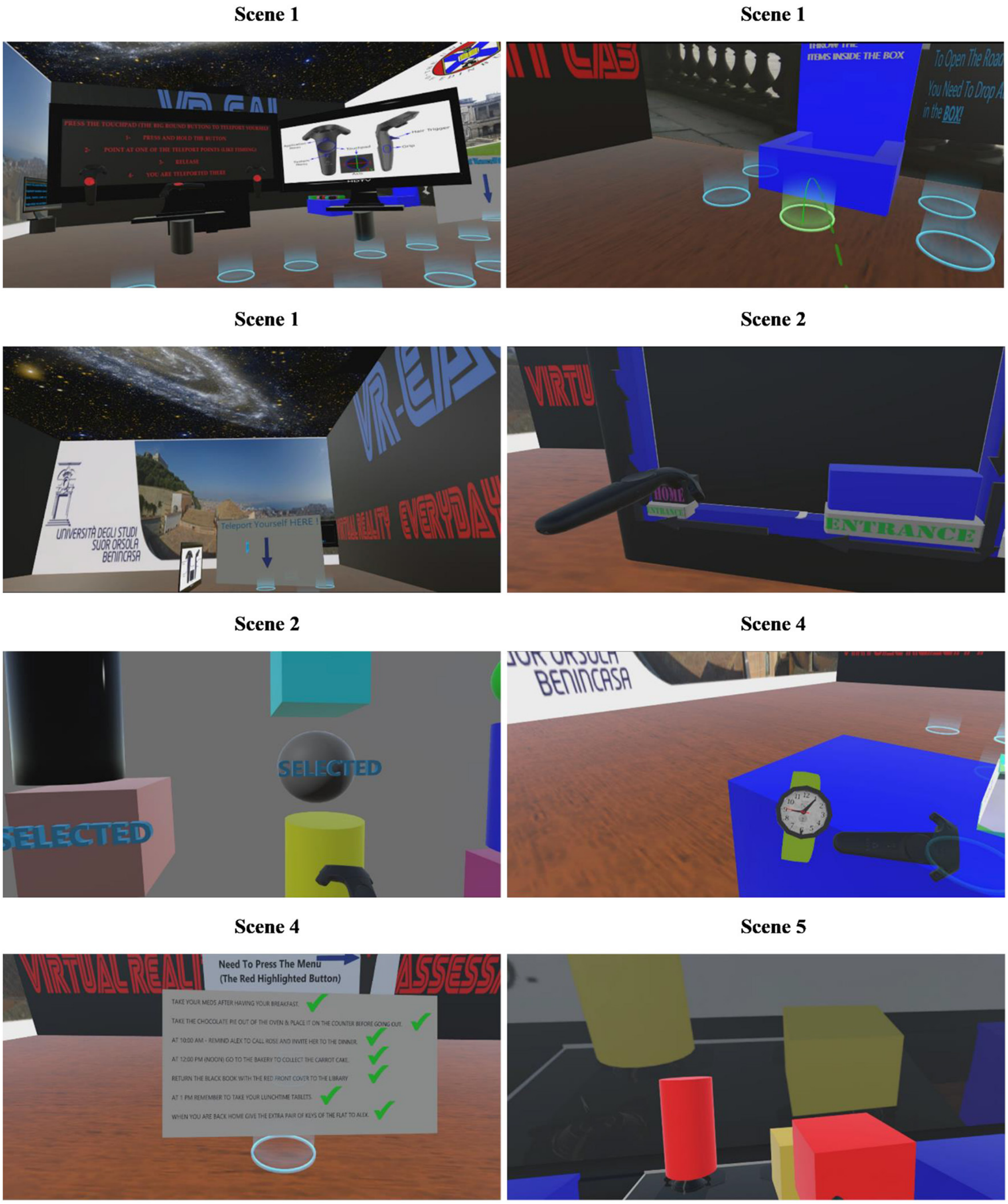


FIGURE 1 | VR-EAL tutorials: scenes 1–5.

At Home

Bedroom: immediate recognition and planning

The storyline commences in a bedroom (i.e., scene 3; see **Table 1**), where the user receives an incoming call from his/her close friend, Sarah, asking the user to carry out some errands for her (e.g., buy some shopping from the supermarket, collect a carrot cake from the bakery, return a library book). All the errands are prospective memory tasks except the shopping task. In this scene, the user should perform three different tasks. The first task is the prospective memory notes (i.e., PM-Notes) task, where the user responds affirmatively or negatively to three prompts asking the user to write down the errands (i.e., PM-tasks). The response of the user indicates his/her intention to use external tools (i.e., notes) as reminders.

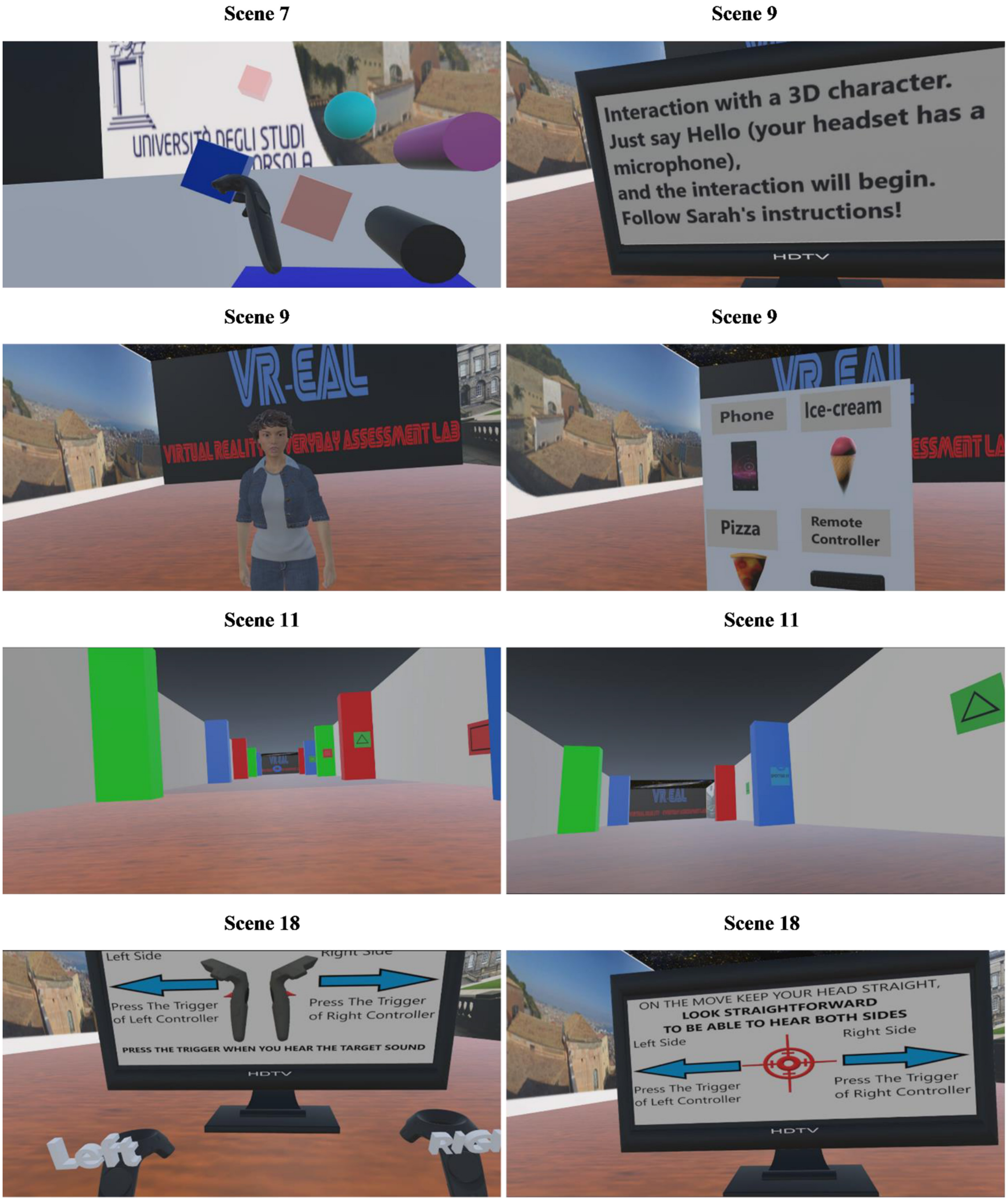


FIGURE 2 | VR-EAL tutorials: scenes 7–18.

The second task is the immediate recognition task where the user should choose the 10 target items (i.e., create the shopping list) from an extensive array of items (see **Figure 3**), which also contains five qualitative distractors (e.g., semi-skimmed milk vs. skimmed milk), five quantitative distractors (e.g., 1 vs. 2 kg potatoes), and 10 false items (e.g., bread, bananas etc.). The user gains 2 points for each correctly chosen item, 1 point for choosing a qualitative or quantitative distractor, and 0 points for the false items. The maximum possible score is 20 points.

The third task is the planning task. The user should draw a route on a map to visit three destinations (i.e., the supermarket, bakery, and library) before returning home. The road system comprises 23 street units (see **Figure 3**). When the user selects a unit, 1 point is awarded. The ideal route to visit all three destinations is 15 units; hence, any extra or missing units are

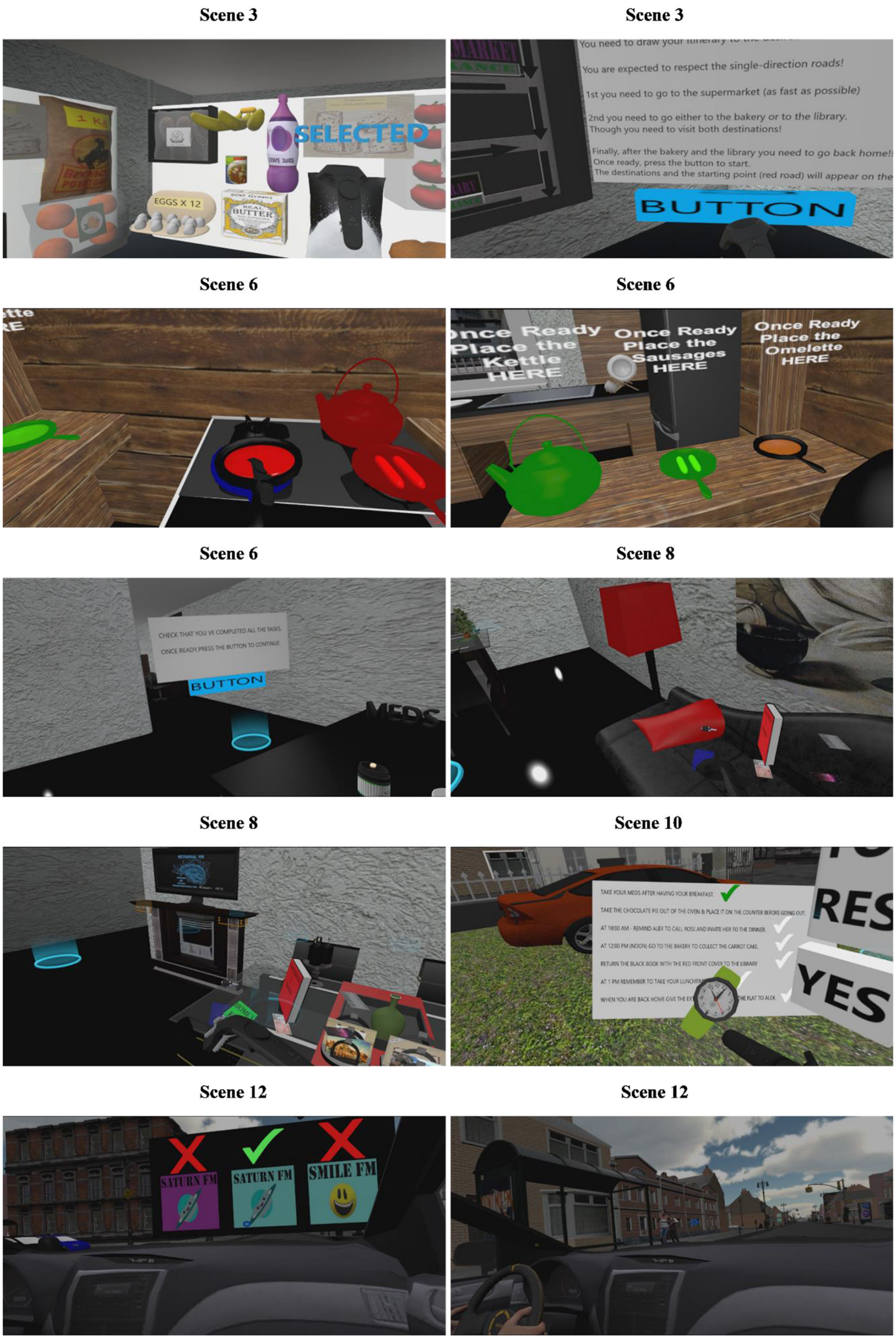


FIGURE 3 | VR-EAL storyline: scenes 3–12.

subtracted from the total possible score of 15. For example, if the user draws 18 units, then the distance from the ideal route is calculated as 3 (i.e., 18 − 15 = 3). Three is then subtracted from the ideal score of 15, resulting in a score of 12. If the user draws 12 units, the distance from the ideal route is also 3 and again 3 is subtracted from 15, resulting in a score of 12. The planning task score is also modified by the planning task completion time (e.g., a completion time two standard deviations below the average time of the normative sample is awarded 2 points while two standard deviations above the average time is subtracted 2 points).

Kitchen: multitasking and prospective memory task

In the kitchen (i.e., scene 6; see **Table 1**), the user should complete two main tasks: the cooking task (i.e., preparing breakfast) and a prospective memory task. The cooking task encompasses frying an omelet and sausages, putting a chocolate pie in the oven, as well as boiling some water for a cup of tea or coffee. The user must handle two pans (one for the omelet and one for the sausages) and a kettle. Images of the omelet and sausages are presented above the cooker to display what their appearance should be when they are ready. Scoring relies on the animations from each game object (i.e., the omelet and the sausages). At the beginning of the animation, both items have a reddish (raw) color which gradually turns to either a yellowish (omelet) or brownish (sausages) color, and finally both turn to black (burnt). The score for each pan hence depends on the time that the user removes the pans from the stove (pauses/stops the animation) and places them on the kitchen worktop (for calculation of the score, see **Figure 4**). Equally, the score for boiling the kettle is measured in relation to the stage of the audio playback (e.g., the water is ready when the kettle whistles; see **Figure 4**) when the kettle is placed on the kitchen worktop.

After breakfast, the user needs to take his/her meds (i.e., a prospective memory task). When the user has had his/her breakfast, the final button of the scene appears (see **Figure 3**). The user should press this button to confirm that all the tasks in the scene are completed. If the user has already taken his/her medication before pressing the final button, then the scene ends, and the user receives 6 points. Otherwise, the first prompt appears (i.e., "You Have to Do Something Else"). If the user then follows the prompt and takes their medication, they receive 4 points. If the user presses the final button again, then the second prompt appears (i.e., "You Have to Do Something After Having your Breakfast"). If the user follows this prompt and takes their medication, they receive 2 points. If the user presses the final button again, then the third prompt appears (i.e., "You Have to Take Your Meds"). If the user follows this prompt and takes their medication, they then receive 1 point. If the user represses the final button without ever taking their medication, they get zero points, and the scene ends.

Living room: selective visuospatial attention and prospective memory task

In the living room (i.e., scene 8; see **Table 1**), the user should collect six items (i.e., a red book, £20, a smartphone, a library card, the flat keys, and the car keys) that are placed in various locations within the living room (see **Figure 3**). The user is not required to memorize the items since there is a reminder list on one of the walls of the living room. The user collects the items by attaching them to the snap-drop-zones attached to the left controller. The user receives 1 point for each item collected. However, there are distractors (e.g., magazines, books, a remote control, a notebook, a pencil, a chessboard, and a bottle of wine) in the room. If the user attempts to collect one of the distractors, the item falls (only the target items can be attached to the snap-drop-zones), which counts as an error. After collecting all the objects, the user needs to take the chocolate pie out of the oven and place it on the kitchen worktop before leaving the apartment (prospective memory task; see scoring for medication above).

Garden: prospective memory task

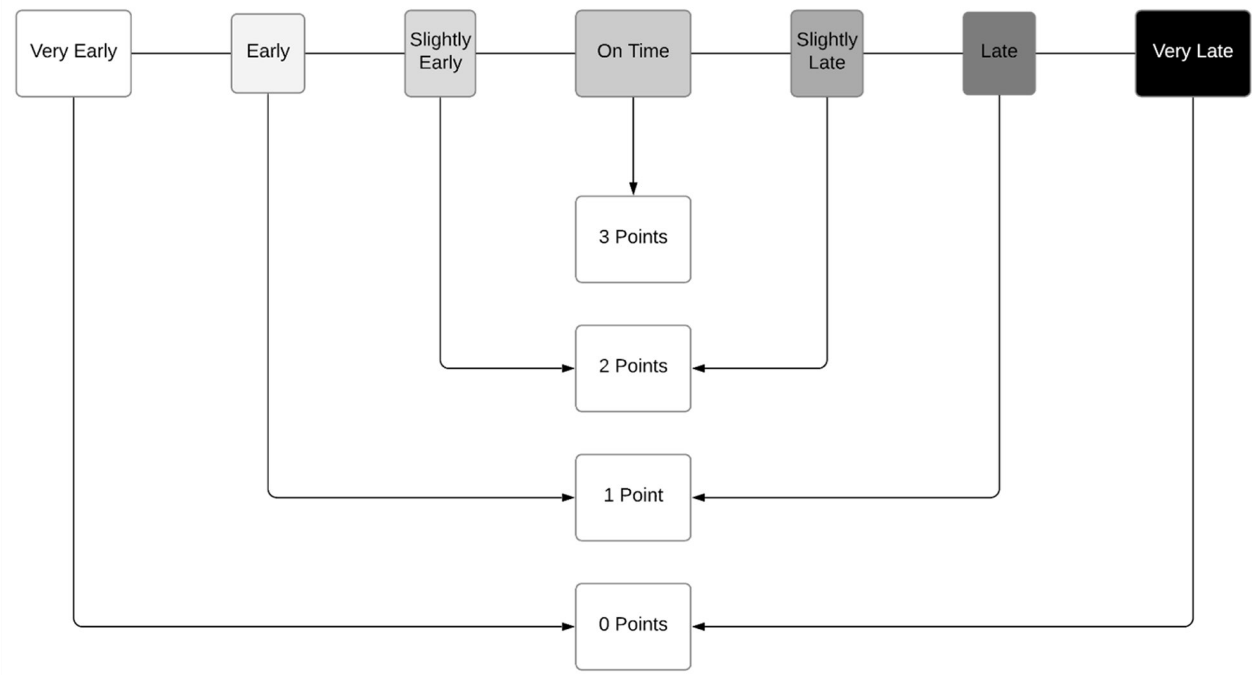
In the garden (i.e., scene 10; see **Table 1**), the user initiates a conversation with Alex (an NPC), to perform a distractor task (i.e., to open the gate). The conversation continues after this distractor task, where the user needs to respond to Alex's question (i.e., "Do we need to do something else at this time?") by selecting either "yes" or "no" (see **Figure 3**). This action is considered as the first prompt for the prospective memory task, and if the user responds "yes," then the second interactive board appears (see **Figure 5** for scoring). If the user selects "no," then the second prompt is given by Alex (i.e., "Are you sure that we do not have to do something around this time?"). If the user selects "no," then the third prompt is provided by Alex (i.e., "I think that we have to do something around this time."). If the user again selects "no," clarification is provided by Alex (i.e., "Oh yes, we need to call Rose"), and the user receives 0 points (see **Figure 5**).

When the user chooses "yes," the second interactive board appears. This second interactive board displays eight items (see **Figure 3**). There is one item, which presents the correct prospective memory response (i.e., the smartphone). There are also three items which are responses related to the other prospective memory tasks (i.e., a red book, carrot cake, flat keys). There is one item, which is semantically related to the correct prospective memory response (i.e., a tablet computer). Also, there are three items which are unrelated distractors, which are neither related to the other prospective memory tasks, nor are in the same semantic category as the correct prospective memory response (i.e., ice cream and a smartphone). Scoring depends on the user's responses on the first and the second interactive boards (see **Figure 5**).

In the City

On the road: selective visual attention

In this scenario, the user is a passenger in a car with Alex driving (i.e., scene 12; see **Table 1**). The radio is on, and the speaker announces a competition where the user needs to identify all the radio stations' target posters and avoid the distractor ones (see **Figure 3**), which are hung along the street. There are 16 target posters and 16 distractors equally allocated on both sides of the street. Eight of the distractors have the same shape as the target poster, but a different background color. The other eight distractors have the same background color as the target posters, but they are a different shape (see **Figure 3**). The user is awarded



Cooking Item	Very Early	Early	Slightly Early	On Time	Slightly Late	Late	Very Late
Pan-Omelette	0-13.99 secs	14-15.99 secs	16-17.99 secs	18-22 secs	22.01-23.99 secs	24-26 secs	> 26 secs
Pan-Sausages	0-17.99 secs	18-19.99 secs	20-21.99 secs	22-26 secs	26.01-27.99 secs	28-30 secs	> 30 secs
Kettle	0-10.99 secs	11-12.99 secs	13-14.99 secs	15-17 secs	17.01-18.99 secs	19-21 secs	> 21 secs

FIGURE 4 | A schematic representation of the cooking task scoring.

1 point when a target poster is "spotted" and subtracted 1 point when a distractor poster is "spotted." The maximum score is 16, and the number of correctly identified posters and distractors (for each type) identified on each side of the road is recorded.

Supermarket: delayed recognition and prospective memory task

The user arrives at the supermarket (i.e., scene 14; see **Table 1**), where s/he should buy the items from the shopping list. The user navigates within the shop by following the arrows, and collects the items using the snap-drop-zones attached to the left controller (see **Figure 6**). The items on the shelves of the supermarket are the same items as the immediate recognition task in scene 3 (see Bedroom). The scoring system is identical to the immediate recognition task in scene 3 (see Bedroom), and the score is calculated when the user arrives at the till to buy the items. Outside the supermarket (i.e., scene 15), the user has another conversation with Alex, where s/he needs to remember that they must collect the carrot cake at 12 noon (i.e., a prospective memory task). The conversation is performed and scored in the same way as the prospective memory task in scene 10 (see Garden). The user then goes with Alex to the bakery to collect the carrot cake.

Bakery and library: prospective memory tasks

The user is outside the bakery (i.e., scene 16; see **Table 1**), after already collecting the carrot cake. Here, they have another interaction with Alex where he asks, "Do we need to do something else at this time?" However, this time, there is no prospective memory task to perform and the user should respond negatively. This deception helps to examine whether the user is simply responding affirmatively to all prospective memory task prompts. If the user responds affirmatively (i.e., "yes"), then s/he loses points (see **Figure 7**). This conversation is similar to the prospective memory task in scene 10 (see Garden). However, the scoring is now inverted, where the user should choose "no" three times in response to Alex's prompts to avoid points being subtracted. In the prospective memory task that follows in the

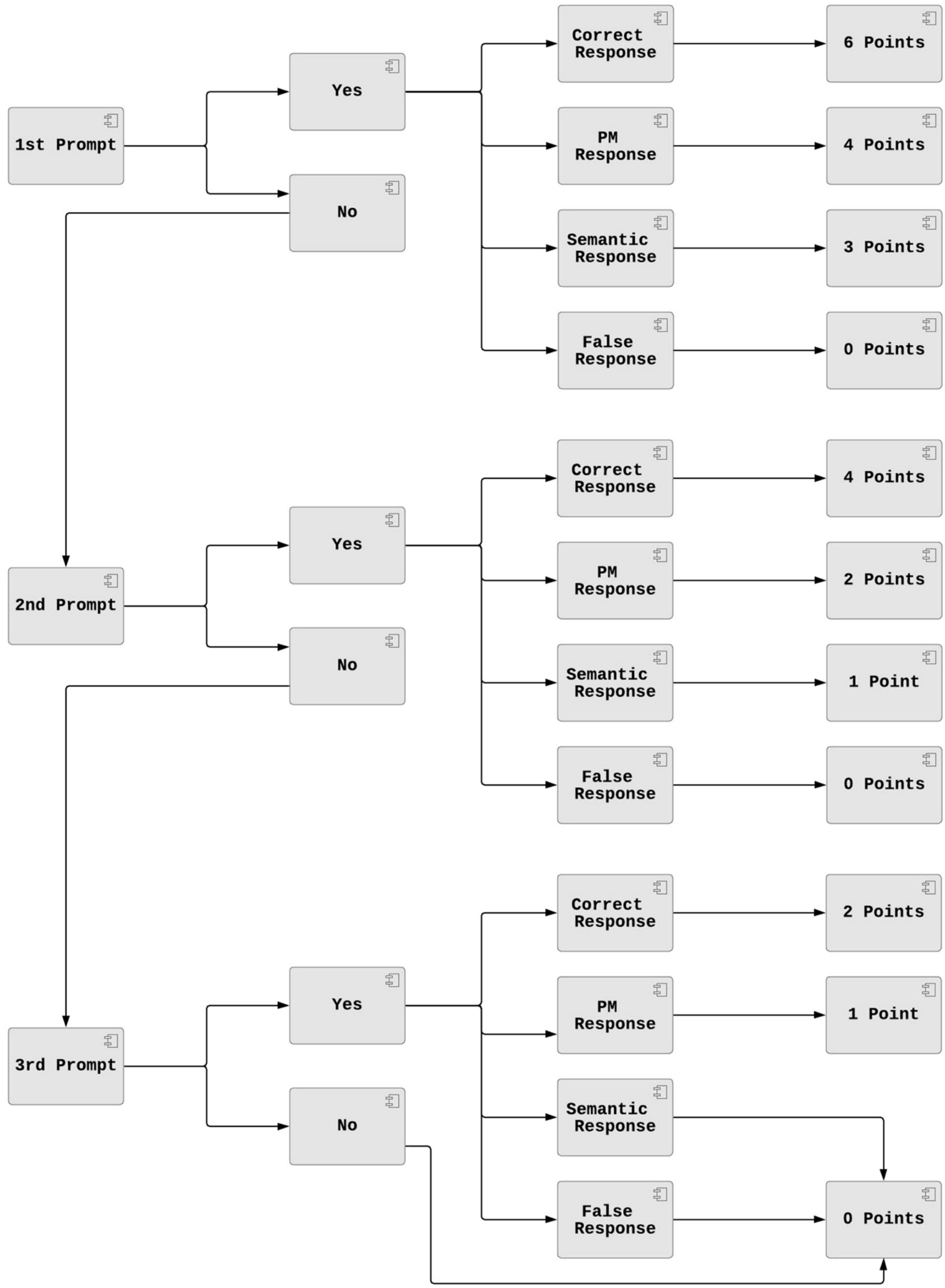


FIGURE 5 | Prospective memory: positive scoring system.

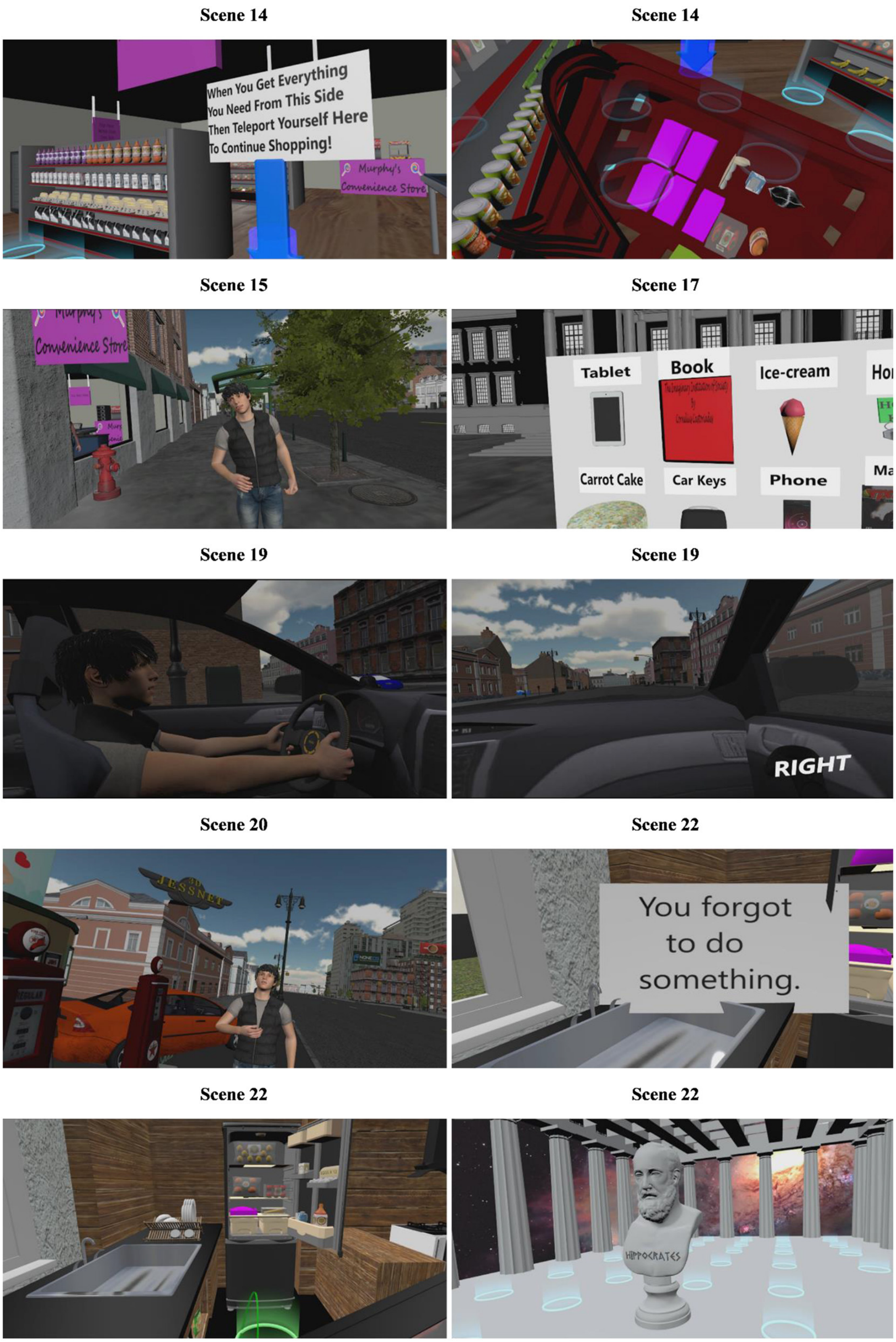


FIGURE 6 | VR-EAL storyline: scenes 14–22.

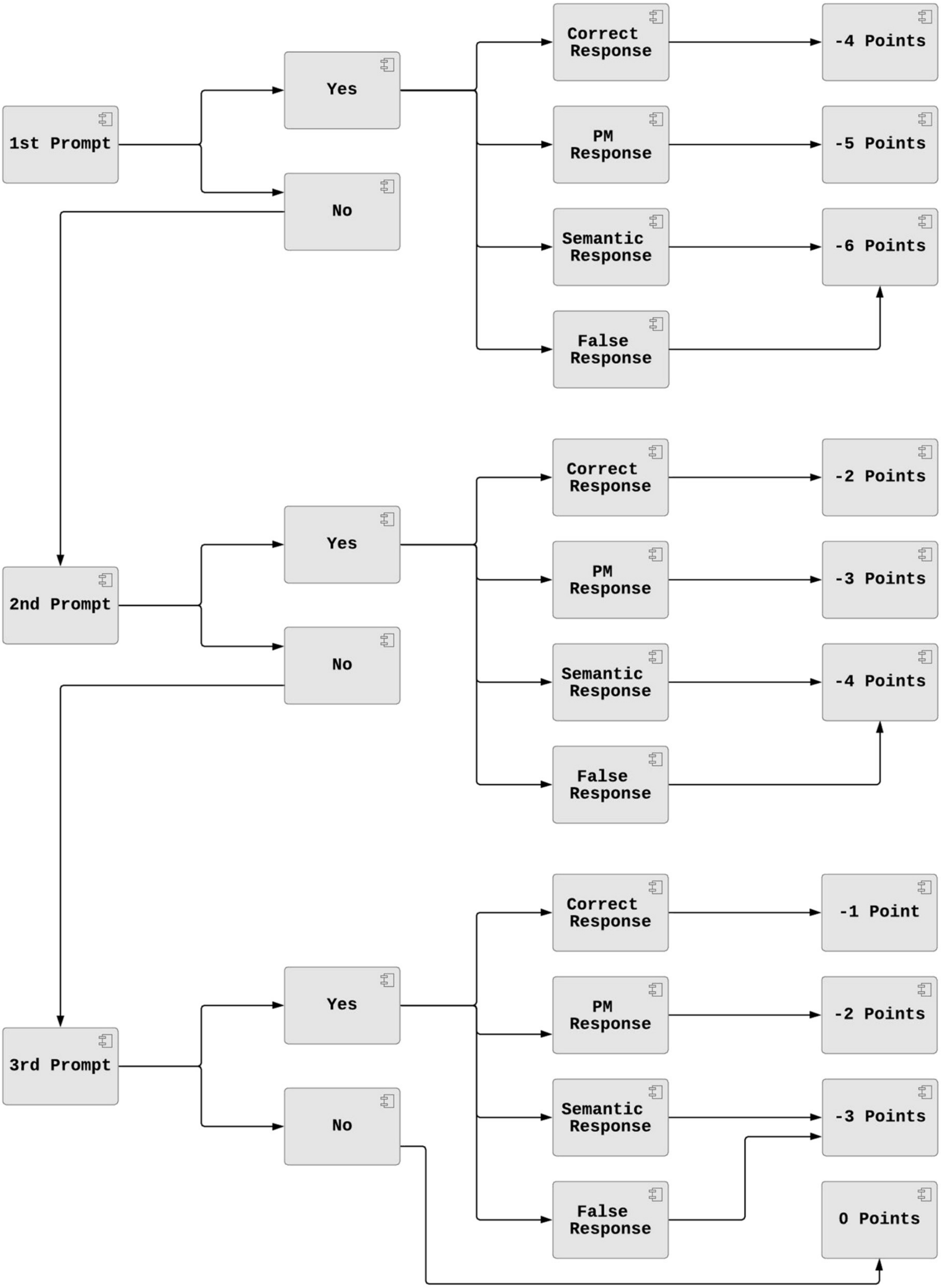


FIGURE 7 | Prospective memory: negative scoring system.

next scene, the user should again respond negatively to avoid a maximum of 3 points being deducted (see **Figure 7**). Therefore, in this task, up to 6 points may be subtracted. Then, the user arrives at the library (i.e., scene 17; see **Table 1**), where s/he has another interaction with Alex (i.e., a prospective memory task), which is performed and scored in the same way as the prospective memory task in scene 10 (see Garden and **Figure 5**). After leaving the library, Alex and the user proceed to the petrol station to refill the car.

On the road home: selective auditory attention and prospective memory tasks

The user is in the car with Alex and the radio station has another challenge (i.e., scene 19; see **Table 1** and **Figure 6**). This time small speakers playing different sounds have been placed on both sides of the street. The user should detect the target sounds and avoid the false high-pitched and low-pitched sounds while Alex drives along the street. As in the tutorial, the user presses the controller trigger when they detect a sound. If the user presses the trigger on the right controller to detect a target sound originating on the right side, then s/he gets 2 points. If the user presses the trigger on the left controller to detect a target sound originating on the left side, s/he also gets 2 points. If the user presses the trigger on the right controller to detect a target sound originating on the left side or a trigger on the left controller to detect a target sound originating on the right side, s/he gains only 1 point. On the other hand, if the user responds to a distractor sound, irrelevant of its origin or the controller used to respond, 1 point is deducted. The stored data summarize the number of detected sounds of each type (i.e., target sounds, low pitched distractor sounds, high pitched distractor sounds), the number of sounds detected on the left and right sides, and how many times the wrong controller (i.e., false side) was used to detect a target sound.

After the car ride, the user is at the petrol station with Alex (i.e., scene 20; see **Table 1**). The user has another conversation with Alex, where s/he receives false prompts (i.e., there is not a prospective memory task to perform). This prospective memory task is performed and scored in the same way as the Bakery prospective memory task (i.e., scene 20, see **Figure 7**). Then, the user returns back home with Alex (i.e., scene 21), where the user has their last interaction with Alex, and should give him the extra pair of keys to the flat. This prospective memory task is also performed and scored as the prospective memory task in scene 10 (i.e., see Garden and **Figure 5**).

Back Home: distractor and prospective memory task

In the final scene (i.e., scene 22), the user is back home, where s/he is required to perform two tasks (see **Figure 6**). The first task is a distractor task, where the user needs to put away the items that s/he has bought from the supermarket. While doing this, s/he needs to remember that s/he should take his/her medication at 1 p.m. If the user performs the task on time, then s/he receives 6 points. If the user fails to remember the prospective memory task after 70 s, a prompt appears. If the user performs the task after this first prompt, s/he receives 4 points. If the user ignores the first prompt, after another 10 s, a second prompt appears. If the user performs the task after the second prompt, s/he receives 2 points. If the user ignores the second prompt, after a further 10 s, a third and final prompt appears. If the user performs the task after the third prompt, s/he receives 1 point. If the user ignores the third prompt and presses the final button, s/he receives 0 points.

Once the user presses the final button, the scenario finishes and the credits appear. Here, the user is informed of the aims of VR-EAL. The VR-EAL attempts to be an extensive neuropsychological assessment of prospective memory, episodic memory, executive functions, and attentional processes by collecting various data pertinent to these cognitive functions (see **Supplementary Material I** for an example of VR-EAL data).

Development of VR Software in Unity

The scenario provides the main framework for developing the VR application. VR-EAL was developed to be compatible with the HTC Vive, HTC Vive Pro, Oculus Rift, and Oculus Rift-S to be aligned with the minimum hardware technological specifications for guaranteeing health and safety standards and the reliability of the data (Kourtesis et al., 2019a). The quality of VR-EAL was assessed in terms of user experience, game mechanics, in-game assistance, and VRISE using the Virtual Reality Neuroscience Questionnaire (VRNQ; Kourtesis et al., 2019b). The total duration for the VR neuropsychological assessment is ∼60 min, which falls within the suggested maximum duration range for VR sessions (Kourtesis et al., 2019b). Long VR sessions appear more susceptible to VRISE, though, long (50–70 min) VR sessions which exceed the parsimonious cut-offs from the VRNQ do not induce VRISE (Kourtesis et al., 2019b). For this reason, the parsimonious cut-offs for the VRNQ (see **Table 2**) will be used to ensure that VR-EAL users do not suffer from VRISE (Kourtesis et al., 2019b).

The development of VR-EAL should be proximal to commercial VR applications. The first step of the development is to select Unity's settings to support the development of VR software. For the development of VR-EAL, Unity version 2017.4.8f1 was used. Unity supports VR software development kits (SDK). The built-in support for the SDKs is for the OpenVR SDK and the Oculus SDK. In the player settings of Unity, the developer may select the VR/XR supported box, which allows the addition of the aforementioned SDKs. For VR-EAL, Unity's

TABLE 2 | VRNQ minimum and parsimonious cut-offs.

Score	Minimum cut-offs	Parsimonious cut-offs
User experience	≥25/35	≥30/35
Game mechanics	≥25/35	≥30/35
In-game assistance	≥25/35	≥30/35
VRISE	≥25/35	≥30/35
VRNQ Total Score	≥100/140	≥120/140

The median of each sub-score and total score should meet the suggested cut-offs to determine that the evaluated VR software is of adequate quality without any significant VRISE. The utilization of the parsimonious cut-offs more robustly supports the suitability of the VR software. Derived from Kourtesis et al. (2019b).

support for both the OpenVR SDK and the Oculus SDK were added, though, priority was given to the OpenVR SDK.

Navigation and Interactions

VR software for the cognitive sciences may require intensive movement and interactions. However, the development of such interactions demands highly advanced programming skills in C# and expertise in VR software development in Unity. Nonetheless, on Unity's asset store and GitHub's website, there are some effective alternatives that facilitate the implementation of intensive interactions without the requirement of highly advanced software development skills. The utilization of the SteamVR SDK, Oculus SDK, Virtual Reality Toolkit (VRTK) or similar toolkits and assets are options which should be considered. For the development of VR-EAL, the SteamVR SDK and VRTK were selected to develop accurate interactions compatible with the capabilities of the 6DoF controllers of HTC Vive and Oculus Rift. The advantage of SteamVR SDK, which was developed based on OpenVR SDK, is that is compatible with both the HTC Vive and Oculus Rift, though, it does not offer a wide variety of interactions or good quality physics. Nonetheless, the VRTK mounts the SteamVR SDK and offers better quality physics and plenty of interactions that support the development of VR research software for cognitive sciences.

A fundamental interaction in the VE is navigation. HTC Vive and Oculus Rift offer a play area of an acceptable size, which permits ecologically valid scenarios and interactions to be developed (Porcino et al., 2017; Borrego et al., 2018). However, the VR play area is restricted to the limits of the physical space and tracking area; hence, it does not allow navigation which is based on physically walking (Porcino et al., 2017; Borrego et al., 2018). A suitable solution is the implementation of a navigation system based on teleportation. Teleportation enables navigation exceeding the boundaries of the VR play area and delivers high-level immersion, a pleasant user experience, and decreases the frequency of VRISE. Typically, a navigation system of a VR software which depends on a touchpad, keyboard, or joystick, substantially increasing the frequency and intensity of VRISE (Bozgeyikli et al., 2016; Frommel et al., 2017; Porcino et al., 2017). In VR-EAL, a combination of teleportation and physical movement (i.e., free movement of the upper limbs and walking in a small-restricted area) is used (see **Figures 1–3**, **6**).

The VRTK provides scripts and tools that aid the developer to build a teleportation system. The VRTK is compatible with 6DoF controllers, which are necessary to provide naturalistic and ergonomic interactions. In addition, the implementation of 6DoF controllers facilitates familiarization with their controls and their utilization, because they imitate real life hand actions and movements (Sportillo et al., 2017; Figueiredo et al., 2018). The VR-EAL user learns the controls in the tutorials, though, there are also in-game instructions and aids that assist even a non-gamer user to grab, use, and manipulate items. These in-game assistance methods significantly alleviate the occurrence of VRISE, while increasing the user's level of enjoyment (Caputo et al., 2017; Porcino et al., 2017). Finally, the VRTK offers additional gamified interactions through the snap-drop-zones. The snap-drop-zones are essentially carriers of game objects and their mechanics are similar to the trigger-zones. For example, when a game object (i.e., a child object of a controller) enters the zone, if the game object is released (i.e., stops being a child object of the controller), this game object is attached to the snap-drop-zone (i.e., it becomes a child object). In VR-EAL, the snap-drop-zones are extensively used, allowing the scoring of tasks, which otherwise would be less effective in terms of accuracy of response times.

The interaction and navigation systems are essential to increase immersion. However, immersion depends on the strength of the placement, plausibility, and embodiment illusions (Slater, 2009; Slater et al., 2010; Maister et al., 2015; Pan and Hamilton, 2018). An ecologically valid neuropsychological assessment necessitates genuine responses from the user. Robust placement and plausibility illusions ensure that the user will genuinely perform the tasks as s/he would perform them in real life (Slater, 2009; Slater et al., 2010; Pan and Hamilton, 2018). The placement illusion is the deception of the user that s/he is in a real environment and not in a VE (Slater, 2009; Slater et al., 2010). However, the placement illusion is fragile because the VE should react to the user's actions (Slater, 2009; Slater et al., 2010). This is resolved by the plausibility illusion, which is the deception of the user that the environment reacts to his/her actions. Therefore, the user believes the plausibility of being in a real environment (Slater, 2009; Slater et al., 2010). The naturalistic interactions in the VE that VRTK and SteamVR SDK offer are pertinent to the plausibility illusion.

Graphics

A strong placement illusion relies on the quality of the graphics and 3D objects (Slater, 2009; Slater et al., 2010). Correspondingly, the quality of the graphics principally depends on the rendering (Lavoué and Mantiuk, 2015). The rendering comprises the in-game quality of the image (i.e., perceptual quality), and the omission of unnecessary visual information (i.e., occlusion culling) (Lavoué and Mantiuk, 2015). The advancement of these rendering aspects ameliorates both the quality of graphics and the performance of the VR software (Brennesholtz, 2018). Likewise, the amplified image refresh rate and resolution decrease the frequency and intensity of VRISE (Brennesholtz, 2018). However, the rendering pipeline and shaders in Unity are not optimized to meet VR standards. The VR software developer should select different rendering options, so the quality of graphics is good and the image's refresh rate is equal to or above 90 Hz, which is the minimum for high-end HMDs like the HTC Vive and Oculus Rift. For example, the "Lab renderer" is an asset that allows VR optimized rendering and replaces the common shaders with VR optimized ones. Additionally, the "Lab renderer" supports an extensive number of light sources (i.e., up to 15), which otherwise would not be feasible in VR. However, the developer needs to build a global illumination map (i.e., lightmap), which substantially alleviates the cost of lights and shadows on the software's performance (Jerald, 2015; LaViola et al., 2017; Sherman and Craig, 2018). Usually, the lightmapping process is the final step in the development process.

The acquisition of 3D objects may be expensive or time-consuming. However, there are several free 3D objects on Unity's asset store and webpages, such as TurboSquid and Cgtrader, which can be used for the development of VR research software. Importantly, the license for these 3D objects obliges the developer not to use them for commercial purposes. However, research VR software like VR-EAL is free, and research software developers usually do not commercialize their products. Although there are several free 3D objects on the websites mentioned above, it is likely that these 3D objects are not compatible with VR standards. In VR, the 3D objects should comprise a low number of polygons (Jerald, 2015; LaViola et al., 2017; Sherman and Craig, 2018). A decrease in polygons may be achieved using software like 3DS Max. The optimization of the 3D objects (to meet VR standards) may be achieved by simply importing the 3D objects, optimizing them, and then exporting them with a low number of polygons in a Unity compatible format (i.e., fbx and obj).

Nevertheless, developers often aim to create large VEs such as cities, towns, shops, and neighborhoods. Each 3D object, whether it be small (e.g., a pen), medium (e.g., a chair), or large (e.g., a building), may comprise several mesh renderers. Unity requires one batch (i.e., draw call) for each mesh renderer. In large environments, the batching may significantly lower the image's refresh rate and the overall performance of the software (Jerald, 2015; LaViola et al., 2017; Sherman and Craig, 2018). However, assets like MeshBaker are designed to solve this problem. MeshBaker merges all the selected textures and meshes into a clone game object with a small number of meshes and textures. For example, the town that was designed for VR-EAL required >1,000 draw calls. After the implementation of MeshBaker, the draw calls were decreased to 16. However, the disadvantage of MeshBaker is that it does not clone the colliders. Hence, the developer needs to deactivate the mesh renderers of the original game object(s) and leave active all the colliders, while the original game object(s) should be precisely in the same position with the clone(s) so the colliders of the former are aligned with the meshes of the latter. Of note, MeshBaker should be purchased from Unity's asset store in contrast with the other assets used in VR-EAL's development which are freely available (i.e., SteamVR SDK, VRTK, and Lab renderer).

Sound

Another important aspect of VR software development is the quality of the sound. The addition of spatialized sounds in the VE (e.g., ambient and feedback sounds) augments the level of immersion and enjoyment (Vorländer and Shinn-Cunningham, 2014), and significantly reduces the frequency of VRISE (Viirre et al., 2014). Spatialized sounds in VR assist the user to orient and navigate (Rumiński, 2015), and enhance the geometry of the VE without reducing the software's performance (Kobayashi et al., 2015). In Unity, a developer may use tools like SteamAudio, Oculus Audio Spatializer, or Microsoft Audio Spatializer for good quality and spatialization of the audio aspects. In VR-EAL's development, Steam Audio was used. SteamAudio spatializes the sound to the location of the audio source's location and improves the reverberance of sounds (i.e., Unity's reverb zone). Notably, the strength of the plausibility illusion is analogous to the sensorimotor contingency, which is the integration of the senses (i.e., motion, vision, touch, smell, taste) (Gonzalez-Franco and Lanier, 2017). Moreover, the VRTK enables the utilization of a haptic modality. For example, when the user grabs an item in the VE, s/he expects a tactile sense as would be experienced in real life. The haptic feedback of the VRTK allows the developer to activate/deactivate the vibration system of the 6 DoF controllers when an event occurs (e.g., grabbing or releasing a game object) and define the strength and the duration of the vibration. The spatialized audio and the haptics additionally reinforce the plausibility illusion by providing an expected auditory and haptic feedback to the user (Jerald, 2015; LaViola et al., 2017; Sherman and Craig, 2018).

3D Characters

Furthermore, VR research software like VR-EAL, which includes social interactions with virtual characters, should also consider the quality of the 3D characters in terms of realistic appearance and behavior. For example, Morph 3D and Mixamo both offer free and low-cost realistic 3D characters that may be used in VR software development. For VR-EAL, Morph 3D was preferred, though, other virtual humans from Unity's asset store were used to populate the scenes (e.g., individuals waiting for the bus at the bus stop). The 3D characters provided by Morph 3D have modifiable features, which may be used by the developer to customize the character's appearance (e.g., body size) and expressions (e.g., facial expressions which signify emotions such as happiness and sadness). Morph 3D provides two free 3D characters (i.e., female and male) capable of displaying naturalistic behavior (i.e., body and facial animations). The developer may use body animations which derive from motion capture (MoCap) techniques. For the development of VR-EAL, body animations were derived from free sample animations from Unity's Asset Store (e.g., hand movement during talking, and waving) and the MoCap animations library of the Carnegie Mellon University. However, the effective implementation of the animations requires modification and synchronization (e.g., the animation should be adjusted to the length of the 3D character's interaction) using Unity's animation and the animator's windows. The animation window may be used for synchronization, while the animator is a state machine controller that controls the transition between animations (e.g., when this event happens, play this animation, or when animation X ends, play animation Y).

However, the most challenging aspect of realistic 3D characters is the animation of their facial features. The 3D character should have realistic eye interactions (i.e., blinking, looking at or away from the user) and talking (i.e., a realistic voice and synchronized lip movements). Limitations in both time and resources did not allow for seamless face and body animations since that would require multimillion dollars' worth of equipment like those used by big game studios. This limitation can result in an uncanny valley effect (Seyama and Nagayama, 2007; Mori et al., 2012). However, previous research has shown that, when users interact with 3D humanoid embodied agents that have the role of an instructor (like the ones used in VR-EAL), they have less expectations for that character due to their role and limited interactivity (Korre, 2019). The addition of 3D characters was important because they deliver an interaction metaphor resembling human-to-human interactions (Korre, 2019). Even though adding a 3D character in the scene can introduce biases, the illusion of humanness—which is defined as the user's notion that the system (in this case the 3D NPC) possesses human attributes and/or cognitive functions—has been found to increase usability (Korre, 2019).

Realistic voices may be established by employing voice actors to produce the script. However, the employment of temporary staff increases development costs. For VR-EAL, text-to-speech technologies were used as an alternative solution to deliver realistic voices. Balabolka software was used in conjunction with Ivona3D Voices (n.b., Ivona3D has been replaced by Amazon Polly). Balabolka is an IDE for text-to-speech which allows further manipulation of voices (i.e., pitch, rate, and volume), while Ivona3D provides realistic voices. The developer types or pastes the text into Balabolka, Balabolka modifies it with respect to the desired outcome (e.g., high-pitched or low-pitched voice) and exports the file in a.wav format. Additionally, free software like Audacity may be used, which offers greater variety in sound modifications. The second crucial part is to synchronize the eyes and lip movements with the voice clips and body animations. There are assets on Unity's asset store that may be used to achieve this desired outcome. In VR-EAL, Salsa3D and RandomEyes3D were used to attain good quality facial animations and lip synchronization. Salsa3D synchronizes the lips with the voice clip, while RandomEyes3D allows the developer to control the proportion of eye contact with the user for each voice clip.

Summary of the VR-EAL Illusions

Summing up, the described VR-EAL development process facilitates the utilization of ergonomic interactions, a VR compatible navigation system, good quality graphics, haptics, and sound, as well as social interactions with realistic 3D characters. These software features contribute to the lessening or avoidance of VRISE and augmentation of the level of immersion by providing placement and plausibility illusions. However, VR-EAL does not seem to deliver a strong embodiment illusion (i.e., the deception that the user owns the body of the virtual avatar), because it only relies on the presence of the 6 DoF controllers. A possible solution would be the implementation of inverse kinematics, which animates the virtual avatar with respect to the user's movements. In addition, the temporal illusion (i.e., deceiving the user into thinking that the virtual time is real-time) only relies on changes in environmental cues (e.g., the movement of the sun, and changes in lighting). Therefore, a VR digital watch was developed (freely distributed on GitHub) and used in an attempt to increase the strength of the temporal illusion. To conclude, the development of VR research software is feasible mainly using free or low-cost assets from GitHub, Unity Asset's store, and other webpages. However, the suitability and quality of the VR software should be evaluated before its implementation in research settings.

EVALUATION OF VR-EAL

Participants

Twenty-five participants (six female gamers, six male gamers, seven female non-gamers, and six male non-gamers) were recruited for the study via the internal email network of University of Edinburgh as well as social media. The mean age of the participants was 30.80 years (SD = 5.56, range = 20–45) and the mean years of full-time education was 14.20 years (SD = 1.60, range = 12–16). Twelve participants (three female gamers, three male gamers, three female non-gamers, and three male non-gamers; mean age = 30.67 years, SD = 2.87, range = 26–36; mean educational level = 14.75 years, SD = 1.30, range = 12–16 years) attended all three VR sessions (i.e., alpha, beta, and final versions), while the remaining 13 participants only attended the final version session. The gamer experience was a dichotomous variable (i.e., gamer or non-gamer) based on the participants' response to a question asking whether they played games on a weekly basis. The current study has been approved by the Philosophy, Psychology and Language Sciences Research Ethics Committee of the University of Edinburgh. All participants were informed about the procedures, possible adverse effects (e.g., VRISE), data utilization, and the general aims of the study both orally and in writing; subsequently, every participant gave written informed consent.

Material

Hardware and Software

An HTC Vive HMD, two lighthouse-stations for motion tracking, and two 6 DoF controllers were used. The HMD was connected to a laptop with a 2.80 GHz Intel Core i7 7700HQ processor, 16 GB RAM, a 4,095 MB NVIDIA GeForce GTX 1070 graphics card, a 931 GB TOSHIBA MQ01ABD100 (SATA) hard disk, and Realtek High Definition Audio. The size of the VR play area was 4.4 m^2. The software was the alpha version of VR-EAL for session 1, the beta version of VR-EAL for session 2, and the final version of VR-EAL for session 3.

VRNQ

The VRNQ is a paper-and-pencil questionnaire containing 20 questions, where each question refers to one of the criteria necessary to assess VR research/clinical software in neuroscience (Kourtesis et al., 2019b). The 20 questions assess four domains: user experience, game mechanics, in-game assistance, and VRISE. The VRNQ has a maximum total score of 140, and 35 for each domain. VRNQ responses are indicated on a 7-point Likert style scale ranging from 1= extremely low to 7 = extremely high. Higher scores indicate a more positive outcome; this also applies to the evaluation of VRISE intensity. Hence, higher VRISE scores indicate lower intensities of VRISE (i.e., 1 = extremely intense feeling, 2 = very intense feeling, 3 = intense feeling, 4 = moderate feeling, 5 = mild feeling, 6 = very mild feeling, 7 = absent). Additionally, the VRNQ allows participants to provide qualitative feedback, which may be useful during the development process. Lastly, the VRNQ has two cut-off scores, the minimum (i.e., 25 for every sub-score, and 100 for the total score) and parsimonious (i.e., 30 for every sub-score, and 120 for the total score) cut-offs. The median scores derived from the user sample should exceed at least the minimum cut-offs, while for VR software which requires long VR sessions, then the parsimonious cut-offs should be preferred. For the evaluation of VR-EAL, the parsimonious cut-offs were opted to support the suitability of VR-EAL. The VRNQ can be downloaded from **Supplementary Material II**.

Procedures

Twelve participants attended all three VR sessions, while an additional 13 participants only attended the third session. The period between each session was 6–8 weeks. In each session, participants were immersed in a different version of VR-EAL. Each session began with inductions in VR-EAL, the HTC Vive, and the 6 DoF controller. Then, participants played a version of VR-EAL. Lastly, after the completion of VR-EAL, participants were asked to complete the VRNQ. A preview of the final version of VR-EAL can be found in **Supplementary Material III** or by following the hyperlink: https://www.youtube.com/watch?v=IHEIvS37Xy8andt=.

Statistical Analysis

Bayesian statistics were preferred over null hypothesis significance testing (NHST). *P*-values calculate the distance (i.e., the difference) between the data and the null hypothesis (H0) (Cox and Donnelly, 2011; Held and Ott, 2018). The *p*-values assess the assumption of no difference or no effect, while the Bayesian factor (BF_{10}) converts *p*-values into evidence in favor of the alternative hypothesis (H1) against the H0 (Cox and Donnelly, 2011; Held and Ott, 2018). BF_{10} is found robustly more parsimonious than the *p*-value in evaluating the evidence against the H0 (Cox and Donnelly, 2011; Held and Ott, 2018; Wagenmakers et al., 2018a,b). Importantly, the difference between BF_{10} and *p*-values is even greater (in favor of BF_{10}) in small sample sizes, where BF_{10} should be opted for as it is more parsimonious (Held and Ott, 2018; Wagenmakers et al., 2018a,b). For these reasons, the BF_{10} was preferred instead of *p*-values for the assessment of statistical inference, especially while having a relatively small sample size. Moreover, a larger BF_{10} postulates more evidence in support of H1 (Cox and Donnelly, 2011; Marsman and Wagenmakers, 2017; Held and Ott, 2018; Wagenmakers et al., 2018a,b). Specifically, a $BF_{10} \leq 1$ indicates no evidence in favor of H1, while $1 < BF_{10} < 3$ indicates anecdotal evidence for H1, $3 \leq BF_{10} < 10$ indicates moderate evidence for H1, $10 \leq BF_{10} < 30$ indicates strong evidence for H1, $30 \leq BF_{10} < 100$ indicates very strong evidence for H1, and a $BF_{10} \geq 100$ indicates extreme evidence for H1 (Marsman and Wagenmakers, 2017; Wagenmakers et al., 2018a,b). For our analyses, we accept the notion put forward by Marsman and Wagenmakers (2017), Wagenmakers et al. (2018a,b) of $BF_{10} \leq 1$ indicating no evidence in favor of H1, $BF_{10} > 3$ indicating moderate evidence in favor of H1, $BF_{10} \geq 10$ indicating strong evidence for H1, and $BF_{10} \geq 100$ indicating extreme evidence for H1. In this study, a parsimonious threshold of $BF_{10} \geq 10$ was set for statistical inference, which postulates strong evidence in favor of the H1 (Marsman and Wagenmakers, 2017; Wagenmakers et al., 2018a,b), and corresponds to a $p < 0.01$ (e.g., $BF_{10} = 10$) or to a $p < 0.001$ (e.g., $BF_{10} > 11$) (Cox and Donnelly, 2011; Held and Ott, 2018). However, we report both BF_{10} and *p*-values in this study. A Bayesian paired samples *t*-test was performed to compare the VRNQ results for each version of VR-EAL ($N = 12$), as well as to inspect potential differences between gamers ($N = 12$) and non-gamers ($N = 13$). The Bayesian statistical analyses were performed using JASP (Version 0.8.1.2) (JASP Team, 2017).

Results

There was not a significant difference between gamers and non-gamers in VRNQ scores (see **Table 3**). The final version of VR-EAL exceeded the parsimonious cut-off for the VRNQ total score, while the alpha and beta versions of VR-EAL did not (see **Table 4**). Notably, the VRNQ sub-scores of the final version of VR-EAL also exceeded the parsimonious VRNQ cut-offs (see **Table 4**), while the average duration of the VR sessions (i.e., duration of being immersed) was 62.2 min (SD = 5.59) across the 25 participants. The beta version of VR-EAL approached the cut-offs for user experience and game mechanics; however, it was substantially below the cut-offs for in-game assistance and VRISE. The alpha version of VR-EAL was significantly below the cut-offs for every sub-score of VRNQ.

According to the adopted nomenclature (i.e., $BF_{10} \leq 1$ indicating no evidence in favor of H1, $BF_{10} > 3$ indicating moderate evidence in favor of H1, $BF_{10} \geq 10$ for H1, and $BF_{10} \geq 100$ indicating extreme evidence for H1) by Marsman and Wagenmakers (2017) and Wagenmakers et al. (2018a,b), the Bayesian *t*-test analysis ($N = 12$) demonstrated significant differences in the VRNQ scores between the final, beta, and alpha versions of the VR-EAL (see **Table 5**). We observed that the probability of the alternative hypothesis that the VRNQ total score for the final version is greater than the VRNQ total score for the alpha version is 57,794 times greater (i.e., $BF_{10} =$ 57,974; see **Table 5**) than the probability of H0 (i.e., not being greater). Similarly, the probability of the alternative hypothesis that the VRNQ total score for the final version is greater than the VRNQ total score for the beta version is 855 times greater (i.e., $BF_{10} = 855$; see **Table 5**) than the probability of H0. Lastly, the probability of the alternative hypothesis that the VRNQ total score for the beta version is greater than the VRNQ total score for the alpha version is 101 times greater (i.e., $BF_{10} = 101$; see **Table 5**) than the probability of H0. The remaining alternative hypotheses for the comparisons between the versions of VR-EAL

TABLE 3 | Comparison of VRNQ scores between gamers and non-gamers.

VRNQ scores	*p*-value	BF_{10}	Error %
Total VRNQ	$p = 0.631$	0.402	1.052e−4
User experience	$p = 0.289$	0.546	0.001
Game mechanics	$p = 0.459$	0.429	2.003e−4
In-game assistance	$p = 0.841$	0.374	0.030
VRISE	$p = 0.983$	0.368	0.030

$^*BF_{10} > 10$; $^{**}BF_{10} > 30$; $^{***}BF_{10} > 100$; No significant differences observed.

TABLE 4 | VRNQ scores for alpha, beta, and final version of VR-EAL.

	N	Median (MAD)	Cut-off	Maximum score
Total VRNQ—alpha version	12	100 (6)	≥120	140
User experience—alpha version	12	25 (2)	≥30	35
Game mechanics—alpha version	12	23.5 (3.5)	≥30	35
In-game assistance—alpha version	12	24 (3)	≥30	35
VRISE—alpha version	12	25.5 (1.5)	≥30	35
Total VRNQ—beta version	12	109.5 (2.5)	≥120	140
User experience—beta version	12	28 (1)	≥30	35
Game mechanics—beta version	12	29 (1)	≥30	35
In-game assistance—beta version	12	26 (1)	≥30	35
VRISE—beta version	12	26 (1)	≥30	35
Total VRNQ—final version—all	25	128 (5)	≥120	140
User experience—final version—all	25	31 (2)	≥30	35
Game mechanics—final version—all	25	32 (2)	≥30	35
In-game assistance—final version—all	25	32 (3)	≥30	35
VRISE—final version—all	25	33 (1)	≥30	35
Total VRNQ—final version—gamers	12	129.5 (5)	≥120	140
User experience—final version—gamers	12	32.5 (1.5)	≥30	35
Game mechanics—final version—gamers	12	32 (1.5)	≥30	35
In-game assistance—final version—gamers	12	32.5 (2)	≥30	35
VRISE—final version—gamers	12	33 (1)	≥30	35
Total VRNQ—final version—non-gamers	13	128 (4)	≥120	140
User experience—final version—non-gamers	13	31 (1)	≥30	35
Game mechanics—final version—non-gamers	13	31 (2)	≥30	35
In-game assistance—final version—non-gamers	13	32 (3)	≥30	35
VRISE—final version—non-gamers	13	33 (2)	≥30	35

MAD, Median Absolute Deviation.

TABLE 5 | Bayesian paired sample *t*-test results.

Alternative Hypothesis (H1)	*p*-value	BF_{10}	Error %
Total VRNQ—alpha < Total VRNQ—beta	$p < 0.001$	101.651***	~ 2.226e–5
Total VRNQ-alpha < Total VRNQ-final	$p < 0.001$	57974.267***	~ 9.361e-35
Total VRNQ-beta < Total VRNQ-final	$p < 0.001$	855.603***	~ 1.506e-17
User experience-alpha < User experience-beta	$p < 0.001$	21.221*	~ 9.875e–5
User experience-alpha < User experience-final	$p < 0.001$	681.518***	~ 8.429e-24
User experience-beta < User experience-final	$p < 0.001$	17.597*	~ 2.172e–4
Game mechanics-alpha < Game mechanics-beta	$p < 0.001$	47.214**	~ 1.820e–4
Game mechanics-alpha<Game mechanics-final	$p < 0.001$	487.798***	~ 2.337e-19
Game mechanics-beta <Game mechanics-final	$p < 0.001$	17.262*	~ 2.288e–4
In-game assistance-alpha < In-game assistance-beta	$p = 0.098$	1.095	~ 9.459e–4
In-game assistance-alpha <In-game assistance-final	$p < 0.001$	224.329***	~ 1.110e-18
In-game assistance-beta < In-game assistance-final	$p < 0.001$	139.994***	~ 5.188e–5
vrise-alpha < vrise-beta	$p = 0.111$	0.988	~ 0.001
VRISE-alpha <VRISE-final	$p < 0.001$	1912.328***	~ 3.643e-24
VRISE-beta < VRISE-final	$p < 0.001$	1277.335***	~ 7.819e-21

$^*BF_{10} > 10$; $^{**}BF_{10} > 30$; $^{***}BF_{10} > 100$; Alpha, Alpha version of VR-EAL; Beta, Beta version of VR-EAL; Final, Final Version of VR-EAL.

and their probabilities against the corresponding null hypotheses are displayed in **Table 5**.

Moreover, the final version was substantially better than the alpha version in terms of every sub-score and total score of the VRNQ. The beta version was better than the alpha version in terms of the VRNQ total score as well as the user experience and game mechanics sub-scores. However, there was not a significant difference between the VRNQ in terms of the VRISE or in-game assistance sub-scores. Moreover, the final version was also significantly improved compared to the beta version in terms of the VRNQ total score and all sub-scores. Though, the difference between them was smaller in the game mechanics and user experience sub-scores (see **Table 5**). Importantly, in the final version of the VR-EAL, all users ($N = 25$) experienced mild (i.e., five in VRNQ) to no VRISE (i.e., seven in VRNQ), while the vast majority ($N = 22$) experienced very mild (i.e., six in VRNQ) to no VRISE (see **Figure 8**).

DISCUSSION

The VR-EAL Versions

The present study attempted to develop a cost-effective VR research/clinical software (i.e., VR-EAL) of a high enough quality for implementation in cognitive studies and that does not induce VRISE. The development included three versions of VR-EAL (i.e., alpha, beta, and final) until the attainment of these desired outcomes. The alpha version of VR-EAL revealed several limitations. It had low frames per second (fps), which increased the frequency and the intensity of VRISE. Also, the alpha version did not include haptics during the interactions, and the in-game assistance props were low in number. Lastly, the shaders of the 3D models were not converted to VR shaders (i.e., the function

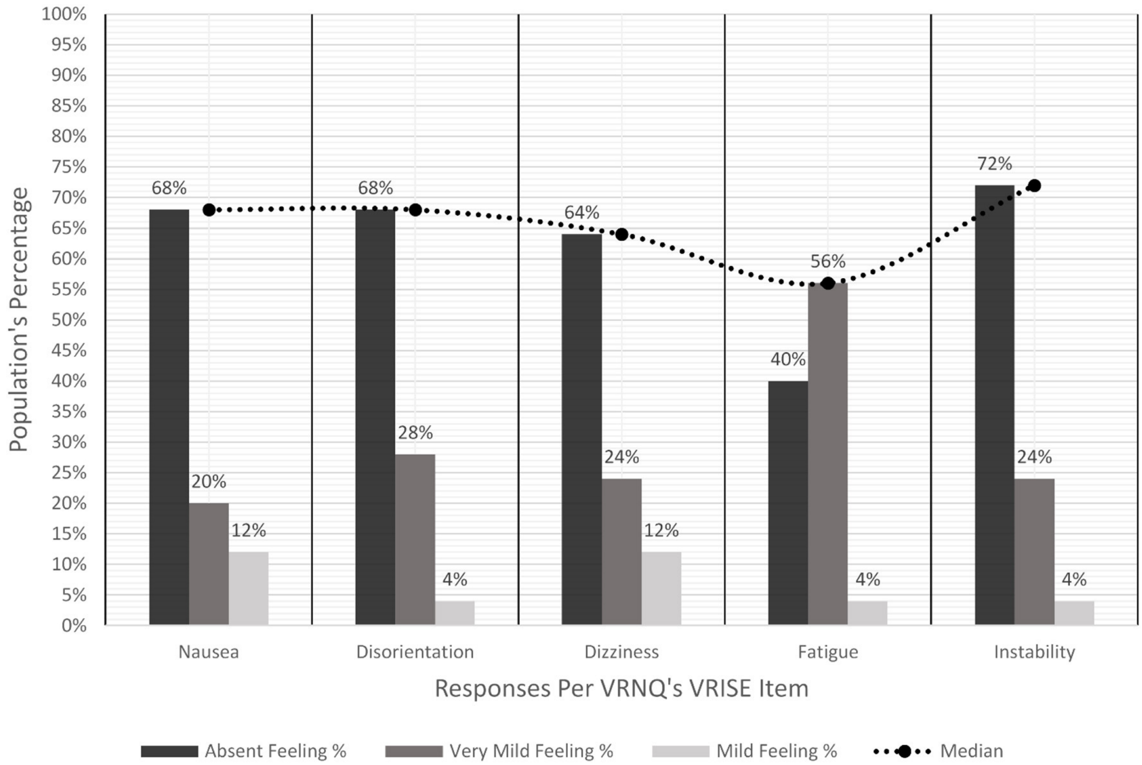


FIGURE 8 | VRISE in the final version of VR-EAL.

of the Lab renderer) and numerous game objects were defined as non-static. As a result, the quality of the graphics was below average, and the fps were substantially below 90 (i.e., 70–80) which is the lowest threshold for VR software targeting high-end HMDs such as HTC Vive and Oculus Rift. However, the feedback also confirmed that several game mechanics and approaches (e.g., tutorials) were in the right direction, which was encouraging for further VR-EAL development.

The principal improvements in the beta version of VR-EAL were pertinent to the alpha version's shortcomings. The shaders for all the game objects were converted to VR shaders, and several game objects, with which the user does not interact, were defined as static. The fps for the beta version were above 90, though, there were various points where the fps dropped for a couple of seconds. Although these fps drops were brief, their existence negatively affected the users who reported moderate to intense VRISE. Nonetheless, the beta version provided haptic and visual (i.e., highlighters) feedback to the users during the interactions, which further improved the quality of the game mechanics. In addition, the number of in-game aids was dramatically increased (e.g., more signs, labels, and directional arrows) and the duration of the tutorials was substantially prolonged (i.e., the inclusion of more explicit descriptions), which improved the quality of the users' experience. However, while the beta version was an improvement, it still failed to meet the parsimonious cut-offs of the VRNQ.

In the final version of VR-EAL, further improvements were conducted. The programming scripts of VR-EAL were re-assessed and correspondingly refined. Various chunks of code were expressed more compactly. For example, part of the code which had several Boolean values and/or float numbers were replaced by events and delegates (i.e., the features of object-oriented programming languages like C# that have substantially lower costs toward the performance of the software). Furthermore, the lightmapping of the 3D environments of scenes was upgraded by calculating high-resolution lightmaps instead of the medium resolution used in previous versions of VR-EAL. Redundant shadows were also deactivated to improve the performance of VR-EAL without degrading the quality of the graphics.

Moreover, major parts of the 3D environments were baked together (i.e., merged) through the implementation of MeshBaker's predominant functions to significantly reduce the draw-calls of VR-EAL. Interestingly, the result was a stable number of fps during gameplay. Specifically, the final version of VR-EAL has 120–140 fps during gameplay. Lastly, there was an improvement and enrichment of in-game assistance. In the tutorials, video screens and videos were added, which show the user how to use the controllers and perform each task. This visual and procedural demonstration allowed users to learn the respective controls and task trials faster and more effectively. This audio-visual demonstration using videos is feasible in VR since it can integrate the benefits of all mediums (e.g., video, audio, audio-visual). Furthermore, in the storyline scenes, where the user performs the actual tasks, several visual aids were added to provide additional guidance and alleviate confusion (see **Figures 3**, **6**).

Our results demonstrated that the VRNQ total and sub-scores exceeded the parsimonious cut-offs of the VRNQ for the final VR-EAL version. The improvements pertinent to graphics substantially increased the quality of the user experience, while they almost eradicated VRISE (see **Figure 8**). This substantial decrease of VRISE also highlights the importance of fps in VR. A developer should use the Unity profiler to check whether the VR software has a steady number of fps during gameplay, which the HMD requires. Also, the final version of VR-EAL appeared to have better in-game assistance and game mechanics. However, there was not any upgrade pertinent to the game mechanics. The increase in the evaluation of the game mechanics probably resulted due to the addition and improvement of in-game aids in both tutorial and storyline scenes. This finding also supports that in-game assistance has a paramount role in VR software. This is especially the case when the software is developed for clinical or research purposes, where the users could be either gamers or non-gamers. The quality of the tutorials and in-game aids should be cautiously designed to ensure the usability of the VR research software. To sum up, the final version of VR-EAL seems to deliver a pleasant testing experience and without the presence of significant VRISE.

VR Software Development in Cognitive Sciences

The current study demonstrated the procedure for the development of immersive VR research/clinical software (i.e., VR-EAL) with strong placement and plausibility illusions, which are necessary for collecting genuine responses (i.e., ecological valid) from users (Slater, 2009; Slater et al., 2010; Maister et al., 2015; Pan and Hamilton, 2018). The implementation of good quality 3D models (e.g., objects, buildings, and artificial humans) in conjunction with optimization tools (e.g., Lab Renderer and MeshBaker) facilitated an analogous placement illusion. Also, VR-EAL incorporates naturalistic and ergonomic interactions with the VE facilitated by the VR hardware (e.g., HTC Vive and 6 DoF controllers), SDKs (e.g., SteamVR and VRTK), and Unity assets pertinent to spatialized audio (e.g., Steam Audio) and artificial characters' animations (e.g., Salsa3D). These naturalistic and ergonomic interactions with the VE are capable of inducing a robust plausibility illusion.

Furthermore, a predominant concern for the implementation of VR in cognitive sciences is the presence of VRISE (Bohil et al., 2011; de França and Soares, 2017; Palmisano et al., 2017), which may compromise health and safety standards (Parsons et al., 2018), as well as the reliability of cognitive (Nalivaiko et al., 2015), physiological (Nalivaiko et al., 2015), and neuroimaging data (Arafat et al., 2018; Gavgani et al., 2018). Equally, the high cost of VR software development may additionally deter the adoption of VR as a research tool in cognitive sciences (Slater, 2018). However, the development of VR-EAL provides evidence that the obstacles above can be surpassed to implement VR software in cognitive sciences effectively.

The users of the final version of VR-EAL reported mild to no VRISE, with the average value in the VRISE sub-score being very mild to no VRISE. Importantly, these reports were offered by the users after spending around 60 min uninterrupted in VR. Typically, VRISE are intensified in longer VR sessions (Sharples et al., 2008). However, the utilization of the parsimonious cut-offs from the VRNQ guaranteed the significant alleviation of VRISE, which was also supported by the users' reports. Notably, the results of this study are in line with our previous work (Kourtesis et al., 2019b), where the gaming experience (i.e., gamer or non-gamer) did not affect the responses on the VRNQ. Also, the results support that the gaming experience does not affect the presence or intensity of VRISE in software of adequate quality. Therefore, VR software with technical features similar to VR-EAL would be suitable for implementation in cognitive sciences.

Cognitive scientists already implement computational approaches to investigate cognitive functions at the neuronal and cellular level (Sejnowski et al., 1988; Farrell and Lewandowsky, 2010; Kriegeskorte and Douglas, 2018), develop computerized neuropsychological tasks compatible with neuroimaging techniques (Peirce, 2007, 2009; Mathôt et al., 2012), as well as conducting flexible statistical analyses and creating high-quality graphics and simulations (Culpepper and Aguinis, 2011; Revelle, 2011; Stevens, 2017). The development of VR-EAL was achieved by using C# and Unity packages (i.e., SteamVR SDK, VRTK, Lab renderer, MeshBaker, Salsa3D, RandomEyes3D, 3D models, 3D environments, and 3D characters) on the Unity game engine, which is a user-friendly IDE equivalent to OpenSesame, PsychoPy, and MATLAB.

The majority of these Unity packages are cost-free, while the remainder are relatively low-cost, and could be used in future VR software development. Also, the acquisition of VR development skills by cognitive scientists with a background in either psychology or computers science can be realized in a moderately short period. Although, collaboration with a psychologist who has the required knowledge and clinical experience is crucial for a computer scientist with VR skills. Likewise, psychologists should either collaborate with a computer scientist with VR expertise or acquire VR development skills themselves. For the acquisition of VR skills by a computer scientist or a psychologist, there are online and on-campus interdisciplinary modules (e.g., Unity tutorials and documentation, game development courses, programming workshops, and specializations in VR) which further support the feasibility of acquiring the necessary skills. However, training cognitive scientists in VR software development should be prioritized for institutions which aspire to implement VR technologies in their studies. To summarize, this study demonstrated that the development of usable VR research software by a cognitive scientist is viable.

Limitations and Future Studies

This study, however, has some limitations. The implementation of novel technologies may result in more positive responses toward them (Wells et al., 2010). A future replication of the current results would elucidate this issue. Also, the study did not provide validation of VR-EAL as a neuropsychological tool. Future work will consider validating the VR-EAL against traditional paper-and-pencil and computerized tests of prospective memory, executive function, episodic memory, and attentional processes. A future validation study should also include a larger and more diverse population than the sample in this study. Regarding the quality of VR-EAL, it is not able to induce a strong embodiment illusion. The future version of the VR-EAL should include a VR avatar that corresponds to the user's movements and actions. Also, the integration of better 3D models, environments, and characters may be beneficial, which will additionally improve the quality of placement illusion and the user's experience. Finally, since VR-EAL is ultimately intended for implementation in cognitive neuroscience and neuropsychology, the future version of VR-EAL should include compatibility with eye-tracking measurements and neuroimaging techniques (e.g., event-related potentials measured by electroencephalography).

Conclusion

This study provided guidelines for the development of immersive VR research software that can be implemented in cognitive sciences to improve the ecological validity of the cognitive tasks and automate the administration and scoring of the neuropsychological assessment. The results substantially support the feasibility of the development of low-cost and effective immersive VR software without the presence of VRISE during a 60 min VR session by cognitive scientists who have skills in VR software development. Technologically competent cognitive scientists are able to develop cost-effective immersive VR research software that guarantees the safety of the users and the reliability of the collected data (i.e., neuropsychological, physiological, and neuroimaging data).

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Philosophy, Psychology and Language Sciences Research Ethics Committee of the University of Edinburgh. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

PK was the developer of VR-EAL. VR-EAL can be used by a third party by contacting the PK. PK had the initial idea and contributed to every aspect of this study. DK, SC, LD, and SM contributed to the methodological aspects and the discussion of the results.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomp.2019.00012/full#supplementary-material

Supplementary Material III | A brief preview of VR-EAL.

REFERENCES

Anderson, F. T., McDaniel, M. A., and Einstein, G. O. (2017). Remembering to remember: an examination of the cognitive processes underlying prospective memory. *Learn. Mem.* 2, 451–463. doi: 10.1016/B978-0-12-809324-5.21049-3

Arafat, I. M., Ferdous, S. M. S., and Quarles, J. (2018). "Cybersickness-provoking virtual reality alters brain signals of persons with multiple sclerosis," In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Reutlingen: IEEE), 1–120. doi: 10.1109/VR.2018.8446194

Bohil, C. J., Alicea, B., and Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nat. Rev. Neurosci.* 12:752. doi: 10.1038/nrn3122

Borrego, A., Latorre, J., Alcañiz, M., and Llorens, R. (2018). Comparison of oculus rift and HTC vive: feasibility for virtual reality-based exploration, navigation, exergaming, and rehabilitation. *Games Health J.* 7, 151–156. doi: 10.1089/g4h.2017.0114

Bozgeyikli, E., Raij, A., Katkoori, S., and Dubey, R. (2016). "Point and teleport locomotion technique for virtual reality," in *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (ACM), 205–216. doi: 10.1145/2967934.2968105

Brade, J., Dudczig, M., and Klimant, P. (2018). "Using virtual prototyping technologies to evaluate human-machine-interaction concepts," in *AW&I Conference*, Vol. 3 (Chemnitz).

Brennesholtz, M. S. (2018). "3-1: invited paper: vr standards and guidelines," in *SID Symposium Digest of Technical Papers,* Vol. 49 (Pleasantville, NY), 1–4. doi: 10.1002/sdtp.12476

Burgess, P. W., Alderman, N., Evans, J., Emslie, H., and Wilson, B. A. (1998). The ecological validity of tests of executive function. *J. Int. Neuropsychol. Soc.* 4, 547–558. doi: 10.1017/S1355617798466037

Canty, A. L., Fleming, J., Patterson, F., Green, H. J., Man, D., and Shum, D. H. (2014). Evaluation of a virtual reality prospective memory task for use with individuals with severe traumatic brain injury. *Neuropsychol. Rehab.* 24, 238–265. doi: 10.1080/09602011.2014.881746

Caputo, F., Greco, A., Egidio, D. A., Notaro, I., and Spada, S. (2017). "A preventive ergonomic approach based on virtual and immersive reality," in *International Conference on Applied Human Factors and Ergonomics* (Cham: Springer), 3–15. doi: 10.1007/978-3-319-60582-1_1

Chandra, K. (2012). A comparison of objective-C and C#: implementing Cbox class. *J. Comput. Sci. Coll.* 27, 178–186.

Chaytor, N., and Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: a review of the literature on everyday cognitive skills. *Neuropsychol. Rev.* 13, 181–197. doi: 10.1023/B:NERV.0000009483.91468.fb

Cox, D. R., and Donnelly, C. A. (2011). *Principles of Applied Statistics*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139005036

Culpepper, S. A., and Aguinis, H. (2011). R is for revolution: a cutting-edge, free, open source statistical package. *Org. Res. Methods* 14, 735–740. doi: 10.1177/1094428109355485

de França, A. C. P., and Soares, M. M. (2017). "Review of virtual reality technology: an ergonomic approach and current challenges," in *International Conference on Applied Human Factors and Ergonomics* (Cham: Springer), 52–61. doi: 10.1007/978-3-319-60582-1_6

Dickson, P. E., Block, J. E., Echevarria, G. N., and Keenan, K. C. (2017). "An experience-based comparison of unity and unreal for a stand-alone 3D game development course," in *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education* (ACM), 70–75. doi: 10.1145/3059009.3059013

Elkind, J. S., Rubin, E., Rosenthal, S., Skoff, B., and Prather, P. (2001). A simulated reality scenario compared with the computerized Wisconsin card sorting test: an analysis of preliminary results. *CyberPsychol. Behav.* 4, 489–496. doi: 10.1089/109493101750527042

Farrell, S., and Lewandowsky, S. (2010). Computational models as aids to better reasoning in psychology. *Curr. Dir. Psychol. Sci.* 19, 329–335. doi: 10.1177/0963721410386677

Figueiredo, L., Rodrigues, E., Teixeira, J., and Techrieb, V. (2018). A comparative evaluation of direct hand and wand interactions on consumer devices. *Comput. Graphics* 77, 108–121. doi: 10.1016/j.cag.2018.10.006

Frommel, J., Sonntag, S., and Weber, M. (2017). "Effects of controller-based locomotion on player experience in a virtual reality exploration game," in *Proceedings of the 12th International Conference on the Foundations of Digital Games* (ACM), 30. doi: 10.1145/3102071.3102082

Gavgani, A. M., Wong, R. H. X., Howe, P. R. C., Hodgson, D. M., Walker, F. R., and Nalivaiko, E. (2018). Cybersickness-related changes in brain hemodynamics: a pilot study comparing transcranial Doppler and near-infrared spectroscopy assessments during a virtual ride on a roller coaster. *Physiol. Behav.* 191, 56–64. doi: 10.1016/j.physbeh.2018.04.007

Gehrke, L., Iversen, J. R., Makeig, S., and Gramann, K. (2018). "The invisible maze task (IMT): interactive exploration of sparse virtual environments to investigate action-driven formation of spatial representations," in *German Conference on Spatial Cognition* (Cham: Springer), 293–310. doi: 10.1007/978-3-319-96385-3_20

Gonzalez-Franco, M., and Lanier, J. (2017). Model of illusions and virtual reality. *Front. Psychol.* 8:1125. doi: 10.3389/fpsyg.2017.01125

Grewe, P., Lahr, D., Kohsik, A., Dyck, E., Markowitsch, H. J., Bien, C. G., et al. (2014). Real-life memory and spatial navigation in patients with focal epilepsy: ecological validity of a virtual reality supermarket task. *Epilepsy Behav.* 31, 57–66. doi: 10.1016/j.yebeh.2013.11.014

Gromala, D., Tong, X., Shaw, C., Amin, A., Ulas, S., and Ramsay, G. (2016). Mobius floe: an immersive virtual reality game for pain distraction. *Electr. Imaging* 2016, 1–5. doi: 10.2352/ISSN.2470-1173.2016.4.ERVR-413

Haines, S., Shelton, J., Henry, J., Terrett, G., Vorwerk, T., and Rendell, P. (2019). *Prospective Memory and Cognitive Aging*. Oxford Research Encyclopedia of Psychology. Available online at: https://oxfordre.com/psychology/view/10.1093/acrefore/9780190236557.001.0001/acrefore-9780190236557-e-381 (accessed December 7, 2019).

Held, L., and Ott, M. (2018). On p-values and Bayes factors. *Ann. Rev. Stat. Appl.* 5, 393–419. doi: 10.1146/annurev-statistics-031017-100307

Higginson, C. I., Arnett, P. A., and Voss, W. D. (2000). The ecological validity of clinical tests of memory and attention in multiple sclerosis. *Arch. Clin. Neuropsychol.* 15, 185–204. doi: 10.1093/arclin/15.3.185

Jansari, A. S., Devlin, A., Agnew, R., Akesson, K., Murphy, L., and Leadbetter, T. (2014). Ecological assessment of executive functions: a new virtual reality paradigm. *Brain Imp.* 15, 71–87. doi: 10.1017/BrImp.2014.14

JASP Team (2017). *JASP (Version 0.8.1.2)* [Computer software]. JASP Team

Jerald, J. (2015). *The VR Book: Human-Centered Design for Virtual Reality*. New York, NY: Morgan and Claypool. doi: 10.1145/2792790

Jerald, J., LaViola Jr, J. J., and Marks, R. (2017). "VR interactions," in *ACM SIGGRAPH 2017 Courses* (Los Angeles, CA: ACM), 19. doi: 10.1145/3084873.3084900

Kobayashi, M., Ueno, K., and Ise, S. (2015). The effects of spatialized sounds on the sense of presence in auditory virtual environments: a psychological and physiological study. *Presence* 24, 163–174. doi: 10.1162/PRES_a_00226

Korre, D. (2019). *Usability evaluation of spoken humanoid embodied conversational agents in mobile serious games* (PhD thesis). University of Edinburgh, Edinburgh, United Kingdom.

Kourtesis, P., Collina, S., Doumas, L. A. A., and MacPherson, S. E. (2019a). Technological competence is a precondition for effective implementation of virtual reality head mounted displays in human neuroscience: a technological review and meta-analysis. *Front. Hum. Neurosci.* 13:342. doi: 10.3389/fnhum.2019.00342

Kourtesis, P., Collina, S., Doumas, L. A. A., and MacPherson, S. E. (2019b). Validation of the virtual reality neuroscience questionnaire: maximum duration of immersive virtual reality sessions without the presence of pertinent adverse symptomatology. *Front. Hum. Neurosci.* 13:417. doi: 10.3389/fnhum.2019.00417

Kriegeskorte, N., and Douglas, P. K. (2018). Cognitive computational neuroscience. *Nat. Neurosci.* 21, 1148–1160. doi: 10.1038/s41593-018-0210-5

LaViola, J. J. Jr., Kruijff, E., McMahan, R. P., Bowman, D., and Poupyrev, I. P. (2017). *3D User Interfaces: Theory and Practice*. Boston, MA: Addison-Wesley Professional.

Lavoué, G., and Mantiuk, R. (2015). "Quality assessment in computer graphics," in *Visual Signal Quality Assessment* (Cham: Springer), 243–286. doi: 10.1007/978-3-319-10368-6_9

Maister, L., Slater, M., Sanchez-Vives, M. V., and Tsakiris, M. (2015). Changing bodies changes minds: owning another body affects social cognition. *Trends Cogn. Sci.* 19, 6–12. doi: 10.1016/j.tics.2014.11.001

Makeig, S., Gramann, K., Jung, T. P., Sejnowski, T. J., and Poizner, H. (2009). Linking brain, mind and behavior. *Int. J. Psychophysiol.* 73, 95–100. doi: 10.1016/j.ijpsycho.2008.11.008

Marsman, M., and Wagenmakers, E. J. (2017). Bayesian benefits with JASP. *Euro. J. Dev. Psychol.* 14, 545–555. doi: 10.1080/17405629.2016.1259614

Mathôt, S., Schreij, D., and Theeuwes, J. (2012). OpenSesame: an open-source, graphical experiment builder for the social sciences. *Behav. Res. Methods* 44, 314–324. doi: 10.3758/s13428-011-0168-7

McDaniel, M. A., and Einstein, G. O. (2000). Strategic and automatic processes in prospective memory retrieval: a multiprocess framework. *Appl. Cogn. Psychol.* 14, S127–S144. doi: 10.1002/acp.775

McDaniel, M. A., and Einstein, G. O. (2007). *Prospective Memory: An Overview and Synthesis of an Emerging Field.* Thousand Oaks, CA: Sage Publications.

Mlinac, M. E., and Feng, M. C. (2016). Assessment of activities of daily living, self-care, and independence. *Arch. Clin. Neuropsychol.* 31, 506–516. doi: 10.1093/arclin/acw049

Mori, M., MacDorman, K. F., and Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robot. Automat. Mag.* 19, 98–100. doi: 10.1109/MRA.2012.2192811

Nalivaiko, E., Davis, S. L., Blackmore, K. L., Vakulin, A., and Nesbitt, K. V. (2015). Cybersickness provoked by head-mounted display affects cutaneous vascular tone, heart rate and reaction time. *Physiol. Behav.* 151, 583–590. doi: 10.1016/j.physbeh.2015.08.043

Palmisano, S., Mursic, R., and Kim, J. (2017). Vection and cybersickness generated by head-and-display motion in the Oculus Rift. *Displays* 46, 1–8. doi: 10.1016/j.displa.2016.11.001

Pan, X., and Hamilton, A. F. D. C. (2018). Why and how to use virtual reality to study human social interaction: the challenges of exploring a new research landscape. *Br. J. Psychol.* 109, 395–417. doi: 10.1111/bjop.12290

Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Front. Hum. Neurosci.* 9:660. doi: 10.3389/fnhum.2015.00660

Parsons, T. D., McMahan, T., and Kane, R. (2018). Practice parameters facilitating adoption of advanced technologies for enhancing neuropsychological assessment paradigms. *Clin. Neuropsychol.* 32, 16–41. doi: 10.1080/13854046.2017.1337932

Peirce, J. W. (2007). PsychoPy–psychophysics software in Python. *J. Neurosci. Methods* 162, 8–13. doi: 10.1016/j.jneumeth.2006.11.017

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Front. Neuroinf.* 2:10. doi: 10.3389/neuro.11.010.2008

Phillips, L. H., Henry, J. D., and Martin, M. (2008). "Adult aging and prospective memory: the importance of ecological validity," in *Prospective Memory: Cognitive, Neuroscience, Developmental, and Applied Perspectives*, eds M. Kliegel, M. A. McDaniel, and G. O. Einstein (Taylor & Francis Group, Lawrence Erlbaum Associates), 161–185.

Porcino, T. M., Clua, E., Trevisan, D., Vasconcelos, C. N., and Valente, L. (2017). "Minimizing cyber sickness in head mounted display systems: design guidelines and applications," in *Serious Games and Applications for Health (SeGAH), 2017 IEEE 5th International Conference on* (IEEE), 1–6. doi: 10.1109/SeGAH.2017.7939283

Rand, D., Basha-Abu Rukan, S., Weiss, P. L., and Katz, N. (2009). Validation of the Virtual MET as an assessment tool for executive functions. *Neuropsychol. Rehab.* 19, 583–602. doi: 10.1080/09602010802469074

Revelle, W. (2011). *An Overview of the Psych Package*. Evanston, IL: Department of Psychology, Northwestern University.

Rizzo, A. A., Schultheis, M., Kerns, K. A., and Mateer, C. (2004). Analysis of assets for virtual reality applications in neuropsychology. *Neuropsychol. Rehab.* 14, 207–239. doi: 10.1080/09602010343000183

Robertson, I. H., Ward, T., Ridgeway, V., and Nimmo-Smith, I. (1994). *The Test of Everyday Attention (TEA)*. Bury St. Edmunds: Thames Valley Test Company.

Rosenberg, L. (2015). The associations between executive functions' capacities, performance process skills, and dimensions of participation in activities of daily life among children of elementary school age. *Appl. Neuropsychol.* 4, 148–156. doi: 10.1080/21622965.2013.821652

Rumiński, D. (2015). An experimental study of spatial sound usefulness in searching and navigating through AR environments. *Virt. Real.* 19, 223–233. doi: 10.1007/s10055-015-0274-4

Sejnowski, T. J., Koch, C., and Churchland, P. S. (1988). Computational neuroscience. *Science* 241, 1299–1306. doi: 10.1126/science.3045969

Seyama, J. I., and Nagayama, R. S. (2007). The uncanny valley: effect of realism on the impression of artificial human faces. *Presence* 16, 337–351. doi: 10.1162/pres.16.4.337

Sharples, S., Cobb, S., Moody, A., and Wilson, J. R. (2008). Virtual reality induced symptoms and effects (VRISE): comparison of head mounted display (HMD), desktop and projection display systems. *Displays* 29, 58–69. doi: 10.1016/j.displa.2007.09.005

Sherman, W. R., and Craig, A. B. (2018). *Understanding Virtual Reality: Interface, Application, and Design*. San Francisco, CA: Morgan Kaufmann. doi: 10.1016/B978-0-12-800965-9.00010-6

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philos. Trans. R. Soc. B* 364, 3549–3557. doi: 10.1098/rstb.2009.0138

Slater, M. (2018). Immersion and the illusion of presence in virtual reality. *Br. J. Psychol.* 109:431. doi: 10.1111/bjop.12305

Slater, M., Spanlang, B., and Corominas, D. (2010). Simulating virtual environments within virtual environments as the basis for a psychophysics of presence. *ACM Trans. Graphics* 29:92. doi: 10.1145/1778765.1778829

Smith, R. E. (2003). The cost of remembering to remember in event-based prospective memory: investigating the capacity demands of delayed intention performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 347. doi: 10.1037/0278-7393.29.3.347

Smith, R. E., Hunt, R. R., McVay, J. C., and McConnell, M. D. (2007). The cost of event-based prospective memory: Salient target events. *J. Exp. Psychol.* 33:734. doi: 10.1037/0278-7393.33.4.734

Sportillo, D., Paljic, A., Boukhris, M., Fuchs, P., Ojeda, L., and Roussarie, V. (2017). "An immersive Virtual Reality system for semi-autonomous driving simulation: a comparison between realistic and 6-DoF controller-based interaction," in *Proceedings of the 9th International Conference on Computer and Automation Engineering* (Sydney: ACM), 6–10. doi: 10.1145/3057039.3057079

Stevens, J. R. (2017). Replicability and reproducibility in comparative psychology. *Front. Psychol.* 8:862. doi: 10.3389/fpsyg.2017.00862

Toschi, N., Kim, J., Sclocco, R., Duggento, A., Barbieri, R., Kuo, B., et al. (2017). Motion sickness increases functional connectivity between visual motion and nausea-associated brain regions. *Auton. Neurosci.* 202, 108–113. doi: 10.1016/j.autneu.2016.10.003

Viirre, E., Price, B. J., and Chase, B. (2014). "Direct effects of virtual environments on users," in *Handbook of Virtual Environments* (CRC Press), 527–535. doi: 10.1201/b17360-28

Vorländer, M., and Shinn-Cunningham, B. (2014). "Virtual auditory displays," in *Handbook of Virtual Environments* (CRC Press), 107–134. doi: 10.1201/b17360-7

Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., et al. (2018a). Bayesian inference for psychology. Part II: example applications with JASP. *Psychon. Bull. Rev.* 25, 58–76. doi: 10.3758/s13423-017-1323-7

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., et al. (2018b). Bayesian inference for psychology. Part I: theoretical advantages and practical ramifications. *Psychon. Bull. Rev.* 25, 35–57. doi: 10.3758/s13423-017-1343-3

Wells, J. D., Campbell, D. E., Valacich, J. S., and Featherman, M. (2010). The effect of perceived novelty on the adoption of information technology innovations: a risk/reward perspective. *Dec. Sci.* 41, 813–843 doi: 10.1111/j.1540-5915.2010.00292.x

Wilson, B. A., Cockburn, J., and Baddeley, A. (2008). *The Rivermead Behavioural Memory Test*. London: Pearson Assessment.

Wilson, B. A., Emslie, H., Foley, J., Shiel, A., Watson, P., Hawkins, K., et al. (2005). *The Cambridge Prospective Memory Test: CAMPROMPT*. London: Harcourt Assessment.

Wilson, B. A., Evans, J. J., Emslie, H., Alderman, N., and Burgess, P. (1998). The development of an ecologically valid test for assessing patients with a dysexecutive syndrome. *Neuropsychol. Rehab.* 8, 213–228. doi: 10.1080/713755570

Zaidi, S. F. M., Duthie, C., Carr, E., and Maksoud, S. H. A. E. (2018). "Conceptual framework for the usability evaluation of gamified virtual reality environment for non-gamers," in *Proceedings of the 16th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry* (Tokyo: ACM), 13. doi: 10.1145/3284398.3284431

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Computer Science

ORIGINAL RESEARCH
published: 04 June 2020
doi: 10.3389/fcomp.2020.00017

External Assistance Techniques That Target Core Game Tasks for Balancing Game Difficulty

Jawad Jandali Refai[1,2], *Scott Bateman*[1]* *and Michael W. Fleming*[1]

[1] *HCI Lab, Faculty of Computer Science, University of New Brunswick, Fredericton, NB, Canada,* [2] *Kabam Games, Vancouver, BC, Canada*

Game balancing is a time consuming and complex requirement in game design, where game mechanics and other aspects of a game are tweaked to provide the right level of challenge and play experience. One way that game designers help make challenging mechanics easier is through the use of External Assistance Techniques—a set of techniques outside of games' main mechanics. While External Assistance Techniques are well-known to game designers (like providing onscreen guides to help players push the right buttons at the right times), there are no guiding principles for how these can be applied to help balance challenge in games. In this work, we present a design framework that can guide designers in identifying and applying External Assistance Techniques from a range of existing assistance techniques. We provide a first characterization of External Assistance Techniques showing how they can be applied by first identifying a game's Core Tasks. In games that require skill mechanics, Core Tasks are the basic motor and perceptual unit tasks required to interact with a game, such as aiming at a target or remembering a detail. In this work we analyze 54 games, identifying and organizing 27 External Assistance Techniques into a descriptive framework that connects them to the ten core tasks that they assist. We then demonstrate how designers can use our framework to assist a previously understudied core task in three games. Through an evaluation, we show that the framework is an effective tool for game balancing, and provide commentary on key ways that External Assistance Techniques can affect player experience. Our work provides new directions for research into improving and maturing game balancing practices.

Keywords: game balancing, external assistance techniques, core tasks, video games, difficulty adjustment

OPEN ACCESS

Edited by:
Ioanna Iacovides,
University of York, United Kingdom

Reviewed by:
Alena Denisova,
City University of London, United Kingdom
Julian Frommel,
University of Saskatchewan, Canada

****Correspondence:***
Scott Bateman
scottb@unb.ca

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Computer Science

Received: *01 February 2020*
Accepted: *23 April 2020*
Published: *04 June 2020*

Citation:
Refai JJ, Bateman S and Fleming MW (2020) External Assistance Techniques That Target Core Game Tasks for Balancing Game Difficulty. Front. Comput. Sci. 2:17. doi: 10.3389/fcomp.2020.00017

INTRODUCTION

Challenge is an important part of what makes a game entertaining (Chen, 2007). Striking the right level of challenge is critical for a game design to be successful: if it is too difficult, it can become frustrating; if it is too easy, players may become disengaged and uninterested (Vazquez, 2011). Traditionally, game designers try to find the right level of challenge through the activity of *game balancing*, where aspects of the rules are tweaked to target the right play experience (Schell, 2019).

Game balancing is extremely challenging, because many of a game's parameters are interconnected (Baron, 2012). For example, imagine a platformer game where in playtesting it is uncovered that players find jumping over large pits too difficult (the split-second timing required is hard for players to master). The designer may consider increasing the character's jumping distance by tweaking aspects of the game's physics. However, this would have many side effects in other parts of the game; e.g., changing running speed may make avoiding enemies too easy and changing gravity will affect the behavior of many other game objects. In such games with interconnected mechanics (such as platformers or first-person shooters), game balancing that operates in the space of the game world is extremely difficult and time consuming because of the interconnected nature of a game's in-world mechanics and properties. Tweaking game mechanics for balance commonly leads to unintended consequences, and can lead to mechanics operating in unsatisfying ways (Baron, 2012). Because of the limited research and reports of practice, game balancing still remains more of an art than science (Schell, 2019).

One way that game designers balance challenge is to use *external assistance techniques*. We define external assistance techniques (or EA techniques) as a set of approaches that work outside of the main mechanics of the game world, but allow a player to more easily complete challenges that are tightly connected to the game world by allowing them to better perform *core tasks*. From previous work in game design, we borrow and refine the definition of "core tasks," the "basic motor and perceptual tasks" that games require in order to interact with game mechanics (Flatla et al., 2011). Our focus on "external" means that the assistance techniques we consider in this work do not need to change the main game mechanics to be effective, rather they can be added separately or distinctly from a game's core mechanics (e.g., a game's rules, physics, or other character or game object behaviors can remain unchanged). Previous work has proposed the use of "assistance techniques" (Bateman et al., 2011b; Cechanowicz et al., 2014; Vicencio-Moreira et al., 2014, 2015); however, the previously proposed techniques all focused on "internal" approaches for balancing gameplay, by directly making it easier for a player to perform a game's main mechanic. One example of this is making it easier to aim at a target by automatically moving the aiming reticule (Bateman et al., 2011b; Vicencio-Moreira et al., 2014). Our distinction between internal and external techniques, places the focus on the idea that with external techniques we are assisting the player in performing a particular challenging task, rather than changing the task itself (changing the task would be internal to the game).

To illustrate the idea of external assistance techniques, consider the hypothetical platform game described above. Using the framework described in this paper, a game designer may identify a core "reaction time" task to jump over a pit in the game. Based on this, they may identify the "advance warning" technique as an appropriate approach and provide an additional visual indicator onscreen as the player approaches a pit. This allows players to learn exactly when they need to press the jump button, by providing an external cue. Another example, is in a large open world game like World of Warcraft, where players must recall the location of places they need to visit for a quest. In this case the core task required of the player is a "spatial memory task," which can be assisted using the common "map" technique (i.e., providing a map to guide the player from their current location to their intended destination).

Providing external assistance techniques (EA techniques) to balance difficulty has advantages over balancing other in-world mechanics and game object properties. This is because EA techniques can be used more selectively (e.g., an external cue might only be shown for the first encounter with a pit that is difficult to jump across) and are less likely to have unforeseen side-effects in the game world (e.g., if, say, jumping height was increased). Further, they are particularly attractive because they allow otherwise difficult tasks to remain in place, but assist players to better perform a required skill mechanic, like pressing the right button combination at exactly the right time, or remembering the location of an object in the game world; this allows players to better stay immersed in a game while having a satisfying experience (Weihs, 2013).

Game designers are generally aware of many of the assistance techniques that have been previously used (Burgun, 2011); however, it is not always clear how they can be applied, because every game is unique (Bourtos, 2008), and there is little work to help organize and describe the range of assistance techniques that might be possible. In this paper, we address the challenge confronting game designers to identify and select the best EA technique for their game. To do this we propose a generalized way of identifying appropriate EA assistance techniques in video games that require the core tasks (e.g., precise timing or recalling a specific detail). Our approach involves identifying the fundamental actions players perform in many games—*core tasks*—such as signal detection, reaction time, or pointing, we can identify a range of *EA techniques* that can improve play experience independently from the characteristics of a specific game.

In order to characterize the space of core tasks and their relationship to EA techniques, we performed a grounded analysis of 54 games. Our analysis started with an existing characterization of core tasks (Flatla et al., 2011), refining them based on the results of our grounded theory study. This resulted in a design framework of 10 core tasks and a description of 27 external assistance techniques that have been previously used to balance challenge in games. Next, to demonstrate the effectiveness and generalizability of our framework, we built three different games (a puzzle-like game, a third-person adventure game, and a sniper simulation) that share a common core task. We implemented three different assistance techniques that target the shared core task in each of the three games, showing that even though the games are seemingly different, assistance techniques can be adapted to fit all of them. Finally, to demonstrate the effectiveness of the resulting techniques for game balancing, we conducted an experiment on two of the games comparing the games to versions without an assistance technique. The results of our experiment show that the EA techniques increased player performance and that they were effective in reducing challenge in the games, meaning that they are effective tools for balancing challenge.

While approaches to game balancing have previously been studied (Bateman et al., 2011b; Cechanowicz et al., 2014; Vicencio-Moreira et al., 2014, 2015), this work has mostly identified specific assistance techniques for certain activities. There has been very little work that has organized and characterized the range of previously proposed EA techniques. It is important to note that the number of techniques and ways that game designers might choose to balance challenge is large. However, in this work we focus on what we call *external* assistance techniques, techniques that operate outside a game's main, in-world mechanics. To focus our initial work in this area, we necessarily exclude many other ways that games may be balanced; for example, providing guidance/advice in strategy games, or by changing in-game mechanics (e.g., by adjusting physics or character or game object properties). In this paper, we provide a demonstration of how EA techniques can be applied to help designers target a desired play experience. Our work provides game designers with a valuable new resource for understanding game balancing practices, guidance for identifying and applying external assistance techniques, and discovering new assistance techniques that can be applied in their games. Ultimately, our work contributes to the advancement of game design and development.

RELATED WORK

Difficulty and Flow

The fact that overly difficult games cause frustration, whilst easy ones lead to boredom, is considered "common knowledge" amongst game designers (Vazquez, 2011). As such, how to design for the appropriate level of difficulty has become an increasingly popular subject with game designers (Vazquez, 2011), as well as what aspects of a game can be manipulated to control difficulty (e.g., time limits, damage scaling, HUD restrictions, etc.) (Bourtos, 2008).

Commonly in game design, difficulty has been related to the concept of *flow*, which can be defined as the state in which people are so immersed in an activity that everything else ceases to matter, and their perception and experience of time become distorted (Csikszentmihalyi, 2008). Flow is achieved when a person's skill ideally aligns with the difficulty of the task at hand, which promotes a high level of engagement and focus on the task (Baron, 2012). In general, game designers want players to remain in a state of flow throughout their experience, which would represent a rich and meaningful engagement with the game (Salen and Zimmerman, 2003). To achieve this, Csikszentmihalyi outlines four task "characteristics" that increase the probability of achieving a flow state. One such characteristic is to "… demand actions to achieve goals that fit within the person's capabilities" (Baron, 2012).

Denisova et al. provide a more nuanced account of player experience arguing that play experience can better be characterized through the "challenges" a game presents. Challenges "… describe a stimulating task or problem," while "difficulty" simply implies that "… something is hard to do" (Denisova et al., 2017). So, in game design, balancing player skill and the challenges they face, culminate in their experience of a game's difficulty, and should be done in a way to maintain equilibrium between stress/arousal and performance. Each player has a unique stress-performance curve, and thus a gradual increase in overall difficulty (easy, medium, and hard) is not necessarily optimal. EA techniques can be used to influence a player's performance and their perception of challenge, so that it corresponds with an appropriate difficulty to promote cognitive flow (Baron, 2012).

While game balancing is a necessary activity in almost any successful game, details around existing practices are not always widely shared in the game industry (Felder, 2015a). This may be, at least partially, because game balancing practices have not reached the same level of maturity as other design and development practices. However, it is generally understood game balancing is an iterative process that takes place throughout development, usually following feedback from play testing (Felder, 2015b; Schell, 2019). Much effort is often placed into balancing activities, though, since there are few well-established practices. Further, as discussed, the fact that game elements are often interconnected means that balancing games is complex and any changes to a games mechanics needs to be tested thoroughly to ensure that other interrelated aspects of the game have not been adversely affected (Felder, 2015b; Schell, 2019).

Games Mechanics, Skill and Core Tasks

Game mechanics can concisely be described as the rules of a game and how players interact with the game. Schell describes "mechanics of skill" as one of the six main types of mechanics, since "Every game requires players to exercise certain skills" (2019). Games most frequently require a range of skills, which can be categorized as (Schell, 2019):

- *physical skills*: skills requiring dexterity, movement, speed, etc.; such as using a game controller.
- *mental skills:* skills including memory, observation, insight, problem solving, developing and following a strategy, etc.
- *social skills:* building trust, guessing an opponent's strategy, team communication/coordination, etc.

Similarly, Adams describes that the challenges a player must overcome can be considered as being either mental or physical (Adams, 2013). These categorizations of skill mechanics open up huge number of ways that games might be out of balance because of a mismatch between player skill and a design, both in low-level interactions (the need to click a button quickly) or higher-order cognitive tasks (e.g., developing a strategy in a game of Chess). In our work, we were initially interested in determining how game balancing practices might be facilitated and improved through further focusing on a particular subset of the physical and mental skills described above.

We were interested in providing a concrete characterization of how a specific set of skills could be assisted. When considering a range of skills fundamental to interacting with games, we found the work of Flatla et al.'s on "calibration games" to be helpful (Flatla et al., 2011). Calibration games are essentially gamified calibration tasks that are designed to encourage people to perform necessary calibration steps needed for many input technologies to operate reliably (e.g., calibrating an eye-tracker for a particular user). In this work, the authors use the idea of *core tasks:* "the core perceptual and motor tasks that … match common game mechanics.…" These included a list of 10 core tasks such as reaction time, visual search, and spatial memory. When relating the core tasks to Schell's skill categorization (enumerated above), it can be seen that Flatla et al.'s core tasks relate to lower level physical and mental skills [described by Newell (1994) in his "Time Scale of Human Action" as unit tasks, operations or deliberate acts], but not to higher level tasks that involve rationalization (e.g., making a choice or developing a strategy). This means that supporting these skills is more tractable and success is more easily measured, since there are fewer sources of variation (that can arise from, say, how an individual engages in conscious deliberation or in human-to-human communication), which would be prevalent for social or higher level mental skills (MacKenzie, 2013). For these reasons, our work leverages Flatla et al.'s list as a concrete and tractable subset of game tasks that represent common skills in games.

Research and Practice in External Assistance Techniques

One way of helping players who are struggling with a game challenge is to assist them with the task preventing their progress, effectively increasing their skill. For example, suppose a player is having a tough time hitting a target with a set number of bullets. Instead of making more bullets available (a typical *internal* game balancing approach), we could instead assist them with their aiming skill to increase the probability of a successful shot. Bateman et al. (2011b) used the term, "target assistance techniques" to describe a set of algorithms that helped players acquire and shoot targets in a multiplayer target shooting game. This work showed that several "target assistance techniques" were effective for helping to balance competition between players of different skill levels. Likely predating this, game designers use the idea of "aim assists" to describe techniques to help players acquire techniques in first-person shooters (Weihs, 2013).

Games have used specific assistance techniques to assist certain core tasks that people find difficult, and these are well-known to designers. As described, first-person shooters often incorporate some form of aim assistance or "auto aiming" [Auto-Aim (Concept), 2019] to help players deal with the difficult task of aiming a reticule at a rapidly moving target, especially when using a thumbstick on a gamepad where control is more difficult than a mouse (Vicencio-Moreira et al., 2014). Common aim assists that can be employed to improve play experience are techniques such as bullet magnetism, reticule magnetism, and auto-locking [Bateman et al., 2011b; Vicencio-Moreira et al., 2014; Auto-Aim (Concept), 2019].

Researchers have explored the concept of EA techniques improving play experience and player performance in several specific types of games including racing games (Bateman et al., 2011a; Cechanowicz et al., 2014), shooting games (Bateman et al., 2011b), and first-person shooters (Vicencio-Moreira et al., 2014, 2015), and have compared the effectiveness of several visual search assistance techniques in an AR game (Lyons, 2016). Also of note is work looking at balancing player skill in traditional multiplayer games (Bateman et al., 2011b; Cechanowicz et al., 2014; Vicencio-Moreira et al., 2014), or between players of different physical abilities (Gerling et al., 2014). Here we keep our review of the existing techniques brief and refer to relevant literature from research and current practice as we introduce the individual EA techniques in our framework.

Previous research provides valuable information comparing different techniques that allow players to better perform a certain core task. However, while this work has proposed (sometimes novel) assistance techniques within a particular context (i.e., using specific input or display devices, a certain type of game, etc.), it is still difficult for game designers to consider the wide range of possibilities for balancing games (Bourtos, 2008; Burgun, 2011; Vazquez, 2011; Baron, 2012; Felder, 2015a; Schell, 2019). Through the characterization of core tasks, our goal is to discuss a range of assistance techniques at a general level that could be applied to any game, irrespective of context. We believe this conceptual organization will provide both game designers and researchers with a starting point to explore and consider a range of EA techniques that can be applied to games to help target a desired level of challenge.

A FRAMEWORK OF CORE TASKS AND EXTERNAL ASSISTANCE TECHNIQUES

To characterize both core tasks and EA techniques we conducted a grounded theory study, which resulted in a framework describing the EA techniques that have been leveraged to assist certain core tasks. In this section we first describe our methodology, then describe our resulting framework, and, finally, we describe the general steps that can be used to adapt EA techniques to existing games using our framework as a guide.

Methodology

Our work used a grounded theory study to create a framework of external assistance techniques that can assist players in completing core tasks in games. Grounded theory is comprised of qualitative practices used to characterize a new domain through the development of codes that are derived from data (Glaser, 1998; Glaser and Strauss, 2017). Grounded theory has been commonly used as a methodology for identifying frameworks from games artifacts (Toups et al., 2014; Alharthi et al., 2018; Wuertz et al., 2018), and our work follows the processes described in this previous work. We adopted a multi-phase process, whereby the research team identified codes from several iterative rounds of data collection and open coding. While Glaser and Straus describe how this process can be supported and informed by existing theory, we also leaned on multi-grounded theory (Goldkuhl and Cronholm, 2010). Multi-grounded theory follows the standard Glaser type approach, but in the structuring step describes how the process can be both inductive (to inform and refine existing theory) and deductive (drawing on existing theory to guide the process). Our process involved three general phases established in previous research (Wuertz et al., 2018):

- Phase 1: identifying and selecting game examples that contain core tasks,
- Phase 2: open coding from initial observations, and
- Phase 3: revision of our coding scheme, and development of axial codes.

All phases involved the research team engaging in discussion to explore the similarities and differences between their codes, concepts and our list of core tasks and external assistance techniques. Below we elaborate on each of the phases.

Phase 1: Finding and Selecting Games

In selecting games for analysis we followed the process from other recent work that led to the creation of a framework to inform game design using grounded theory. Our initial selection process involved selecting games that the authors were familiar with (Wuertz et al., 2018).

Our goal with game selection was to identify games that contained core tasks, but we also believed that different genres may involve different core tasks (perhaps that had not previously been identified) and might also use vastly different EA techniques. We used a high-level taxonomy of games to assist in getting a mix of genres (Wikipedia, 2020), and we initially selected games from 16 of the genres and subgenres that we believed represented a good mix of games; the initial list of genres and games is available in **Supplementary Materials**. This list was only used to help diversify our initial game selection. The non-exhaustive list of genres and sub-genres comes from a list of video game genres on (Wikipedia, 2020), and game examples are drawn from the genre descriptions on this page. Subsequent iterations relied on selecting games that maximized variability based on our identified codes and did not use game genres.

Our inclusion criteria for games in our sample was relatively loose, in that a game only needed to have one core task and one EA technique to be included. To define core tasks we pre-determined that a core task must be a "basic motor and perceptual task" (Flatla et al., 2011) that takes place within the cognitive band of human action (i.e., excludes detailed deliberation, communications, or social processes) (Newell, 1994; MacKenzie, 2013). EA techniques were considered to be any feature in the game that was not related to game mechanics that operate in the game world.

Researchers frequently returned to Phase 1 after Phase 3 to seek out core tasks and EA techniques that were hypothesized about as potential codes. This also led us to have the following stopping criteria:

- Our existing axial codes did not suggest new core tasks or EA techniques that were not already represented in our dataset.
- We no longer found game examples that provided new core tasks or EA techniques.

This process resulted in 54 games that can be found in our Ludography, which is available as **Supplementary Material** to this paper. The games were analyzed either directly (through gameplay) or indirectly (by watching gameplay videos on YouTube).

Phase 2: Observations and Coding

Data were collected through experience reports from playing the games or through watching gameplay videos on YouTube. As new data were added, each was first evaluated for the core tasks involved, followed by EA technique identification. For each game we collected the game name, genre, descriptions for each core task, a listing and description of all observed game features that might be considered as EA techniques for each of the game's core tasks, and the data source (e.g., where it can be found in gameplay or a link to the YouTube video). As more games were added, we increasingly saw saturation in the data. The initial coding of core tasks and EA techniques was done without considering any existing theory, allowing us to later consider whether previous characterizations could accurately describe our data (which occurred as part of Phase 3).

Phase 3: Axial Coding and External Assistance Technique Classification

Through discussions, we iteratively refined our list of core tasks and the external assistance techniques used to assist them. Recall that we relied on existing theory to help narrow the scope of our interest (as described in Related Work; see section Games Mechanics, Skill and Core Tasks). We initially considered the descriptions of each core task that we collected and determined their relationship to other core tasks. We then considered whether the core tasks could be reconciled with the core tasks described by Flatla et al., who identified a list of 14 core tasks (Flatla et al., 2011). Our process resulted in 10 core tasks, since we found that several of Flatla at al.'s core tasks were conceptually similar and could be supported by the same assistance techniques, thus we merged them. Our list of core tasks in the end is pragmatic reflection of our data collection, rather than having direct correspondence to, say, individual (or atomic) psychomotor control tasks (Schmidt et al., 2018).

We repeated a similar process with assistance techniques. Here, we iteratively grouped and labeled our assistance technique descriptions. Here we used the existing literature (described in Related Work) to consider our identified assistance techniques. This resulted in the 27 external assistance techniques, where each technique was aligned with one or more core task.

The Descriptive Framework

Below we present the descriptive framework that resulted from our analysis. **Figure 1** displays the results of our analysis, relating the 10 core tasks with the 27 identified assistance techniques. In the subsections below, we first define the core tasks, followed by our description of EA techniques. For each, EA technique we provide an example of the technique from an existing commercial or research game. In our description of the framework below, we reference examples from our Ludography (available in **Supplementary Material**) using the notation [L#].

CT1 Signal Detection

Definition: The conscious perception of a stimulus, such as sound, light, or vibration.

External Assistance Techniques for Signal Detection

1.1 *Companion Signals*: Accompanying the original signal with an additional signal, often of a different medium (Johanson and Mandryk, 2016). E.g., the 3 stages of the charged attack in Monster Hunter: World [L13] use a visual signal (pulsing energy around the character), an audio signal (a sound effect of varying pitch per tick), and a haptic signal (controller vibration with every tick).

1.2 *Augmented Parameters*: Modifying parameters of a signal to make it more or less noticeable. E.g., In Overwatch [L9], Lucio uses a Sonic Amplifier ability that fires a burst of soundwaves as projectiles. Each shot of Lucio's weapon is accompanied by a sound effect. The pitch of this sound effect increases with every shot made, making it stand out.

1.3 *Announcements/Emphasis*: Feedback to point out a signal and/or for directing the player's attention to it. E.g., When players mount a monster in Monster Hunter: World [L13], a message telling the player what to do is displayed in the corner of the screen; this message glows making it more visually salient.

1.4 *Signal Priority*: Giving the signal more access to its media, such as more space on a visual display. E.g., When a player is injured in Call of Duty [L27] an overlay creates a blood spatter effect on the screen, which increases as the player's health goes down.

1.5 *Transformation/Replacement*: Altering a signal to provide one that is more effective or more informative. E.g., In Call of Duty [L27] players are rewarded for kill streaks by having access to a power-up that changes the representation of enemies on the mini-map from a blip to a triangle; this additional information makes it easier for players to identify enemy location and movement.

1.6 *Noise Reduction*: Modifying other elements in the game scene to make the signal more discernible. E.g., Whenever someone speaks using the in-game voice chat in Counter Strike Global Offensive [L25], the volume of everything else is reduced slightly so that the voice is loud enough to be heard. Note that signal priority (1.4) strengthens the signal being communicated by a particular source, while noise reduction reduces other signals in order to highlight another information source.

CT2 Signal Discrimination

Definition: Determining that there is a difference between two stimuli (e.g., determining that two colors or two sounds are different).

External Assistance Techniques for Signal Discrimination

2.7 *Additional Cues*: Presenting additional signals to convey complementary information. E.g., First-person shooters, such as Overwatch [L9], display a red outline around enemies when targeted. This helps distinguish them from friendlies, as players on both teams can look nearly identical.

2.8 *Augmented Parameters*: Modifying parameters of a signal to make it more noticeable. E.g., In Overwatch [L9], enemies' footsteps are louder than friendlies' footsteps to help ensure players can discriminate between the two.

2.9 *High Contrast*: Transforming/replacing a target signal to vary the contrast between it and other signals. E.g., Ultimate abilities in Overwatch [L9] are accompanied by a unique battle cry, which differs depending on whether the character is an ally or an enemy.

2.10 *Noise Reduction*: Reducing noise to better notice certain details about the target signal. E.g., When aiming through the scope in Sniper Elite [L45], the sounds of the surrounding battle are nearly muted to help players concentrate and line up a perfect shot.

CT3 Body Controls

Definition: Using muscle activation such as flexing or movement of a body part.

External Assistance Techniques for Body Controls

3.11 *Input Modulation*: Amplifying or diminishing the input from sensors to reach a desired result. E.g., In "The Falling of Momo," a prosthesis training game (Tabor et al., 2017) there is an auto-calibration feature that automatically sets the gains on myoelectric sensors, so that the character can be controlled consistently between players with different muscle strengths.

3.12 *Detection Threshold*: varying the margin of error while posing or activating muscles. E.g., In Kinect Star Wars [L51], even minimal movements allow the players' input to be correctly interpreted as a goal movement.

CT4 Reaction Time

Definition: Reacting to a perceptual stimulus as quickly as possible.

External Assistance Techniques for Reaction Time

4.13 *Companion Signals*: Additional signals to alert the player that they are required to react. E.g., In NBA 2K19 [L54], on-screen indicators assist in timing the release of the button when shooting.

4.14 *Advance Warning*: Notifying players that a time-sensitive action is imminent. E.g., In Dead by Daylight [L5], skill checks (which require quick reaction) are announced by a gong before they appear.

CT5 Visual Search

Definition: Finding a visual target in a field of distractors; includes pattern recognition (determining the presence of a pattern amongst a field of distractors).

External Assistance Techniques for Visual Search

5.15 *Highlights*: Highlighting the target to make it stand out. E.g., Many quests in World of Warcraft [L7] require finding and collecting items. Target quest items are highlighted with a sparkle effect.

5.16 *Visual Connection*: A guide to the target via a visual connection (similar to Renner and Pfeiffer, 2017). E.g., Before the start of an Overwatch [L9] match, a line is shown on the ground to guide defending players from their starting location to the first defense point.

5.17 *Target Details*: Providing additional information about the target. E.g., In one mission of Grand Theft Auto V [L47] the player is provided with clues to identify an enemy, as the mission proceeds more clues are provided to facilitate identifying the enemy.

5.18 *Compass*: Pointer guiding the player in the target's direction. E.g., In Rocket League [L43], an arrow is always available pointing in the direction of the ball when it is off-screen.

5.19 *Directional Cues*: An auditory/visual effect to guide the players' attention. E.g., In Dead by Daylight [L5], when an escape hatch (which is difficult to find) is opened a sound of rushing air can be heard as the player nears it, guiding them to it.

CT6 Pointing

Definition: Accurately pointing at a target with feedback about current pointing position.

External Assistance Techniques for Pointing

6.20 *Sticky Targeting*: Slowing down the pointer when passing over a target (see Balakrishnan, 2004; Bateman et al., 2011b). E.g., Halo 5 [L1] uses a combination of aim assist techniques, one of which slows down the reticle as it passes over a target

6.21 *Target Gravity*: A force pulling the pointer toward the target based on distance (see Bateman et al., 2011b; Vicencio-Moreira et al., 2014). E.g., Fans of Battlefield 4 [L18] have analyzed the game and discovered that it includes a form of assistance to help players aim better. Two techniques were uncovered; one matches the Sticky Targeting technique, discussed earlier, and the other matches Target Gravity. In commercial games this is often referred to as reticule magnetism [Auto-Aim (Concept), 2019].

6.22 *Target Lock*: Instantly snapping the pointer to the target location [Auto-Aim (Concept), 2019]. E.g., In Monster Hunter: World [L5] (using slinger shot mode), players hold down a button that automatically targets a potential enemy; hitting another button will cycle through other targets.

CT7 Aiming

Definition: Accurately pointing at a target (possibly using a device) and/or predicting the collision between two objects, without feedback.

External Assistance Techniques for Aiming

7.23 *Target Lock*: Readjusting player position toward the target. E.g., In Naruto: Ultimate Ninja Storm [L16], target locking is provided. Players are free to roam around and even break the lock, but most actions players take reorient them toward their opponent.

7.24 *Projectile Magnetism*: Changing the projectile trajectory toward the target [Auto-Aim (Concept), 2019]. E.g., In Halo 5 [L1], bullets are pulled toward the target even if the shot was made slightly off-target. It should be noted that while projectile magnetism operates on a game object inside the game world, we consider it as an "external" technique because it does not change the game's main mechanics (i.e., firing a bullet with a particular trajectory), and can be done with little-to-no effect on other aspects of the game.

CT8 Steering

Definition: Moving or guiding an object along a path.

External Assistance Techniques for Steering

8.25 *View*: Giving the player a view into the game world that provides a better awareness of the environment. E.g., Dead by Daylight [L5] uses asymmetric view to create additional challenge. Survivors play the game in a third-person view, while the killer is given a first-person view. This makes it easier for the survivors to steer around obstacles and plan their routes as they can see more of their surroundings.

8.26 *Steering Adjustment*: Adjusting player velocity toward the optimal path (e.g., Cechanowicz et al., 2014). E.g., In Harry Potter: Quidditch World Cup [L33], to end a match, players chasing the snitch (a flying golden ball) are assisted to stay on a path that follows it.

8.27 *Path Guidance*: Guiding the player toward the optimal path. E.g., In TrackMania Turbo [L35], players race against a phantom car of the same color, representing the optimal path.

CT9 Memory

Definition: Memorizing and/or retrieving sets of items, sequences, and/or mappings.

External Assistance Techniques for Memory

9.28 *Real-Time Reminders*: Actively reminding the player of information to be recalled. E.g., In Pokémon Leaf Green [L21], players are often reminded of details they need to complete actions in the game.

9.29 *Information Archive*: A store of relevant information that may need to be recalled. E.g., The Witcher 3 [L14] has a Bestiary Guide that keeps track of all the monsters and creatures the player has encountered in the game so far. It includes information such as monster descriptions, and weaknesses.

9.30 *Announcements*: Highlighting relevant information, reinforcing the fact that they may need to be recalled in the future. In Professor Layton and the Curious Village [L32], important game events are automatically documented in a journal, which can be reviewed later to inform about future challenges.

CT10 Spatial Memory

Definition: Remembering the location of items in a space without persistent visual cues.

External Assistance Techniques for Spatial Memory

10.31 *Maps*: A visual representation of the game environment. E.g., Many open world RPGs (like The Witcher 3 [L14] or World of Warcraft [L7]) provide both a detailed world map and a mini-map.

10.32 *Markers*: Markers in the game environment to inform the player of a location. E.g., When playing a healer in Overwatch [L9], the locations of friendlies are visible in the environment (even through obstacles), so that the healer can easily find them.

Applying External Assistance Techniques to Existing Games

In this work, we propose the use of external assistance techniques in games to aid in balancing challenge. The idea is that by understanding the core tasks that exist in a game, using our design framework, a logical set of starting points can be identified for consideration as assistance. Most, if not all, of the techniques we examined should be well-known to game designers, and, many of them would be expected as a part of the functionality of any good modern game. For example, it would be hard to imagine a large open-world RPG without a map feature. However, the insight is that further assistance can be offered to a game by providing additional assistance techniques, or stronger assistance versions of already used techniques. Below we enumerate a basic process for applying assistance techniques to a game using our newly developed framework.

Step 1: Core Task Identification

The first step is to identify the core tasks within the game. To do this we can examine each of the high-level actions in detail and try to describe them directly through associating them with one or more core tasks. For example, "Shooting a gun" requires the player to *point* and click, "Collecting items" may include the player moving (*Steering*) to the item location (*Spatial Memory*) and then finding the item within an environment with many distractor objects (*Visual Search*).

Step 2: Assistance Goals

To choose an appropriate technique (or to develop a new one), we first need to identify the goal of the assistance. For example, answers are needed to questions such as "Does this part of the game need assisting?", "How much easier should it be?" and "What aspect of this action needs assisting?" For the last question, we are referring to player actions that consist of several core tasks like shooting a ranged weapon that may require *aiming* and *reaction time*. A further example could be, if a game requires collecting items as the main challenge, like in Animal Crossing [L41], then a technique that provides a weaker amount of assistance for tasks like *visual search* or *reaction time* might be chosen. However, if the item collection is a part of a looting system after a difficult boss encounter, such as the monsters from Monster Hunter: World [L5], then a stronger implementation would make sense for *visual search* since the players have beaten the challenge and picking up the reward should not be difficult at all.

Step 3: Selecting Techniques

When selecting a technique, two important considerations exist that help focus which technique will best match the core task and the assistance goals: *Theme* and *Presentation*. Theme refers to whether and how the assistance technique can fit into the theme of the game. Some techniques may be harder to implement than others, based on the type of game, as there can be a fundamental mismatch between a game theme and a technique. For instance, a basketball game could implement some of the *aiming* techniques to help players score. However, using Projectile Magnetism may not fit, since moving the ball in mid-air would be strange.

Presentation refers to how it will be made available in the game. Will it be available by default, will it be optionally activated by players, or will it be dynamic (i.e., only made available when the system determines that it is needed)?

Step 4: Calibration and Play Testing

Many EA techniques need to be calibrated, so that they provide the right level of help. For many of the techniques, the need for calibration is self-explanatory (such as how strong a Target Gravity effect should be). Calibration is important because simply implementing a technique may not be enough to reach the desired goal, or perhaps it could be too much if a degree of challenge and difficulty is still desired. To fine tune the implementation of a technique, playtesting can be done to see how it affects player performance. Further, as we will see in the evaluation of the example assistance techniques below, we found through testing that one technique did not perform as well as expected. Assistance techniques, like other elements of a game, need to be extensively play tested.

DEMONSTRATION OF APPLYING EXTERNAL ASSISTANCE TECHNIQUES

To demonstrate how our framework can be applied to balancing challenge in a variety of games, we developed three games that all shared a common core task: *Visual Search.* We pre-selected three different techniques and implemented them in each of the three games: Highlights, Target Details, and Compass. The goal of this demonstration was to provide a concrete illustration of how the idea of identifying core tasks is an effective strategy to help guide the selection of an EA techniques, and that different assistance techniques can effectively assist a single core task. We chose to demonstrate three separate EA techniques to highlight their diversity and their application in a range of games. This allowed us to demonstrate how different EA techniques can vary in their appropriateness for different game designs. Visual search was selected as the core task for the demonstration because it has not been closely examined in previous research in assistance techniques or as a target of game balancing activities.

Games Developed

The first game seen in **Figure 2** (Left) is a simple hidden-object game with a top-down view and is similar to aspects of play in many puzzle or point-and-click games. The second game is a simple sniper simulation shown in **Figure 2** (middle). A "sniper" mode is a common element of many action games. The third game, found in **Figure 2** (Right), is a third-person view game with target objects that must be found throughout the environment, which is similar to finding and collecting items in many RPGs.

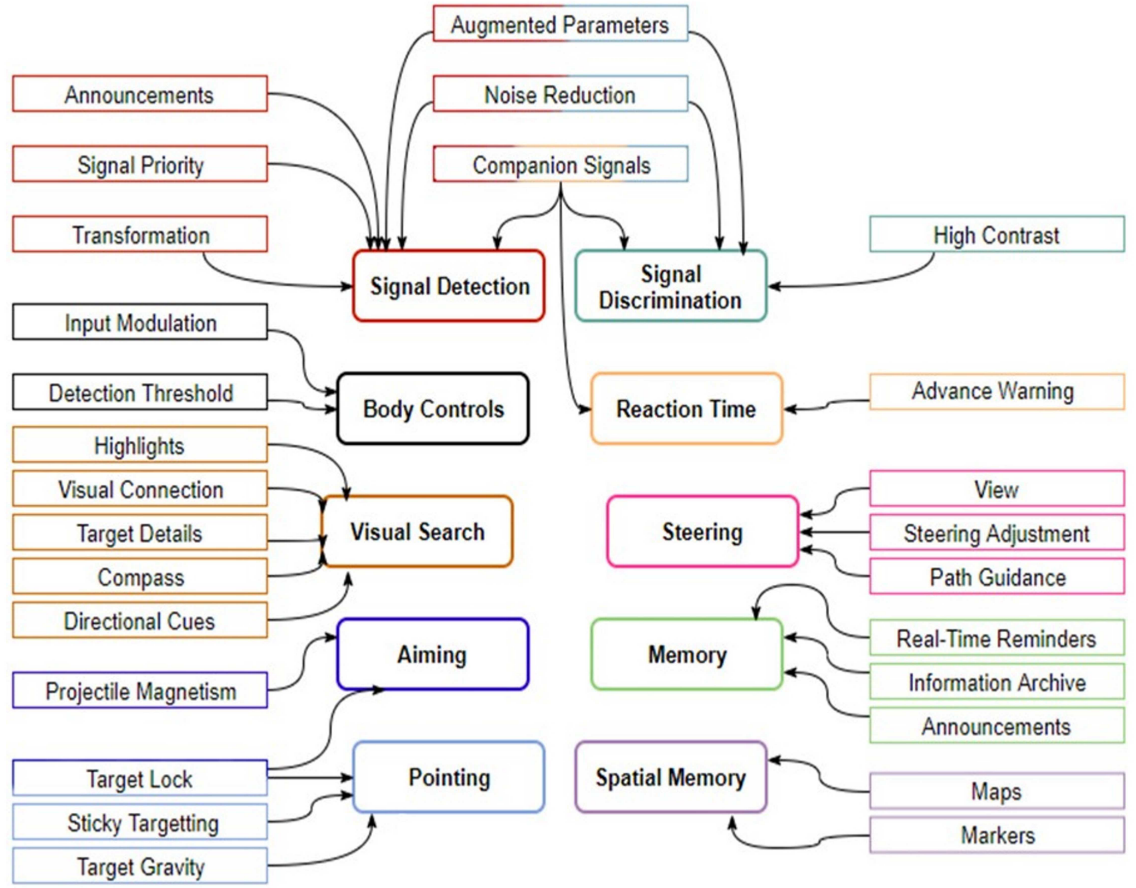


FIGURE 1 | The resulting descriptive framework of Core Tasks (rounded corners) and External Assistance Techniques (square corners). Arrows connect assistance techniques to the core tasks they facilitate.

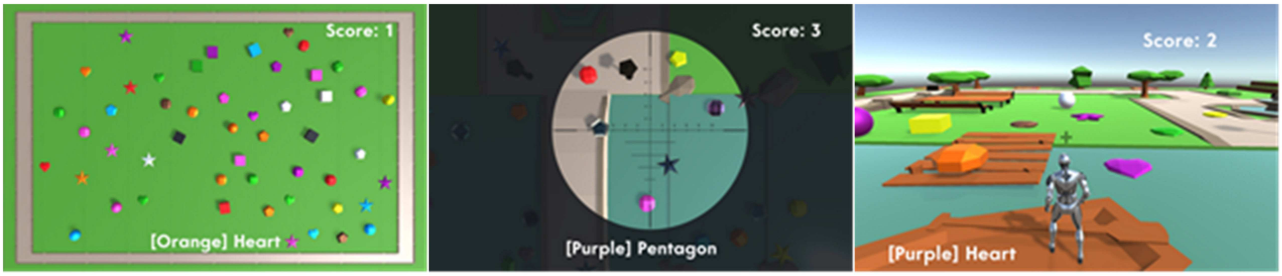


FIGURE 2 | Three games focused on the same core task (Visual Search) developed to demonstrate the application of assistance techniques. (**Left**) A hidden-object game with a top-down view. (**Middle**) A sniper game with a first-person view. (**Right**) A simple adventure game with a third-person view.

The games are not only differentiated by style and view, but also by the other core tasks they require. The top-down, object search game also requires *pointing*. The third-person game (**Figure 2**, right) has elements of *steering* for moving the character and *spatial memory* for remembering where a player has already looked. The sniper game (**Figure 2**, middle) also has an element of *signal discrimination,* as objects can look similar from the first-person view from the gun.

In each game, the target the player needs to select is given in text—centered near the bottom of the screen. Their score is in the top right corner. A correct click grants the player two points, while clicking on any other shape removes one point. The targets used in all three games are simple shapes such as

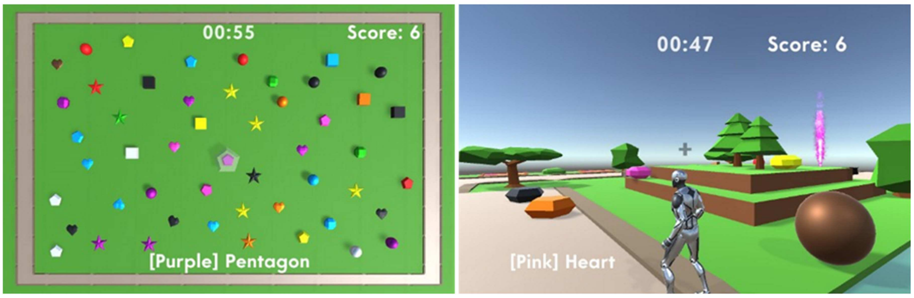


FIGURE 3 | The Highlights assistance technique. A highlighted pink pentagon in the top-down game (**left**), and in the third-person game (rising pink smoke in the distance; **right**).

FIGURE 4 | The target details technique. A large preview hint of the target shape to find in the top-down game (**left**), and in the third-person game (**right**).

FIGURE 5 | The Compass technique. An arrow above appears above the target shape to find, directing the player toward the target.

cubes, spheres, stars, and hearts. However, these could take the form of anything that may be relevant to the game's theme. For example, in **Figure 2**, the targets could be some in-game loot such as weapons or gold like those from Diablo 3 [L8]. Snipers are usually used to hit moving targets—enemies—such as in Sniper Elite [L45].

Evaluation

We conducted an evaluation with 16 players. Our evaluation had two main purposes. First, we aimed to provide a demonstration that our framework can lead to techniques that are actually effective at adjusting the difficulty of a game (i.e., our framework can guide the selection of techniques that actually make a game easier). Second, and more importantly, it allowed us to collect players' views on the specific EA techniques implemented in the games. Thus, our evaluation allows us to evidence the main concerns and details that designers might confront when applying EA techniques in practice.

To limit our study length to ∼1 h, our study used two of the three games that we implemented: Top-down view (**Figure 2**, left), and third-person (**Figure 2**, right). These two games were chosen as they are most distinct from one another, allowing players to comment on how the EA techniques vary across different game designs.

Our experiment was a 4 × 2 within-subject design with assistance technique (highlights, target details, compass, no assistance) and game type (top-down and third-person) as independent variables. Our design allowed us to investigate how

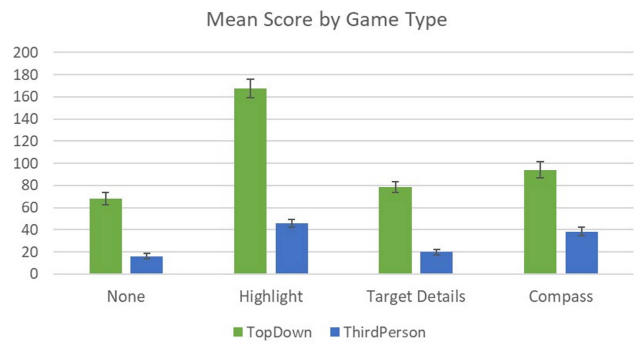


FIGURE 6 | Mean player score (in points) achieved after two minutes of gameplay for each game and technique combination.

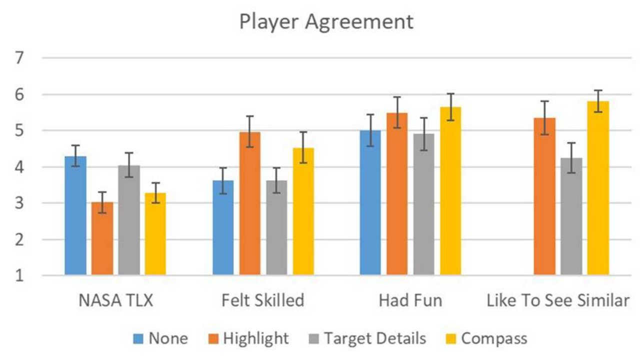


FIGURE 7 | Agreement ratings by Assistance Technique. NASA TLX scores lower is better, other ratings higher is better.

different techniques perform in different game designs, both from a balancing point of view (through objective in-game performance data) and an experience point of view (through subjective responses to questionnaires).

External Assistance Techniques Applied to Games

Below we describe how we developed each of the three external assistance techniques for the Visual Search core task in each of the three games.

Highlights: For the top-down game, highlights were implemented as a "glow" effect around the target, as seen in **Figure 3** (left). For the third-person version, Highlights were redesigned with the players' limited view in mind, and since the glow technique would not be particularly helpful when a target was far away from the player. The Highlights implementation is a beam of colored particles shooting upwards from the target (**Figure 3**, right). These are like the colored markers found in Fortnite [L20].

Target Details: Target Details gives the player more information about their target. Normally, only the name and color are given to the player in the form of text. Our implementation gives players an image of their target, which helps them avoid the need to visualize the target and enables quicker comparisons with what they see on screen. The implementation for both games is nearly identical, except for the location of the hint image; see **Figure 4**.

Compass: A 3D pointer was implemented that is rotated to point toward the next target. The rotation was updated in real-time and moved once a new target was generated, or the player moved (in the third-person view). The compass was placed at the bottom-center of the screen in the top-down version of the game (**Figure 5**, left), and under the playable character in the third-person view (**Figure 5**, right).

Apparatus and Setup

The games were built using the Unity 3D game engine. Game sessions were played on a 64-bit Windows 10 machine, Intel Core i7 CPU and a 19-inch, 1440 × 900 monitor.

Participants

We recruited 16 participants to play the games. Participants ranged in age from 18 to 43 years old (M = 25.375, SD = 6.889), 4 identified as female and 12 as male; 12 were university students. Participants had a wide range of gaming experience, but all had played games using a mouse and keyboard. Most participants preferred the WASD (62.5%) input scheme over arrow keys (12.5%), and the remainder had no preference. Participants that were not familiar with WASD input spent more time in Practice Mode for the third-person view game to get comfortable with the controls.

Procedure

The evaluation required ∼60 min to complete. The game consisted of completing the experimental task 8 times (see below), once for every combination of the independent variables: 4 "technique" levels (no assistance, compass, highlight, and hint) and 2 "game type" levels (top-down and third-person). The presentation of techniques was balanced between participants using an 8 × 8 Latin square.

Before beginning a play-through, participants were given a brief introduction to each new technique and new game, and provided an opportunity to get accustomed to the combination by playing the game in a practice mode. Participants were informed that they could take as much time as necessary to get comfortable with a specific technique in a specific game. Overall, training required ∼10 min per participant.

Participants were asked to complete a demographics questionnaire and subjective questionnaires as follows. The demographics questionnaire was completed at the start of the experiment and collected basic information (age, occupation, etc.) as well as experience with video games and relevant game controls schemes (e.g., the WASD controls that were used in our third-person game). Subjective measures were collected via a brief questionnaire presented after each technique, and a final questionnaire asking participants to reflect on their experiences was completed after gameplay. All questionnaires used Likert-style scales. Post-technique questionnaires collected ratings on their experience and the NASA TLX (to capture cognitive effort/task loading); the individual questionnaire items are presented with the results (section Subjective Measures). The final questionnaire asked participants to compare techniques as to their appropriateness for use in games.

Our experiment was approved by the Research Ethics Board of the University of New Brunswick.

Experimental Task

For each run of the game, participants were given 2 min to locate and click on as many targets as possible amongst a field of distractors. Participants were instructed to score as many points as possible in the 2 min. The amount of time remaining in the game was displayed in the upper left corner during the experiment. We selected 2 min as the time to play each version of the games through piloting with members of our research group not involved in the research. We found that 2 min provided more than enough time to experience the game without fatiguing participants, and allowed the full experiment to be completed comfortably within 1 h.

Recall that players score two points for successful selections and lose one point for incorrect selections. After a successful click, all the shapes in the game scene are randomly regenerated (to ensure that the game required visual search and not spatial memory), and a new target was presented.

Given the amount of practice that participants had and the relatively simple gameplay, 2 min provided more than enough time per game, and ensured that players stayed engaged throughout the experiment. We note that participants provided a consistent level of effort throughout the experiment. We often observed participants racing against the clock to hit one more target, and a few were visibly upset if they were unable to get their final target just as their time ran out.

Results

To assess the effectiveness of the techniques we considered the score participants achieved in each of the games, since maximizing score was their main goal in the experiment.

External Assistance Technique Performance

The grand mean score for all games was 66.0 points (sd = 50.5). The Top-Down game had the higher mean score of 102.1 (sd = 46.7), while Third-Person had the lower mean score at 30.0 points (sd = 17.7). We conducted a two-way ANOVA analysis with within-subject factors (technique and game type) to analyze the data. The main effect of Game Type on player score was statistically significant [$F_{(1,15)} = 235.4$, $p < 0.01$], see **Figure 6**. This difference is unsurprising, since the third-person game requires players to spend time navigating to the next target.

Examining the assistance techniques (independent of game type) on player score, we found that the highest mean score was achieved with Highlights, at 106.6 points, followed by Compass with 68.0 points, and finally Target Details and No Assistance came near the bottom with 49.2 and 42.1 points, respectively. The main effect of assistance techniques was also significant [$F_{(3,45)} = 107.2$, $p < 0.01$], see **Figure 6**.

There was a significant Game Type × Assistance Technique interaction effect [$F_{(3,45)} = 50.7$, $p < 0.01$], which was due mainly to the significant differences between Highlight and the other techniques for the Top-Down game and nearly all the pairs for Third-Person (except for None/Hint and Highlight/Compass, which were not significant; see **Figure 6**), as determined by a Scheffé *post-hoc* analysis.

Subjective Measures

To assess players' views on the use of assistance techniques in the games, and in general, participants filled out several questionnaires. We analyzed the results of both games together, since the techniques were rated similarly in both game conditions. The chart in **Figure 7** shows the mean agreement with Likert-style ratings for a number of statements (see below). All ratings were on a 7-point scale; 1 = strongly disagree, 4 = neutral, 7 = strongly agree.

First, participants' ratings on NASA Task Load Index (TLX) items were aggregated into one value representing "Task Loading" (lower is better) (Hart, 2006). The other three ratings are mean values of agreement with the following statements: "I felt skilled at this game," "I had fun playing this game," and "I would like to see similar assistance techniques implemented in games I play." See **Figure 7**.

We used a Friedman's test to detect differences in rating between each of the four Technique levels. Overall, the same trend can be seen across each item. There was a significant main effect of Assistance Technique on Task Loading—TLX ($\chi^2 = 25.253$, $p < 0.0001$, $df = 3$), Feeling Skilled ($\chi^2 = 22.829$, $p < 0.0001$, $df = 3$), Having Fun ($\chi^2 = 10.602$, $p < 0.05$, $df = 3$), and Liking to See a Similar Technique in Games Played ($\chi^2 = 25.196$, $p < 0.0001$, $df = 3$). *Post-hoc* tests show there were significant differences ($p < 0.05$) for the same pairs in all tests. All pairs were significantly different with the exception of Highlight & Compass, and Target Details & None.

Best Technique for Game: After playing each game, players responded to the question: "Which technique did you feel best fit with this type of game?" For the Top-Down game, of the 16 participants, 7 (44%) felt the Compass technique was most appropriate, 6 (38%) chose Highlights, and 3 (19%) felt that Target Details was best.

When asked about the most suitable assistance technique for the Third-Person game, of the 16 participants, there was an even split, eight felt that the Highlights was the most appropriate and eight felt that compass was most appropriate; none felt that Target Details was best for the Third-Person game. When asked to justify their choice for the best technique for the game, some participants described selecting Highlights because it helped them perform the best. However, others pointed out the Highlights made the game too easy, which is why they chose Compass over Highlights.

Overall, participants found the preview available in the Target Details technique not as helpful. From informal discussions after the experiments, participants attributed this to the fact that the visual representation (a larger, semi-transparent shape) was not always similar to the target they were looking for. In the top-down game, the preview was a closer representation of what they were looking for (e.g., consider the blue star Highlight and the target shape that can be seen in **Figure 4**, left). However, participants found Target Details less helpful in the Third-Person game. This is due to the target often being differently orientated than in the game (note the different appearance of the "Orange Octagon" using Highlights in **Figure 4**, right, and the corresponding target in **Figure 2**, right).

Free-form Questions: We also asked participants several free-form questions at the end of the experiment to solicit their opinions after playing all of the games with the different techniques and to gauge their feeling on the use of EA techniques in games.

All 16 participants felt that EA techniques like the ones they used are well-suited for inclusion in other games. Some players provided further detail as to why, explaining, such as P1 who said, "… some games have very good potential but discourages player by not offering help to proceed when the player is stuck." Other participants agreed but provided some conditions for how Assistance Techniques should be provided, for example saying that "… a toggle menu allowing players to turn on or off the techniques would be beneficial" (P15). When asked, all participants felt they would like to have control of Assistance Techniques and to be able to turn them off, if they wanted to.

When considering whether they would like to know about how different techniques are being used to assist them in games they played, players were a little more split. Most of the participants (75%) felt that they would like to know. However, others felt they would prefer not to know. P15 highlighted this sentiment by pointing out that "Knowing that [an Assistance Technique] was there and performing poorly would not be beneficial as you would feel as though even with assistance, you could not perform well in game."

Summary of Study Results

Our study provides important findings that can help in designing and applying EA techniques.

1. For the previously under-studied core task of Visual Search in games, all techniques helped players perform better and have a better play experience (except for Highlights in the Third-Person game).
2. In general, as an Assistance Technique provided more of a boost to performance, players found the task easier (lower TLX scores), felt more skilled, had more fun, and were more in favor of seeing the technique in a game.
3. Participants rated techniques differently depending on the game: Highlights was rated best for the Top-Down game, and Highlights and Compass were rated best for the Third-Person game.
4. Participants want control of Assistance Techniques, through the ability to toggle them on or off.
5. Revealing the use of Assistance Techniques to provide success should be done carefully, since players might interpret not reaching goals as a personal failure.

DISCUSSION

The focus of our work was not the evaluation of previously understudied EA techniques, but rather the presentation of our framework of EA techniques. Our games and user evaluation were performed to exemplify the use of EA techniques in practice, including how techniques were adapted to different game types and genres and how different techniques can lead to different experiences. Below we elaborate on how the findings from our study inform the application of EA techniques and propose future work to refine and extend our framework.

Explanation for Results of the Study

Performance in the two games was different. While both games were based on a Visual Search core task, and involved clicking on shapes, they played like very different games. The games, while simple, were representative of many common real world games, and the fact that players score differently between the games, with or without assistance suggests that our goal of making different games was effective. We note that all players rated the games as fun to play, regardless of Assistance Technique, and that the techniques improved play experience.

As we expected, participants enjoyed the games more the better they performed. As they found the game easier, they felt more skilled, had more fun, and rated their desire to have a particular technique in their game higher as their performance increased. However, we also found strong evidence that when a game was too easy, it started to become less fun for many players. We can identify exactly when this happened for players in our study, due to the interaction effect detected between Game Type and Assistance Technique, which we can attribute largely to the disproportionate performance of Highlights in the top-down game (see **Figure 6**). In this case, Highlight provided such strong assistance that it removed most of the challenge of the game, meaning that players found the game less fun. This speaks to the challenge of balancing games, with any approach; while our Assistance Techniques were added after the development of the basic game, they still need to be balanced themselves. The hope, however, for external assistance techniques is that since they operate outside of the main in-game mechanics, this balancing activity is greatly simplified since it can work independently of those other mechanics.

We believe that in considering adding any type of assistance to a game, like all other types of game mechanics, it must be done carefully. The very idea of an "Assistance Technique" might suggest to some that there is an expectation of success once assistance has been given. As we saw from on participant's comments, they felt knowing that assistance was being given and still failing might make them feel badly about their abilities. At the same time, in many game designs, failure is an important part of the play experience; we play to be challenged, to accomplish our goals in a safe space, and with this in mind some degree of failure is needed to make the experience meaningful (Juul, 2013). For players who are highly skilled, providing assistance when it removes challenge or reduces the chance of failure might be akin to attaching training wheels to the bicycle of a skilled rider, and might even harm their perceived competence (Wiemeyer et al., 2016). Further work needs to be done to understand the intricacies of providing assistance and how it affects play experience.

Our analysis did not provide a deeper dive into the reasoning behind players' preferences for certain techniques. For example, it is possible that players preferred techniques that they had more familiarity with, rather than the ones that made universally the "best" game. While we do not believe this to be the case, future work should consider such potential factors. As in previous research on assistance techniques, we have found further evidence that players want control over the use of assistance, yet do not necessarily want to be reminded that it is being provided (Bateman et al., 2011b). Revealing the use of assistance techniques for balancing challenge (or competition between players Vicencio-Moreira et al., 2015) should be done carefully.

How the Study Results Inform Game Designers About Incorporating External Assistance Technique Into Games

Our implementations and evaluation demonstrated the utility of the framework for adapting EA techniques to games to provide significant changes in game performance, which we consider. We provided a step-by-step process for adapting EA techniques to existing games, where the first step, "Core Task Identification," is a critical step in determining how a game can be assisted. In our evaluation, the games were comprised of just a few core tasks: Visual Search and Pointing for both games, and, additionally, Steering in the Third-Person game.

Once the core tasks are identified, the appropriate assistance can be chosen to adjust player performance. If the desired effect or difficulty is not achieved, designers may then consider providing assistance to the other tasks in the game. For example, in our evaluation, players could have further benefitted from a Pointing assistance technique to improve their scores even further.

Game designers need to play test their techniques carefully. Sometimes our evaluation results followed conventional wisdom, but sometimes we found unexpected results. For example, Compass performed nearly as well as Highlights, especially in Third-Person, which we did not expect.

Unsurprisingly, but importantly, player preference is not always about performance. EA techniques can make certain tasks too easy, as many participants felt about Highlights. Even though it was effective and increased the players' scores considerably, players often preferred the Compass, which did not have as strong an effect as Highlights. Designers should keep this in mind when deciding when and where to implement certain techniques and consider the level of challenge that is desired.

Importantly, however, our work is fundamentally limited because it did not directly involve a broad set of game designers. While our first author has experience as a game designer, and we leveraged relevant experience reports from designers (in the cited articles from Gamasutra), we have little evidence still to the utility of our framework in actual practice with larger and more complex game designs. In our future work, we would very much like to discuss our framework in an interview study with practicing game designers to understand its utility to them, and how it might actually fit into their design practices.

Future Work

This work provides several new directions for research. Assistance techniques for games have been investigated for a number of years in the HCI community. However, previous work has often focused on input assistance (working at the level of input for steering, pointing, and aiming). In this work we identify a number of understudied ways in which assistance can be provided in games (by beginning with the game's Core Tasks). Future work should confirm the effectiveness of our process for adapting EA techniques to games, both when applying existing techniques and in developing completely new techniques.

More basic research is needed, looking at EA techniques and how they can impact play, in a wider range of game types. By relating new work back to the concept of core tasks, we will get a consistent organizational concept for identifying new directions and understanding performance at a fundamental level, and how techniques impact other aspects of play such as skill development (Gutwin et al., 2016).

Our work focuses on the idea of core tasks from the work of Flatla et al. (2011) to help focus and narrow the mechanics that we looked at. Core tasks are the basic motor and perceptual tasks that are needed to interact with common game mechanics; however, the skills corresponding to core tasks only make up a small subset of the larger sets of skills players might need in games. We believe that focusing for our initial research in this area was a necessary step to make our work tractable. Future work should consider Schell's broader characterization of skills in games (Schell, 2019), which include social (e.g., building trust and relationships) and mental skills (e.g., establishing plans and strategy), as a starting point to identify very different but important skills that EA techniques can target to improve balance.

There is also a wider range of research that can likely be drawn upon and further exemplify EA techniques. In our work, we looked at a wide range of techniques that could be considered as assistance. However, different examples of EA techniques might emerge, and depending on the focus of any process creating a framework, different granularities of concepts and organizing principles will be developed. For example, the work of Alves and Roque (2010) provide a comprehensive list of "sound design patterns" that can help support game designers in developing sound to support their games. Two of the authors have informally discussed all 78 patterns and believe that roughly a third of these could be considered as EA techniques. For example, the "Imminent Death" sound pattern would be an example of a "Companion Signal" in our framework. Of the techniques that might be considered as EA techniques, we believe they represent specific examples of the "Signal Detection," "Signal Discrimination," and "Path Guidance" techniques.

So, while other specific examples of EA techniques might exist, it seems that they fit well into the categories of EA techniques that we identified. The informal exercise described above helps reassure us that our framework provided good generalizability, but that there are likely many examples of the techniques that could help designers identify specific adaptations of a technique for their games. To this end we hope to follow the lead of Alves and Roque, and develop materials that help make concrete examples of EA techniques in games more accessible, similar to Alves and Roque's sound design cards (Alves and Roque, 2011) and companion website (www.soundingames.com). While we believe our list of core tasks has good utility, in the future it is likely that new technological developments may lead to the need for changes and refinements to our initial list. For example, Body Controls is currently a comprehensive category that includes Muscle Activation, Ambidexterity and Movement. These subcategories may become more distinct as games begin to take advantage of body input, especially with the advancements made in Virtual and Augmented Reality technologies (Foxlin et al., 1998).

CONCLUSION

In this paper, we studied 54 games using a grounded theory study, allowing us to identify a framework of 10 different core tasks commonly needed in games, and 27 possible external assistance techniques that can make them easier to complete. Several of those techniques have been previously studied, while others are still to be explored and evaluated.

By organizing video game assistance at a fundamental level, through the lens of core tasks, we assist in the portability and understanding of these techniques across games, regardless of genre or platform. One of the main goals of this work was to create a comprehensive starting point for designers and game developers considering assistance for their games. We have successfully collected and presented a wide range of assistance techniques, exemplifying them and providing clear new language for discussing them.

We also conducted a study on the effectiveness of several techniques pertaining to a previously under-studied core task in games, Visual Search. We evaluated the effectiveness of three techniques (Highlights, Target Details, and Compass) in two different games that share Visual Search as a core task. Our findings show that the techniques improve performance and are suitable for balancing challenge.

In this paper, we provide the first generalization of how the range of core tasks can be assisted in games. Our work gives designers a new language for discussing external assistance techniques and an important starting point for making important, and common design decisions in order to target appropriate level of challenge in their games. Further, we provide a general methodology that can be used in future research that studies and characterizes techniques that can designers can employ in targeting a desired play experience.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Research Ethics Committee University of New Brunswick. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

The research was conceived of mutually by JR, SB, and MF. JR led data collection and analysis with the assistance of SB and MF. JR and SB contributed equally to the writing with assistance from MF.

FUNDING

This research was supported by NSERC and the New Brunswick Innovation Foundation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomp.2020.00017/full#supplementary-material

REFERENCES

Adams, E. (2013). *Fundamentals of Game Design*, *3rd Edn*. Berkeley, CA: New Riders.

Alharthi, S. A., Alsaedi, O., Toups, Z. O., Tanenbaum, J., and Hammer, J. (2018). "Playing to wait: a taxonomy of idle games," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems CHI'18* (Montreal, QC: Association for Computing Machinery), 1–15.

Alves, V., and Roque, L. (2010). "A pattern language for sound design in games," in *Proceedings of the 5th Audio Mostly Conference: A Conference on Interaction With Sound AM'10* (Piteå: Association for Computing Machinery), 1–8.

Alves, V., and Roque, L. (2011). "A deck for sound design in games: enhancements based on a design exercise," in *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology ACE'11* (Lisbon: Association for Computing Machinery), 1–8.

Auto-Aim (Concept). (2019). Giant Bomb. Available online at: https://www.giantbomb.com/auto-aim/3015-145/ (accessed February 1, 2020).

Balakrishnan, R. (2004). "Beating" fitts' law: virtual enhancements for pointing facilitation. *Int. J. Hum. Comput. Stud.* 61, 857–874. doi: 10.1016/j.ijhcs.2004.09.002

Baron, S. (2012). *Cognitive Flow: The Psychology of Great Game Design*. Gamasutra. Available online at: https://www.gamasutra.com/view/feature/166972/cognitive_flow_the_psychology_of_php (accessed January 21, 2020).

Bateman, S., Doucette, A., Xiao, R., Gutwin, C., Mandryk, R. L., and Cockburn, A. (2011a). "Effects of view, input device, and track width on video game driving," in *Proceedings of Graphics Interface 2011 GI'11* (St. John's, NL: Canadian Human-Computer Communications Society), 207–214.

Bateman, S., Mandryk, R. L., Stach, T., and Gutwin, C. (2011b). "Target assistance for subtly balancing competitive play," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI'11* (New York, NY: ACM), 2355–2364.

Bourtos, D. (2008). *Difficulty is Difficult: Designing for Hard Modes in Games*. Gamasutra. Available online at: https://www.gamasutra.com/view/feature/132181/difficulty_is_difficult_designing_.php (accessed January 21, 2020).

Burgun, K. (2011). *Understanding Balance in Video Games*. Gamasutra. Available online at: https://www.gamasutra.com/view/feature/134768/understanding_balance_in_video_.php (accessed April 7, 2020).

Cechanowicz, J. E., Gutwin, C., Bateman, S., Mandryk, R., and Stavness, I. (2014). "Improving player balancing in racing games," in *Proceedings of the First ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play CHI PLAY'14* (Toronto, ON: Association for Computing Machinery), 47–56.

Chen, J. (2007). Flow in games (and everything else). *Commun. ACM* 50, 31–34. doi: 10.1145/1232743.1232769

Csikszentmihalyi, M. (2008). *Flow: The Psychology of Optimal Experience*, New York, NY: Harper Perennial Modern Classics.

Denisova, A., Guckelsberger, C., and Zendle, D. (2017). "Challenge in digital games: towards developing a measurement tool," in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems CHI EA'17* (Denver, CO: Association for Computing Machinery), 2511–2519.

Felder, D. (2015a). *Design 101: Balancing Games*. Available online at: https://www.gamasutra.com/blogs/DanFelder/20151012/251443/Design_101_Balancing_Games.php (accessed February 1, 2020).

Felder, D. (2015b). *Design 101: Playtesting*. Available online at: https://www.gamasutra.com/blogs/DanFelder/20151105/258034/Design_101_Playtesting.php (accessed February 1, 2020).

Flatla, D. R., Gutwin, C., Nacke, L. E., Bateman, S., and Mandryk, R. L. (2011). "Calibration games: making calibration tasks enjoyable by adding motivating game elements," in *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology UIST'11* (New York, NY: ACM), 403–412.

Foxlin, E., Harrington, M., and Pfeifer, G. (1998). "Constellation: a wide-range wireless motion-tracking system for augmented reality and virtual set applications," in *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques SIGGRAPH'98* (New York, NY: Association for Computing Machinery), 371–378.

Gerling, K. M., Miller, M., Mandryk, R. L., Birk, M. V., and Smeddinck, J. D. (2014). "Effects of balancing for physical abilities on player performance, experience and self-esteem in exergames," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI'14* (New York, NY: ACM), 2201–2210.

Glaser, B. G. (1998). *Doing Grounded Theory: Issues and Discussions*. Mill Valley, CA: Sociology Press.

Glaser, B. G., and Strauss, A. L. (2017). *Discovery of Grounded* Theory: *Strategies for Qualitative Research*. Mill Valley, CA: Routledge.

Goldkuhl, G., and Cronholm, S. (2010). Adding theoretical grounding to grounded theory: toward multi-grounded theory. *Int. J. Qual. Methods* 9, 187–205. doi: 10.1177/160940691000900205

Gutwin, C., Vicencio-Moreira, R., and Mandryk, R. L. (2016). "Does helping hurt? Aiming assistance and skill development in a first-person shooter game," in *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play CHI PLAY'16* (Austin, TX: Association for Computing Machinery), 338–349.

Hart, S. G. (2006). Nasa-task load index (NASA-TLX); 20 years later. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 50, 904–908. doi: 10.1177/154193120605000909

Johanson, C., and Mandryk, R. L. (2016). "Scaffolding player location awareness through audio cues in first-person shooters," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems CHI'16* (San Jose, CA: Association for Computing Machinery), 3450–3461.

Juul, J. (2013). *The Art of Failure: An Essay on the Pain of Playing Video Games*. Cambridge, MA: MIT Press.

Lyons, K. (2016). "Visual parameters impacting reaction times on smartwatches," in *Proceedings of the 18th International Conference on Human-Computer Interaction With Mobile Devices and Services MobileHCI'16* (Florence: Association for Computing Machinery), 190–194.

MacKenzie, I. S. (2013). *Human-Computer Interaction: An Empirical Research Perspective*, *1st Edn.* Amsterdam: Morgan Kaufmann is an imprint of Elsevier.

Newell, A. (1994). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Renner, P., and Pfeiffer, T. (2017). "Evaluation of attention guiding techniques for augmented reality-based assistance in picking and assembly tasks," in *Proceedings of the 22nd International Conference on Intelligent User Interfaces Companion IUI'17 Companion* (Limassol: Association for Computing Machinery), 89–92.

Salen, K., and Zimmerman, E. (2003). *Rules of Play: Game Design Fundamentals.* Cambridge, MA: The MIT Press.

Schell, J. (2019). *The Art of Game Design: A Book of Lenses, 3rd Edn*. Boca Raton, FL: A K Peters/CRC Press.

Schmidt, R., Lee, T., Winstein, C., Wulf, G., and Zelaznik, H. (2018). *Motor Control and Learning 6th Edition With Web Resource: A Behavioral Emphasis*, *6th Edn*. Champaign, IL: Human Kinetics.

Tabor, A., Bateman, S., Scheme, E., Flatla, D. R., and Gerling, K. (2017). "Designing game-based myoelectric prosthesis training," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems CHI'17* (Denver, CO: Association for Computing Machinery), 1352–1363.

Toups, Z. O., Hammer, J., Hamilton, W. A., Jarrah, A., Graves, W., and Garretson, O. (2014). "A framework for cooperative communication game mechanics from grounded theory," in *Proceedings of the First ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play CHI PLAY'14* (Toronto, ON: Association for Computing Machinery), 257–266.

Vazquez, R. (2011). *How Tough Is Your Game? Creating Difficulty Graphs.* Gamasutra. Available online at: https://www.gamasutra.com/view/feature/134917/how_tough_is_your_game_creating_.php (accessed January 21, 2020).

Vicencio-Moreira, R., Mandryk, R. L., and Gutwin, C. (2015). "Now you can compete with anyone: balancing players of different skill levels in a first-person shooter game," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems CHI'15* (New York, NY: ACM), 2255–2264.

Vicencio-Moreira, R., Mandryk, R. L., Gutwin, C., and Bateman, S. (2014). "The effectiveness (or lack thereof) of aim-assist techniques in first-person shooter games," in *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems CHI'14* (New York, NY: ACM), 937–946.

Weihs, N. (2013). *Techniques for Building Aim Assist in Console Shooters*. GDC Vault. Available online at: https://www.gdcvault.com/play/1017942/Techniques-for-Building-Aim-Assist (accessed February 1, 2020).

Wiemeyer, J., Nacke, L., Moser, C., and 'Floyd' Mueller, F. (2016). "Player Experience," in *Serious Games: Foundations, Concepts and Practice*, eds R. Dörner, S. Göbel, W. Effelsberg, and J. Wiemeyer (Cham: Springer International Publishing), 243–271.

Wikipedia, A. (2020). *List of Video Game Genres - Wikipedi*a. Wikipedia. Available online at: https://en.wikipedia.org/wiki/List_of_video_game_genres (accessed January 21, 2020).

Wuertz, J., Alharthi, S. A., Hamilton, W. A., Bateman, S., Gutwin, C., Tang, A., et al. (2018). "A design framework for awareness cues in distributed multiplayer games," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (ACM), 243.

Conflict of Interest: At the time of preparing this manuscript JR was employed by Kabam Games. However, all research described was conducted while JR was a graduate student at the University of New Brunswick, and all work was done independently of Kabam Games.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Computer Science

ORIGINAL RESEARCH
published: 11 June 2020
doi: 10.3389/fcomp.2020.00018

What Is It Like to Be a Game?—Object Oriented Inquiry for Games Research, Design, and Evaluation

Katta Spiel[1,2,3]* and *Lennart E. Nacke*[3,4]

[1] eMedia Research Lab, KU Leuven, Leuven, Belgium, [2] Centre for Teacher Education, University of Vienna, Vienna, Austria, [3] HCI Games Group, Games Institute, University of Waterloo, Waterloo, ON, Canada, [4] Stratford School of Interaction Design and Business and Department of Communication Arts, University of Waterloo, Waterloo, ON, Canada

Human-Computer Interaction (HCI) researchers more and more challenge the notion of *technologies as objects* and *humans as subjects*. This conceptualization has led to various approaches inquiring into object perspectives within HCI. Even though the development and analysis of games and players is filled with notions of intersubjectivity, games research has yet to embrace an object oriented perspective. Through an analysis of existing methods, we show how *Object-Oriented Inquiry* offers a useful, playful, and speculative lens to pro-actively engage with and reflect on how we might know what it is like to be a game. We illustrate how to actively attend to a game's perspective as a valid position. This has the potential to not only sharpen our understanding of implicit affordances but, in turn, about our assumptions regarding play and games more generally. In a series of case studies, we apply several object-oriented methods across three methodological explorations on *becoming*, *being*, and *acting as* a game, and illustrate their usefulness for generating meaningful insights for game design and evaluation. Our work contributes to emerging object-oriented practices that acknowledge the agency of technologies within HCI at large and its games-oriented strand in particular.

Keywords: object-oriented inquiry, evaluation, speculation, design, games, play

OPEN ACCESS

Edited by:
Ioanna Iacovides,
University of York, United Kingdom

Reviewed by:
Jussi Holopainen,
University of Lincoln, United Kingdom
Michael Straeubig,
Independent Researcher, Berlin, Germany

***Correspondence:**
Katta Spiel
katta.spiel@kuleuven.be

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Computer Science

Received: *01 May 2020*
Accepted: *07 May 2020*
Published: *11 June 2020*

Citation:
Spiel K and Nacke LE (2020) What Is It Like to Be a Game?—Object Oriented Inquiry for Games Research, Design, and Evaluation. Front. Comput. Sci. 2:18. doi: 10.3389/fcomp.2020.00018

1. INTRODUCTION

Digital games evoke visceral emotions in players. Expressing these feelings often comes with tendencies of ascribing human features to games (Müller et al., 2018): games are presented as deliberately thwarting players' efforts or attributed kindness and collaborative potential. However, thins anthropomorphization asserts human modes of being onto non-human games. One approach allowing us to rigorously engage with the question of "What is it Like to Be a Game?" lies in drawing on epistemologies that allow for object-specifics inquiries.

Object- or technology-centered modes of inquiry emerged in the form of different theoretical perspectives. However, they all share that they fundamentally rethink the ontological role of the material world. Being critical of human exceptionalism, they argue for rejecting the dualism of "nature vs. culture" in favor of a relational ontology that accounts for the agency of things. In other words: They argue to take things and artifacts as well as their embedded knowledge, seriously. Technologies and humans are largely understood as fundamentally different entities, which suggests that it is possible to investigate one or the other independently. Building upon prior theoretical object-oriented work (e.g., Latour, 2005; Bogost, 2012; Hayles, 2014), we argue that human and non-human participants in play mutually shape their relationship and continuously (re-)constitute their ontological (in)separability, i.e., their state of being in the world.

Within Human-Computer Interaction (HCI), we observe a recent surge of research that puts the perspective of objects and technologies explicitly at the core of its inquiry, thus providing a counter-perspective to the strictly human-centered view (e.g., Wakkary et al., 2015; Giaccardi et al., 2016; Chang et al., 2017). However, such work is virtually non-existent in HCI related games and play research. Even though inquiries into game generated data (i.e., logs) exist and are often employed, they are predominantly framed along an explicit interest into players' experiences, effectively decentering the object in their human-centered mode of inquiry. We argue that there is a potential for game design and research practices operating from an object-oriented perspective to generate innovative ideas and insights.

To this extent, we contribute new knowledge to the field of HCI by analyzing existing object-oriented methods and use the notion of *Object-Oriented Inquiry* (OOI) by Hayles (2014) as a theoretical background for our methodological explorations. Our aim is to articulate what we can understand by attending to an object perspective. After outlining the epistemological premises of this theory, i.e. what we can learn through OOI and how it ties into existing HCI and games research. We then explore different methods for *Object-Oriented Inquiry* and exemplify the approach through three *methodological* explorations on *becoming*, *being*, and *acting as* a game. Finally, we illustrate how this approach generates additional parameters for design decisions and the evaluation of digital games. Our work contributes a playful approach to *Object-Oriented Inquiry* stemming from theoretical deliberations with the potential to expand on qualitative methods and understandings for game design practice and research.

2. BACKGROUND

Before diving into the particularities of *Object-Oriented Inquiry* (OOI), we present *Object-Oriented Ontology* (OOO) as the theoretical foundation of the approach. We then show how OOI can be applied to contexts involving digital games.

2.1. Object-Oriented Ontology

Research inquiries, in general, practically position their knowledge paradigmatically (Guba et al., 1994). With situated paradigms come implications for the ontological, epistemological and methodological backdrop thereof. Ontology (the question of "how are things in the world?"), epistemology ("what can we know about things?"), and methodology ("what kinds of methods lead to which kind of knowledge?") comprise individual research areas in their own rights. We draw on *Object-Oriented Ontology* (OOO) as our ontological background, epistemologically position our knowledge as partial and subjective (Haraway, 1988) and offer a *methodological analysis for a range of methods* aimed at identifying object-oriented knowledge in game design and research from these positions.

An early precursor for OOO can be found in Actor-Network Theory (ANT). By arguing for ontological symmetry, ANT emphasizes the agency of things and the interdependent relationships between humans and things: these show themselves in use, practice, maintenance, development, invention, and so on, continuously rearranging each other into networks of relationships (Latour, 2005). These networks are in a state of continuous malleability. ANT is based on highly detailed observations and stories of the series of interactions necessary to sustain a network (e.g., Latour et al., 1999). By placing all actors on the same level and giving them the same amount of attention, ANT lends itself toward the concept of a "flat ontology" Bryant (2011), which blurs the distinction between objects and subjects. This has been made clear particularly for immaterial objects (such as digital objects) even before ANT was formalized. Flusser states that "[t]he future culture of immaterial information (...) will hold objects in contempt: it will consume them without paying any attention to them. In this sense, the human being will no longer be subject to objects" (Flusser, 1986, p. 331), hinting at a dissolution of ontological difference between objects and humans as subjects.

The term *Object-Oriented Ontology* stems from speculative realism (Harman, 2015). It positions *things* central to philosophical inquiry and opposes the consideration that knowledge about them can be potentially absolute or perfectly controllable. OOO is part of a conglomerate of non-humanist theories that reject the different categorizations of humans and objects entirely (Forlano, 2017). The approach focuses on how to engage with objects as they present themselves (Hayles, 2014). In other words, "'objects' does not mean non-humans any more than it means humans. All entities are objects; all have an inscrutable inwardness withdrawn from direct access" (Harman, 2015, p. 407). Understanding all potential actors, including humans, as objects is not meant as a call for passivity, rather this objectification comes from a stance that explicitly seeks to understand the different shapes of agency that are possible from an object perspective (Cussins, 1996). In that regard, OOO is a call for humility in the development of knowledge, a call to be cautious before asserting the specificity of humanity and to acknowledge the material knowledge embedded in things. It cautions us to be humble about the limitations of knowledge production more generally, i.e., even when we decidedly investigate human concerns as humans, our knowledge about these matters cannot be assumed as absolute or complete.

Scholars have used OOO as the inspiration for a variety of methods and tools, e.g., to examine artifacts and digital objects (Hui et al., 2016). As a theoretical framework, it allows researchers to consciously engage with artifacts' perspectives. However, therein also lies the core limitation in that, as humans, we are inherently removed from things and limited in the ways we can inquire about them (Bogost, 2012). To do so, we have to rely on the perceptive apparatus that is available to us, and can only project our own interpretations onto the objects' representations and manifestations when we engage with them (Bryant, 2011). Hence, we cannot separate our knowledge production from human specificity and can only approximate object knowledge, if at all.

2.2. OOO for Games and Play: Object-Oriented Inquiry

We argue that game design and research benefits from an approach grounded in the work of Hayles (2014), who outlines the foundations for an object-oriented, posthuman, narrative and speculative analysis: *Object-Oriented Inquiry* (OOI). Specifically, we suggest following Hayles' engagement with resistance. "The object responds by resisting the human's inquiry, in a continuing dialectic in which the resistance forces the questions to be modified, and the modified questions uncover new forms of resistance" (Hayles, 2014, p. 169). Hayles further argues that despite this limitation, "human imagination is the best way [...] to move beyond anthropocentrism into a more nuanced understanding of the world as comprised of a multitude of world views, including those of [...] inanimate objects," because it requires scholars to be actively and creatively invested in the relation with the thing and the reactions/resistance they get from it. It argues for a decidedly different stance to think from a perspective that aims to decenter human and subject-oriented approaches.

Nonetheless, we argue that taking this perspective from our position as HCI researchers in the area of games and play allows us to encounter games within their socio-technical context on game-oriented terms. This can be understood as an extension of "staying open to multiple meanings in design and evaluation" (Sengers and Gaver, 2006) within third-wave HCI (Harrison et al., 2011) by putting the focus on the plurality of meanings, some of which might be coming from games. *Object-Oriented Inquiry* can open up further potentially conflicting perspectives on the interaction and relationships between games and players.

Games, toys, and playthings offer excellent opportunities for methodological explorations of OOI. Games are already understood as acting by themselves (Zook et al., 2011), and following their own predetermined rules, most notably in the notion of machine vs. operator actions (Galloway, 2006). When encountering games, players often become viscerally passionate and engage with them through anthropomorphization (Müller et al., 2018), implicitly acknowledging and discursively re-iterating a game's agency. Idle games even present an entire game genre that does not necessarily require player input (Alharthi et al., 2018a). They "tend to play themselves, making the player's participation optional or—in some cases—entirely redundant" (Fizek, 2018). Hence, idle games can be understood as games facilitating object-oriented play that decenters players while also facilitating distinct experiences through gameplay (Spiel et al., 2019).

Digital games have been used in a fashion which Bogost (2012), (in reference to Harman) calls *carpentry* as the act of expertly manipulating material explorations to create objects that *do* philosophy through their embodied knowledge. These are objects (sometimes games) that interrogate their environment through their being, conceptualizing "philosophy as a *practice*" (Bogost, 2012, p. 92) and providing "ontographical tools meant to characterize the diversity of being" (Bogost, 2012, p. 94). Games can provide the ideal playground to experiment with ontography: Bogost (2016) carpentered *Cow Clicker* as an investigation into the practice of supposedly social games; which is also understood as a precursor or early representative within the idle game genre. Similarly, Gualeni (2014) created *Haerfest* to philosophically engage with the question of what it might be like to experience the world as a bat (in reference to Nagel, 1974). This means, games are particularly conducive to object-oriented inquiries as artifacts that are understood as having agency more generally *and* as a medium for the carpentry of object-oriented play.

3. OBJECT-ORIENTED INQUIRY AS A PRACTICE

We reviewed existing approaches in HCI that focus on objects instead of human perspectives or the interaction between them. As work within dedicated games and play HCI research from an object-oriented perspective is exceedingly uncommon, we look to the larger field of research we are embedded in as well as to associated work in the realm of speculative design (Auger, 2013). Through our close reading (Martin, 2005) of available works and subsequent analysis thereof, we identified and classified different strands of methods for data acquisition and analysis: namely *schematic*, *narrative*, and *manipulative* inquiries augmented by *descriptive*, *discursive*, and (purely) *speculative* analysis. Additionally, we briefly discuss data (re-)presentation as a particular concern to object-oriented inquiries.

3.1. Data Acquisition

A range of different methods for generating data within OOI can be understood as either *schematic*, *narrative*, or *manipulative* inquiries. By mixing and combining them, we can continuously change the lens and encounter a game within different states and contexts. We collected several methods and approaches that have already been used in HCI or associated literature and have the potential to enable researchers to acquire a variety of perspectives on and from a game.

3.1.1. Schematic Inquiries

We refer to methods aimed at gaining insight into objects as a *crowd* Bryant (2011) or assemblage of other objects as *schematic* inquiries. Within these, researchers focus on the things that come together to make up another thing, the part that forms a whole. Methodological suggestions for schematic inquiries often stem from an ANT background, and range from *listing* parts, creating *photographs* with things as the focus, assembling *exploded* or *cut/away views* or simply drawing *flowcharts*. Additionally, we see examples of schematic inquiries in the tradition of *system log analysis*, though with the intent to understand a given system, rather than the errors generated when someone interacts with it (McVeigh-Schultz et al., 2012), *workbooks* (Gaver, 2011), *annotated portfolios* (Hauser et al., 2018), or the *visualization of actor-networks* (Spiel et al., 2017). We schematically acquire game-related data in our methodological exploration on *being* a game.

3.1.2. Narrative Inquiries

Several approaches allow researchers to generate narratives from an object's perspective. For example, technology can be *anthropomorphisized* to discuss the different roles it takes up in relation to humans and other objects (Buttrick et al., 2014). Narrative inquiries can also rely on multiple human perspectives, be it through *co-speculation* on a thing with distinct groups (Wakkary et al., 2018), *interviews* with actors who enact being a thing from previously collected data (Chang et al., 2017) or entire *speculative enactments* (Elsden et al., 2017) from an object's perspective. We partly acquire data through a narrative inquiry in our methodological exploration on *acting as* a game.

3.1.3. Manipulative Inquiries

The active manipulation of material and objects to *do* philosophy and inquire through an object's perspective is another form of practizing *Object-Oriented Inquiry*. Especially relevant in inquiring into less tangible actors and concepts as objects, *speculative design* can be a form of *doing* philosophy through the creation of speculative virtual technologies (DiSalvo et al., 2016). It follows the tradition of *carpentry*, which has been actualized in playful (Gualeni, 2014; Bogost, 2016) and techno-physical forms (Wakkary et al., 2015, 2017). Explicit *deconstruction* can additionally be a form of engaging with a thing through actively disassembling and re-configuring it into the same or different things (Murer, 2018). In our methodological exploration on *becoming* a thing, we acquire data by manipulating game related objects.

3.2. Data Analysis

The above-mentioned approaches toward data acquisition lead to a range of artifacts, structured and unstructured texts as well as images that serve as potential data points. These different forms of data lend themselves to different modes of analysis; we identified three strands that can be applied, either individually or in combination. We have identified *descriptive*, *discursive*, and *speculative* analyses. A chosen type of analysis results, in turn, in a range of epistemological implications, which we briefly touch upon for each approach.

3.2.1. Descriptive Analysis

A straightforward form of inquiring into a technological object (including digital games) is to implement the recording of log data as an interface for human analysis. Indeed, quantitative analyses allow us to gather insights into the range of complexity surrounding a thing and/or its communication, and to gain initial pointers for potentially relevant areas for qualitative introspection. This notwithstanding, we would argue that the data can also be analyzed phenomenologically and qualitatively (Ädel, 2014). Descriptive results allow for a reductive overview on the complexity of objects, especially across temporal instances, and allow us to illustrate scale with regard to the complexities surrounding the reliance of objects on other objects as parts or required environments. Hence, schematic inquiries lend themselves particularly well to descriptive analyses. We include some descriptive observations with quantitative and qualitative aspects in our methodological exploration on *being* a game.

3.2.2. Discursive Analysis

Seeing all data sources, including non-textual ones such as images, as an instantiation of meta-text allows researchers to then apply textual methods such as thematic analysis (Braun and Clarke, 2006), grounded theory (Strauss and Corbin, 1990), or discourse analysis (Fairclough, 2013). In a classical ANT-inspired approach, the data could also be used to define actors and their actions by following them—including their relations or "associations" (Latour, 1984)—through their manifestations within given networks (Latour, 2005). This analysis invites researchers to practice ANT, which leads to the necessity of translations back into text, of which the resulting "trahison" (Law, 2006) requires active reflection. Basically, we refer here to any type of analysis that aims to contextualize different texts and construct knowledge through this process. As different manifestations of data are all translated and approached as texts, this form of analysis s allows researchers to engage practically with the notion of flat ontology between core texts, images, objects, and other traces that things leave. We analyse our data in the methodological exploration on *becoming* as well as *being* a game discursively.

3.2.3. Speculative Analysis

Speculative analyses have has been proposed as a necessary practice for HCI and ubiquitous computing research (Bardzell and Bardzell, 2014). We see the potential for explicitly speculative analysis in its application to narratives such as design fictions (Tanenbaum, 2014), fictitious designs (Tanenbaum et al., 2010), or simply attending to the object as represented through what is available and exploring possible avenues for different manifestations (Giaccardi and Karana, 2015). The knowledge gathered from this practice is particularly relevant for design purposes. We employ speculative analysis in our methodological exploration on *acting as* a game.

3.3. Data (Re-)presentation

When aiming to represent data, some form of visualization is often already inherent in the process of acquisition or analysis. Visualization can also constitute a core part of engaging with the complexity of things on their own, as shown in the "Anatomy of an AI" map (Crawford and Joler, 2018) as well as some of the case studies below. These visualizations can illustrate the complexity of assembled things to such an extent that in scientific writing and presentation, researchers may be required to only present selective views. We have encountered this issue when preparing this publication and point interested readers to the **Supplemental Material** which provides the full visual and textual context for our methodological explorations. Many of the methods described above readily lend themselves to a visual representation of data. As static media, these allow us to reflect on the temporal fleetingness they represent as they can only ever be snapshots. Hence, any systematic engagement with an object in general and a game in specific remains necessarily incomplete and partial on this account as well. To put this and the approaches above into practice, we explored their methodological potential in three sets of a total of six cases, each of which illustrates different kinds of object-oriented knowledge we could acquire.

4. METHODOLOGICAL EXPLORATIONS

We conducted three different methodological explorations on *becoming*, *being*, and *acting as* a game. Across several stages in the design process we use them to illustrate the feasibility of *Object-Oriented Inquiry* for HCI research in the context of games and play and how different modes of data acquisition and analysis lend themselves to different insights guiding evaluation and/or design decisions.

4.1. Approach

In choosing the methodological explorations, we aimed to cover different manifestations of different games and the objects surrounding them. We specifically set out to explore different methodological notions embedded in the concept of *Object-Oriented Inquiry* in practice. In the first methodological exploration, we focus on physical aspects of a game which are either intended to integrate with a digital element (as is the case with the Nintendo Labo™) or rely heavily on other technology for their construction (which is the case with the 3D-printed figurine). This endeavor was driven by an interest in identifying an appropriate context for manipulative inquiries and include predominantly physical objects that are augmented in play. For the second methodological exploration, we chose to investigate the assemblage of a browser game without requisite physical manifestations to consider a purely digital schematic context. In the last methodological exploration we engage with directly tangible technological games bridging the two previous forms of play and allowing inquiries into contexts that are embedded in interaction.

Each methodological exploration relies on a combination of the previously described methods for data acquisition and analysis, adapted to the particular context in which they are applied. Across them they illustrate different choices for inquiring into a game and the different types of knowledges that might come from doing so. As the material in the first two methodological explorations is much more extensive than can be described in the body of the paper, we point interested readers to the **Supplemental Material** for more detailed insights.

4.2. Becoming a Game

Making or crafting as an activity people do has been a predominant angle of prior HCI research for example, (for example, Blikstein and Krannich, 2013; Tanenbaum et al., 2013; Toombs et al., 2015; Meissner et al., 2017; Frankjær and Dalsgaard, 2018). There, the focus lies on the people who are seen as the primary initiators and "makers" of artifacts. However, by turning the magnifying glass to the *becoming* of a game, we can investigate another perspective on the process (similar to Huvila, 2016, but using a method with humans removed). It allows us to focus on the process in a different way, potentially uncovering new object-centered perspectives into playful crafting and production material. The leading question here is then: *What is it like to become a game?*

We engaged with two different materials and modes of assembly to take a closer look at the processes entailed in *becoming* a game. In the first case, we aided cardboard material in the construction of a new house as part of the Nintendo Labo™ Variety Kit and reconstructed a previously de-assembled piano from the same set. In the second case, we observed machines supporting the becoming of a small three-dimensional figurine. During both of these processes, we took a vast amount of photographs for documentation, namely 150 in the case of the house, 186 with the piano and 341 for the figurine. We then reviewed the photos and created booklets akin to workbooks (Gaver, 2011) illustrating the becoming process (see **Supplemental Material** and **Figure 1**). Hence, we performed a *manipulative* inquiry and analyzed it *discursively*, i.e., data were acquired through conducting and documenting the alterations on the object and the resulting documentation served as the basis for our analysis.

The three things resulting from the becoming process all went on a different journey to arrive at the state that we identified as a preliminary constant. Even though the procedure of arriving at the insights was identical, in assembling the booklets, we could observe different aspects of the process emerging as relevant to each game context.

The house took form from sheets of cardboard with pre-cut parts for assembly (see also, **Figure 1**, left). Hence, before the house became one connected thing, individual objects had to come together. Some of these objects resisted the process, e.g., by clinging to the sheet and only letting go after injury (slight breakage of part of the material). The house itself holds a three-dimensional structure, but the sheets are two-dimensional, which means that external forces had to be exerted on the individual objects to give them the shape that allowed them to interlock with others. The different parts coming out of the cardboard have different relationships to the house. While many came together to build the house, others were left out and did not take part in the process of becoming a house. Instead, they became merged into an assemblage of discarded objects collected as garbage.

As the piano was in the process of becoming *again* instead of merging its parts for the first time, there were no left out pieces, but rather missing ones that left it in a state of permanent incompleteness. The instructions for the piano assumed that it would be built from scratch. These expectations were not met in the particular process of becoming again. On the other hand, some parts had already taken on three-dimensional forms before and presented themselves as such (see also, **Figure 1**, middle). Traces of previous interactions, bends and folds revealed a prior history of the piano, which is independent of the person involved in reassembling. Still, the parts also partly resisted in becoming again—at least in comparison to the expected state given in the instructions. The preliminarily final version of the piano is somewhat crooked as the material consolidates previously known positions with current positions within the piano.

The figurine went through an entirely different process of becoming. It first existed as a digital object, which was virtually malleable. However, the figurine and what it stands for have an entire history of becoming that we were not privy to. This is another point of resistance that illustrates how we can only gain partial insights into the process of becoming due to the limitations of our own embodiment and placement. In the temporal slice we participated in, the figurine engaged the help

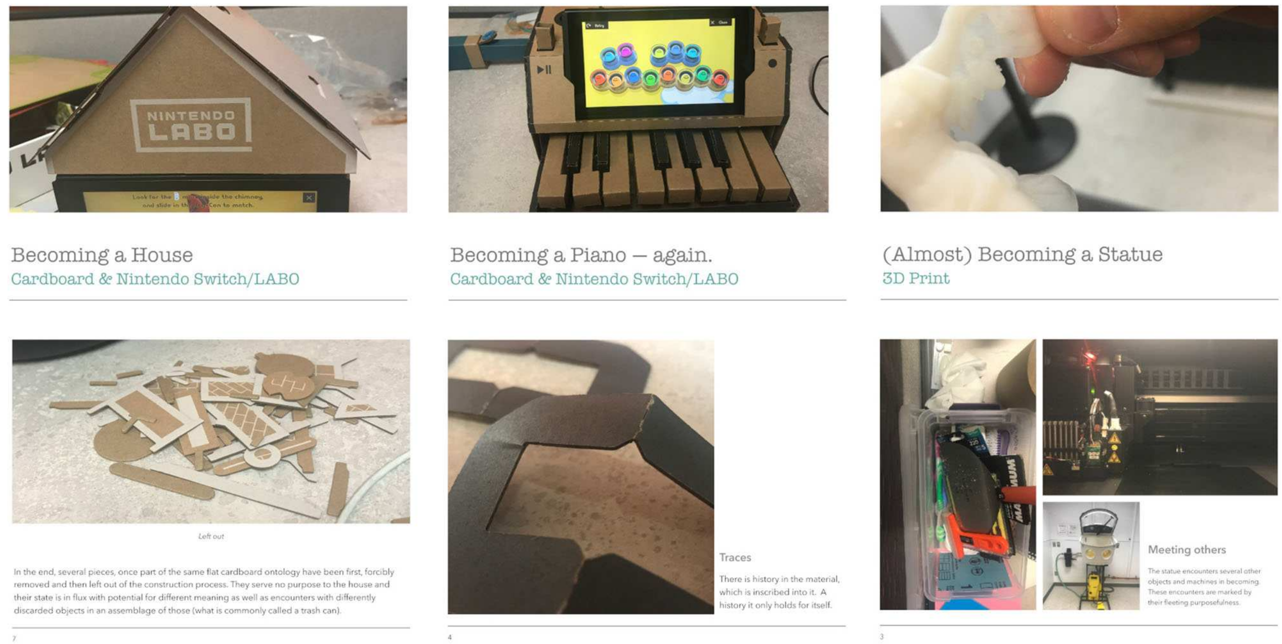


FIGURE 1 | Examples of booklets illustrating the becoming of the cardboard house and piano as well as the figurine.

of several other objects and machines to support the becoming process (see also, **Figure 1**, right). These objects have a primary usefulness in aiding the process, but also resist it in parts as can be seen by the destructive power that the cleaning station exerted on the figurine, breaking part of a wing, which leaves this particular assemblage of material in a state of externally (humanly) assigned incompleteness. We could also observe that the object took on specific meanings for the people involved in its becoming, precisely because it was the focus of our observation. It became a token of its brokenness, instead of being discarded or replaced. Hence, objects are shaped by researchers' observations in a similar way as they are shaped by the actions of humans within such contexts (Obrenović, 2014).

This allows us to consider implications relevant to the potential evaluation and (re-)design of the involved objects, but also, more generally, to technological and material development. The house shows us how, through care for discarded pieces, we could envision alternative futures for these pieces where they have a place outside of garbage and can be sustainably integrated in this or other projects. The piano illustrates design assumptions of an ideal states instead of re-use and appropriation. Instructions and availability of material should be part of design considerations that account for these practices (e.g., Jackson and Kang, 2014). The figurine exemplifies how the design of technology for digital fabrication must not only consider design for use by humans, but also for object-technology interaction to aid the becoming of games appropriately.

Through these three investigations into *becoming* a game, we could observe several aspects of the process being of different relevance to the particular game objects at different times. The method of taking photos during the becoming, and then assembling them into booklets, appears to lend itself to the analysis of different material contexts and processes surrounding the *becoming* of games. We also note that the resulting documentations as workbooks created yet another set of objects that could be inquired into and analyzed in their own right.

4.3. Being a Game

By understanding any game as an assemblage of other objects, the complexity of trying to account for all these entities can increase at an exponential rate. Additionally, considering the different states games might be in at different points in time further increases this complexity. Our leading question through this methodological exploration was thus: *What is it like to be a game of many things?*

We *schematically* inquired into *Kittens Game* through *lists*, which we then analyzed *descriptively*. Concretely, we collected a range of objects contributing to the being of *Kittens Game* in a specific instance at a specific point in time. As a visualization mode that preserves the structural entanglements between the different objects, we used a mind map, parts of which we show in **Figure 2**. On one side (blue), we depict the objects as they are represented at a certain point in time during the game, on the other side (purple), we captured all of the objects as they appear within the code, going down to one level past classes, but covering object instances in arrays. Mirroring these two perspectives allows us to directly contrast between an interface perspective and an assemblage perspective as it pertains to the game. We understand the processes of object collection and visualization as part of acquiring our data, whereas the description constitutes our analysis.

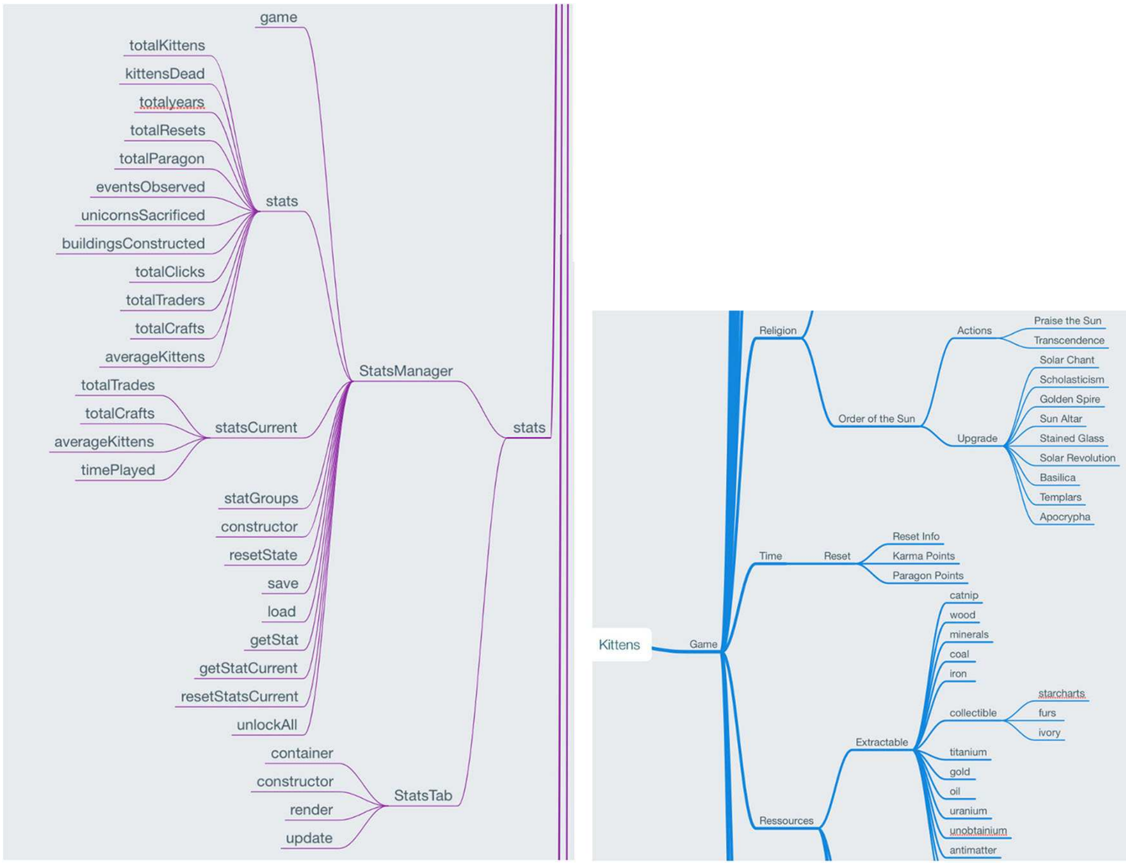


FIGURE 2 | Partial view on the objects who are part of *Kittens Game*. The left side depicts objects as represented in the code structure whereas the right side lists objects as presented to a potential player at a certain time.

Kittens Game[1] was developed in 2014 and belongs to the genre of idle games. At the beginning of *Kittens Game*, human players are represented as a single kitten in a catnip forest. Through gathering and refining catnip, more and more proverbial kittens gather together and advance their civilization beyond even current human technological progress. The system reveals itself gradually, becoming more and more complex over time (Alharthi et al., 2018b).

We extracted 282 objects as they were available to the first authors after five months of interrupted play (including four resets). We also collected 2,034 objects within the code. Individual instantiation and underlying implementation are only two ways in which we could think of the things which contribute to the existence of *Kittens Game*. We ignored several other physical and conceptual objects that might be relevant here, such as texts from players and developers, the genre context, influences from other games, metaphorical references seeping into and out of the game, the range of platforms and technologies the game could be played on or the different instances for each context of play—all of which co-constitute of what *Kittens Game* is. These could provide further alternative perspectives on the manifestations of the game.

On the right hand side of **Figure 2** (blue), there is a selection of objects as they present themselves to a potential human player during a specific state of the game at a specific point in time. Objects can refer to metaphysical as well as physical representations. Resources, buildings, concepts and game mechanics are all considered to be objects in this context. Only in acting with and on each other do they make a *being* of *Kittens Game* possible.

The implementation of the game is additionally tied to an object-oriented perspective through the use of JavaScript as the programming language. The left hand side of **Figure 2** illustrates the objects responsible for collecting, defining and manipulating the statistics of *Kittens Game*. The file is separated into calculations (*StatsManager*) as well as a class for representation (*StatsTab*). The objects themselves range from references to the game instance (*game*) to containers for statistics (e.g., *kittensDead*) and functions which are both specific (e.g., *getStatCurrent*) and general (e.g., *save*). Hence, internally, all virtual objects are declared and instantiated as a flat ontology. Regardless of their later behavior (e.g., variable, container,

[1] Available online at: http://bloodrizer.ru/games/kittens/. In reading the entire graph in the **Supplemental Material**, readers might be confronted with spoilers.

function) they are equally objects. Only by looking at the concrete mechanics can we distinguish their purposes.

Creating such schematic lists and analysing them descriptively enables designers to understand the complexity of the games they aim to create—not just as a complexity of the code base but with the added complexity of the semantic objects presented to players. By aiming to capture the assemblage of parts, designers might find this a useful tool for understanding potential additions and missing objects that can meaningfully alter a given status quo. In that regard, this approach relates somewhat to existing practices in software engineering (Bruegge and Dutoit, 2009), e.g., the use of the Unified Modeling Language (UML) (Medvidovic et al., 2002). However, while the UML is used to specify, structure and document software architectures, whereas our approach aims to understand more ontologically of "what there is" and not necessarily conceptually tied to the code base or its representation in the game. It operates from the concept of ontological lists (Bogost, 2012) decidedly without illustrating relationships or complexities. In differentiating between semantic and structural objects the aim is more to identify differences and commonalities from different perspectives. If a given software is created while making use of UML, this can very well be the starting point that can be reduced ore expanded upon to be suitable for a list based investigation.

Our schematic approach sheds light on the complexity of games, even though this can never be completely captured. Nevertheless, the differences in how *Kittens Game* manifests through its interface and through its implementation provide a basis for further investigation. For example, researchers could consider including some of the other objects we identified as co-constituting a game or evaluate mismatches between objects instantiated in code and toward players, or gain a deeper understanding into how the mental models of players are guided and might, hence, differ from the mental models of developers.

4.4. Acting as a Game

While games might act without humans around them, within HCI we are mostly concerned with how technologies (and, subsequently, games) manifest themselves through interaction with human or other animate actors animals (e.g., animals in Mancini, 2011). However, in these interactions, we focus on animate agency with technologies and games, neglecting other potentially relevant actors and perspectives that contribute to the enactments. Actively erasing these animate actors from our analysis allows us to reflect on the infrastructures (De Angeli et al., 2014) and additional requirements that are relevant to the design of virtual and physical playful artifacts and technologies. The leading question in this methodological exploration is, hence: *What is it like to act as a game?*

In contrast to ontographs, a form of schematic inquiry which focuses the photographic lens solely on objects (Bogost, 2012), we deliberately chose images that originally included humans interacting with technologies. We then redrew the photographs, focusing on the things that contribute to the technological dispositive with which people interact, but decisively cut out human actors (see **Figure 3**). We focus on two different contexts: (1) during the design of musical pads that allow several people to playfully create music through spatial movement (left hand side) and (2) a player engaging with a finished commercial product which projects a virtual environment on a head-mounted screen (right hand side), effectively allowing insights into a more evaluation driven context. In visually creating alternative perspectives on the interaction, we understand this procedure as a form of *narrative* inquiry that also follows certain actors and traces them visually. Through our editorial intrusion into the picture, we change the potentials of narratives it presents

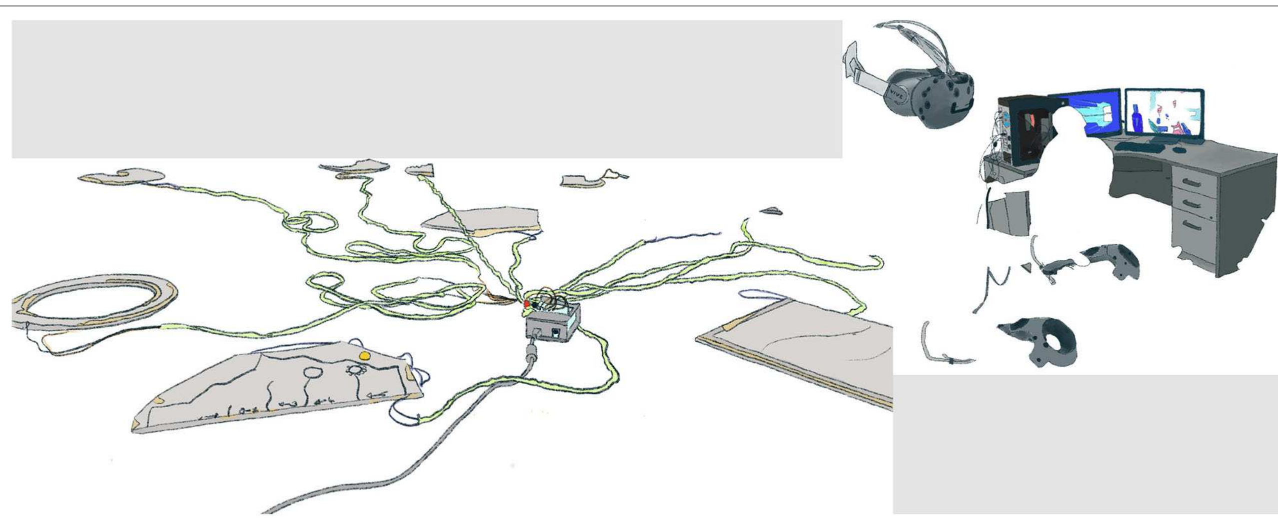


FIGURE 3 | Different illustrations of acting as a game in different contexts with human and non-human actors. Both illustrations have all human actors removed. To the left, a prototype of music pads refocuses our view on the mess of cables, whereas to the right, the context of a VR scenario indicates the relevance of not just the technological artifacts but the necessity of furniture to enable play. Note that non-essential actors as well as human actors create artificial white space.

to us and prepare the data basis on which we *speculate* on these different potentials from the objects' perspectives.

Both cases, individually, point to different aspects relevant to *acting as* a game. However, tracing the relevant objects from a given perspective is limited to exactly that perspective. As such, it might ignore other things which were not in the focus of a photographer who might not necessarily have been attuned to an object-oriented perspective and might have missed aspects that are relevant to the games' perspective.

With the musical pads, we deliberately worked off of such an unattuned image to understand how an *Object-Oriented Inquiry* can support reflection on ongoing design processes. One consequence of this is that there is a cable leading outside the picture, leaving the other objects it might have been attached to (plugs, computers) outside of our analysis. However, this perspective taking is precisely what allows us to reflect on the focus we take when attempting to capture technological objects during interaction. Additionally, some parts were occluded by humans actively engaging with the technology. As we only traced the relevant non-human actors, this creates artificial white spaces that actively remind us of the limited perspective we have available when inquiring into a technology through a snapshot in time.

In particular, this image shows us the messy state (cf. Dourish and Bell, 2011) in which the thing finds itself at this current moment in design. Cables are everywhere, obstructing the freedom of the plates to move into different spaces. They try to distance themselves from a centralized entity, but never manage to get rid of it entirely. A potential design decision following from this is that a wireless version of this design idea might be preferable. While designers might reach this conclusion in other forms as well, this is one way to reach it from the object's perspective.

In the case of the VR play scenario, the white space illustrates the need for another human to position a cable in just the right way. This leads to humans being effectively objectified as assistants to the technology. Even when that human actor is systematically excluded from the representation, they are relevant to the manifestation of the game as an object in play. We further notice that the game is instantiated not just by the apparent technological bits and pieces, but also by more circumstantial objects such as the furniture on which parts of the technology are placed. These are objects that have not been actively designed for, but are instead a matter of happenstance. They are assembled according to availability or convenience as perceived by the people who focus on interacting with what they view as the core technology.

While human actors in this space are visually (and potentially also auditorily) re-placed into a virtual environment outside of the space, the physical aspects of the technology are strongly tied to their environment and have to collaborate with things that might not be ideal to their instantiation. It resists and subjects human actors to do its bidding in cooperation. Otherwise it refuses to collaborate with another human actor. Considering this refusal, designers could target this as an identified weak spot and resolve it to a point where the technology does not require as much intricate attention from humans. Again, these issues can also be reported from humans or identified through other methods, but this is another part in designers' toolsets to do so by engaging speculatively—we dare say, artistically— and productively with the objects in the interaction.

The radical exclusion of human actors and the explicit inclusion of potentially relevant additional objects provides a different view on how a game manifests itself through interaction with players. By removing humans from the picture, we are invited in "speculating about how that object encounters the world" (Hayles, 2014). The illustrations offer active and reflective engagement as the process of redrawing encourages researchers to explicitly focus their attention on inanimate actors. This approach relies on capturing the limited perspective of an in-the-moment snapshot of a thing's manifestation in action. Potentially, a series of drawings along different moments in time or covering a broader range of perspectives could provide further insights, while still only marginally mitigating this limitation.

Across the three methodological explorations, we conducted six case studies probing into a range of different modes of object-oriented knowledge productions and their implications for analysis. We favored the illustration of breadth (in the form of several methods) instead of depth while hinting at further opportunities to dig deeper in specific contexts. Subsequently, we now discuss the epistemological and methodological implications of our explorations.

5. DISCUSSION

Considering our framing of *Object-Oriented Inquiry* and its actualization in our methodological explorations, we now connect our insights to more general epistemological and methodological deliberations. We then shed some light on the usefulness of *Object-Oriented Inquiry* as a productive agenda for game design and research.

5.1. Epistemological Insights

Our exploration on *becoming* a game illustrated the perspective of different types of physical manifestations around two sets of play contexts regarding issues like re-use, repair, object context, and instructional materials. While we started with object-oriented *ontology*, our work was fundamentally oriented toward knowledge production and how we might use the ontological backing to gain insights on games. It was not our aim to establish *what* a given game is, but rather explored *how* we can know what it might be like to be a game and *how* we might know about it differently using speculative object-oriented approaches. It was useful to compare and reflect on how we perceive, define and understand different reconfigurations of the material elements that might be associated with a game. We could know about these through different means but as a decidedly playful and creative approach, we deem this procedure particularly conducive to game design contexts.

Our exploration on *being* a game showed how analysing which concrete parts constitute it can inspire additional features or point out missing ones. It supports thinking about different (re-)presentations of a game and associated scale biases. These types of engagements invite explorations of the complexity associated with games, to think about details within the context of a larger picture. In that regard, we expect this approach to be potentially useful in both, design and research settings.

Our exploration on *acting as* a game provided indications for more holistic game design and offered critique on existing prototypes. It further draws attention to the roles objects of play inhabit when radically reduced to themselves. As a mode of knowing about games, this approach provides design opportunities as a part of iterative game development or evaluation.

Hence, each of the three methodological explorations lead to distinct insights and let us know different aspects of what it might be like to be a game without assuming individual or collective completeness. Additionally, the explorations can be understood as referring to different types of relations in Bryant's flat ontology. Analysing the *becoming* of a game, means taking a look at what Bryant (2011) calls the *endo-relations* of an object as it manifests rapidly through several instances. We closely examine all the things that come together to create another thing—a game, to be specific—be it through a temporary or permanent relationship. Through re-focusing our attention on the *being* of an object, we can switch between *endo- and exo-relations* (in Bryant's terms), where the inwardly and outwardly formed relationships of a thing gain relevance. Finally, in *acting as* a game, we concentrate on its situatedness in the moment of an active *exo-relation* with another human. In all of these methodological explorations, though, it becomes apparent how "all objects are a crowd" (Bryant, 2011, p. 217), an assemblage of other objects manifesting in a temporally and spatially flexible form.

These object-oriented approaches decidedly limit the perspective taken by Human-Computer(game) Interaction and can, hence, not inform us on many matters relevant to human sociality. They are somewhat static snapshots of an objects' perspective on interaction. While not lending themselves easily to an understanding of process of interaction, they do, however, illustrate how taking an object's perspective means following a plan whereas interaction is often signified through situated actions (Suchman, 1987). While objects could feasibly attributed those actions as well (as we have shown in the case on *becoming* a game, our approaches do not (yet) do so. Another relevant methodological limitation lies in how all approaches remove players' perspectives from the analysis—albeit deliberately. However, they do not support questions concerned with players' experiences or are conducive to tackling equity issues (e.g., privileged immersion Passmore et al., 2018) appropriately. As such, it is a somewhat apolitical perspective to take, one that does not lend itself well to transformative research. As every method or set of methods limits how we can know about a specific context, we deem it relevant to point out the limits of the knowledge produced by using the approaches we delineated in our explorations. Given the political and transformative potential speculative design itself has brought forward (cf. de Oliveira, 2016), we see potential in the development of object-oriented methods that include such considerations.

5.2. Methodological Insights

These different perspectives on a range of game contexts were subjugated to different methods —albeit all of them sharing a *speculative* core. In our case study on *becoming* a game, we performed a manipulative inquiry into physical objects and digital fabrication, which we analyzed discursively. In our case study on *being* a game, we descriptively analyzed a schematic inquiry into an idle game. Finally, in our case study on *acting as* a game, we speculated on a visual narrative inquiry. These states and inquiries are not necessarily tightly coupled, though. One could imagine a schematic inquiry into becoming as much as a manipulative inquiry into acting, a speculative analysis of being and a descriptive analysis of becoming (and many other combinations). A mix of potential inquiries and analyses on the same thing yield different perspectives on it, which potentially become disruptive and disjoint between them, opening up the option of creative action for resolving these multiple meanings coming from the same thing. We chose our cases along the options they illustrate.

Part of our contribution also lies in identifying the strands of existing speculative object-oriented approaches as *schematic*, *narrative* and *manipulative* inquiries for data acquisition as well as *descriptive*, *discursive*, and (purely) *speculative* analysis in section 3. By categorizing them as such and situating our explorations within them, we aimed to show how game design and research could adapt these to different contexts. In that, we invite further adaptations and explorations that might illustrate more breadth in these approaches as well as how they might be combined with more classical methods to contribute to a range of insights from different perspectives. For example, we envision our approaches to be used in practice alongside more established methods such as contextual inquiry through interviews and observations (Holtzblatt et al., 2005) or other approaches oriented on gathering data for interaction design from humans (Preece et al., 2015).

A core challenge in Object-Oriented Inquiry remains in being *humble* (Hayles, 2014) about the insights we gain from these endeavors. The knowledge and perspectives we have access to remain necessarily partial. As human researchers we engage with games and inquire into them through our distinct perceptive apparatus, resulting in fundamentally limited access to appropriately claim an understanding of a game *as a game*. While this is a core methodological limitation (with adjacent epistemological implications), there is also no way around it. In addition, there was an inherent focus on visually charged representations and inquiries. All methods assume some kind of textual or visual engagement, ignoring the knowledge we could gather through analysing smell, haptics, taste, and other sensations potentially acting on us through a technology. While we encountered these modalities in our research, we somewhat discarded them incidentally during our analysis, due in part to being lured by the temptation of textual and visual representation modes as relevant to communicating this research through academic papers.

What these methodological explorations offer, however, are insights into different manifestations of things through distinct perspectives. They contribute to an understanding of the complexity of the assemblage of games and, together with other methods of inquiries into humans, interaction and conceptual relevance, provide us with a toolset that augments the perspectives, questions and analyses of classical research, both in lab settings and in the wild.

5.3. Revisited: Object-Oriented Inquiry for Games and Play

Object-Oriented Inquiry has a place in game design and research akin to speculative methods within HCI (Bardzell and Bardzell, 2014). However, while traditional speculation is oriented toward alternative potential futures, *Object-Oriented Inquiry* speculates about the present, and the role of currently existing technologies within it (Hayles, 2014). In that, it can be a structured approach for creating design heuristics, especially when prototypes are not refined enough yet for playtesting.

Further, through attending to the games as technological objects, designers and researchers can use *Object-Oriented Inquiry* to critically engage with the limitations of any perspective they encounter in their respective and shared practices. With objects, it becomes painfully obvious that a complete picture about their 'experience' is never achievable (Bogost, 2012). *Object-Oriented Inquiry* can function as an exercise to reflect on the boundaries of empathy (Spiel et al., 2017) toward other actors (animate or inanimate) but particularly to our games.

5.3.1. Design

Each of the methodological explorations provided us with some indications on how to step forward in game design. The methodological exploration on *becoming* a game illustrated issues around dealing with *remaining* or leftover materials, issues around instructions for *re-building* and the associated messiness of having some parts in different states than others as well as issues around *destruction* and the becoming of an incomplete object. Engaging with these issues inspire investigations into how processes can be altered to avoid material and time waste while keeping physical components interesting and relevant to players.

The methodological exploration on *being* a game provided insights into their assemblage and how different structured ways of conceptualizing these can guide a deeper understanding of potential mismatches, new solutions and alternative re-presentations. Particularly for highly detailed and complex games, this can directly lead to improvements in code that make further development easier through mindful refactoring. It can be seen as a potentially playful adaptation of already existing object-oriented software engineering practices (Bruegge and Dutoit, 2009). However, in contrast to those, our approach deliberately leaves out aspects of the system architecture (particularly relations) as to leave room for speculation and imagination and has the potential to include objects not related to a game's software implementation. It takes a structured activity to allow for creative freedoms (Makhaeva et al., 2016) with familiar tools, but serves an entirely different function within the design process.

Finally, the methodological exploration on *acting as* a game makes way for deliberations about incidental objects that are a necessary part of a playful technology setup, but not deliberately designed for. It further leads to an understanding of potential avenues for redesign by speculating about the emotional state of the game but also identifying opportunities for meaningful change.

Hence, *Object-Oriented Inquiry* lends itself to a range of different insights that can be beneficial as part of a well-rounded design practice. We do not claim that this is not already happening and showed that, indeed, it is, as in Murer (2018), but we offer a vocabulary and useful theoretical context to articulate these kinds of knowledges by presenting a speculative thought experiment of how we might approach games and play from an object oriented perspective not just pragmatically (like UML does), but also ontologically. In that regard, future work could conduct empirical studies investigating whether there is actually an epistemological difference between the two approaches.

5.3.2. Evaluation

We understand evaluation as a form of inquiring into game use with the intent to understand particularities about the interaction and to inform future re-design and improvements. It can be a part of iterative game design as well as research into games and play. In particular, through speculating from a game's perspective, we can gain additional insights compared to relying on eloquent and available humans to convey their perspective. For example, people might tell researchers how they enjoyed interacting with a given design, whereas logs indicate that this was rarely the case. By putting these logs into a first-person narrative statement (e.g., "I was barely used." Spiel et al., 2017), we can uncover frictions that not only tell us about the current stage of a design, but also give way to further developments (cf. Sengers and Gaver, 2006).

Investigating the *becoming* of a game might only be relevant to contexts in which others are expected to drive the becoming of an object (as is the case with the Nintendo Labo™, but not necessarily with 3D printing devices). Looking at how the material rebels against manipulation provides an additional perspective to inquiring into how long people took to assemble a certain object or which steps they followed, and in which way. It can qualitatively aid us in understanding *why* we observe certain behaviors and patterns that might be part of playful interactions.

Analysing the *being* of a game allows for an in-depth analysis of heuristics for evaluation and can inform other studies by generating specific questions about a game and subsequently trying to address them. Hence, while not directly lending itself to definitive results, *Object-Oriented Inquiry* can be used as a starting point for evaluation.

More directly, inquiring into the *acting as* a game can be a way to further consider the particularities of an interaction by decidedly focusing on the game in that interaction. By taking a step back from privileging players' perspectives, we might just reach into a space that could, in return, become relevant to all actors, animate, or inanimate.

This is not to say that game designers and researchers could not arrive at such insights in a different manner as well. However, there is something inherently *playful* in a speculative engagement with games as objects. Such an approach might lend support to those who prefer to inquire into their environment with a more playful mindset. Hence, these explorations and methodological suggestions are not meant to replace existing ones but rather expand the toolset game designers and researchers can use to understand (their) games.

6. CONCLUSION

We have provided an overview of existing object-oriented practices in HCI research and applied them to a set of methodological explorations in the context of games and play to structurally inquire into the kinds of knowledges that is embodied and materialized within games. Focusing on *becoming*, *being* and *acting as* a game, we took a look on three different manifestations of games. We showed that *Object-Oriented Inquiry* can provide an opportunity for game design and research activities by allowing us to gather holistic insights into different perspectives pertaining play, players, and playful engagements between them.

Future work in this area could investigate and critique our analysis through additional methodological explorations and the investigation of the usefulness of these perspectives as part of larger studies. Additionally, it could be fruitful to find methods that address other human modalities through which we experience objects and subsequently inquire into them, such as smell, sound, and haptics. Another line of research could look into making *Object-Oriented Inquiry* applicable to animate actors.

Our work illustrated the feasibility of *Object-Oriented Inquiry* for game design and research from the perspective of HCI games researchers. It provides a range of indications on the kinds of knowledges games and, potentially, other technologies embody, and presents several methodological explorations as examples for OOI oriented practices. We encourage game designers and researchers to take on an object-oriented perspective to gain deeper insights into the intricacies of all parts pertaining to the interaction between games and players.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Office of Research Ethics—University of Waterloo. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

The work described in the is manuscript has been conducted by KS under mentorship and guidance from LN. KS then led the process of producing this manuscript with iterative feedback and contributions by LN.

FUNDING

This work was supported by NSERC Discovery Grant 2018-06576 (LN), the Canada Foundation for Innovation Infrastructure Fund 35819 SURGE–The Stratford User Research and Gameful Experiences Lab as well as by a Marietta Blau-Grant of the Austrian Academic Exchange Service (Ö-AD) for KS.

ACKNOWLEDGMENTS

The authors thank all the games that enriched their lives and this research. Further, we are in deep gratitude to Emeline Brulé, Judith Good, Fares Kayali, and Christopher Frauenberger for providing supportive feedback.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomp.2020.00018/full#supplementary-material

REFERENCES

Ädel, A. (2014). Selecting quantitative data for qualitative analysis: a case study connecting a lexicogrammatical pattern to rhetorical moves. *J. Engl. Acad. Purposes* 16, 68–80. doi: 10.1016/j.jeap.2014.09.001

Alharthi, S. A., Alsaedi, O., Toups, Z. O., Tanenbaum, T. J., and Hammer, J. (2018a). "Playing to wait: a taxonomy of idle games," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18* (New York, NY: ACM), 621:1–621:15. doi: 10.1145/3173574.3174195

Alharthi, S. A., Toups, Z. O., Alsaedi, O., Tanenbaum, T. J., and Hammer, J. (2018b). *The Pleasure of Playing Less-A Study of Incremental Games Through the Lens of Kittens*. Pittsburgh, PA: Carnegie Mellon University; ETC Press.

Auger, J. (2013). Speculative design: crafting the speculation. *Digital Creativ*. 24, 11–35. doi: 10.1080/14626268.2013.767276

Bardzell, J., and Bardzell, S. (2014). "a great and troubling beauty": cognitive speculation and ubiquitous computing. *Pers. Ubiquit. Comput*. 18, 779–794. doi: 10.1007/s00779-013-0677-8

Blikstein, P., and Krannich, D. (2013). "The makers' movement and fablabs in education: experiences, technologies, and research," in *Proceedings of the 12th International Conference on Interaction Design and Children, IDC '13* (New York, NY: ACM), 613–616. doi: 10.1145/2485760.2485884

Bogost, I. (2012). *Alien Phenomenology, Or, What It's Like to be a Thing*. Minneapolis, MN: University of Minnesota Press. doi: 10.5749/minnesota/9780816678976.001.0001

Bogost, I. (2016). *Play Anything: The Pleasure of Limits, the Uses of Boredom, and the Secret of Games*. New York, NY: Basic Books.

Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. *Qual. Res. Psychol*. 3, 77–101. doi: 10.1191/1478088706qp063oa

Bruegge, B., and Dutoit, A. H. (2009). Object-oriented software engineering. using UML, patterns, and Java. *Learning* 5:7. doi: 10.5555/1795808

Bryant, L. R. (2011). *The Democracy of Objects*. London: Open Humanities Press. doi: 10.3998/ohp.9750134.0001.001

Buttrick, L., Linehan, C., Kirman, B., and O'Hara, D. (2014). "Fifty shades of CHI: the perverse and humiliating human-computer relationship," in *CHI '14 Extended Abstracts on Human Factors in Computing Systems, CHI EA '14* (New York, NY: ACM), 825–834. doi: 10.1145/2559206.2578874

Chang, W.-W., Giaccardi, E., Chen, L.-L., and Liang, R.-H. (2017). ""interview with things": a first-thing perspective to understand the scooter's everyday socio-material network in Taiwan," in *Proceedings of the 2017 Conference on*

Designing Interactive Systems, DIS '17 (New York, NY: ACM), 1001–1012. doi: 10.1145/3064663.3064717

Crawford, K., and Joler, V. (2018). Anatomy of an AI system: the Amazon echo as an anatomical map of human labor, data and planetary resources. *AI Now Institute and Share Lab.* Available online at: https://anatomyof.ai/

Cussins, C. (1996). Ontological choreography: agency through objectification in infertility clinics. *Soc. Stud. Sci*. 26, 575–610. doi: 10.1177/030631296026003004

De Angeli, A., Bordin, S., and Blanco, M. M. (2014). "Infrastructuring participatory development in information technology," in *Proceedings of the 13th Participatory Design Conference: Research Papers*? *Vol. 1, PDC '14* (New York, NY: ACM), 11–20. doi: 10.1145/2661435.2661448

de Oliveira, P. J. V. (2016). Design at the earview: decolonizing speculative design through sonic fiction. *Design Issues* 32, 43–52. doi: 10.1162/DESI_a_00381

DiSalvo, C., Jenkins, T., and Lodato, T. (2016). "Designing speculative civics," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16* (New York, NY: ACM), 4979–4990. doi: 10.1145/2858036.2858505

Dourish, P., and Bell, G. (2011). *Divining a Digital Future: Mess and Mythology in Ubiquitous Computing*. Boston, MA: MIT Press. doi: 10.7551/mitpress/9780262015554.001.0001

Elsden, C., Chatting, D., Durrant, A. C., Garbett, A., Nissen, B., Vines, J., et al. (2017). "On speculative enactments," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17* (New York, NY: ACM), 5386–5399. doi: 10.1145/3025453.3025503

Fairclough, N. (2013). *Critical Discourse Analysis: The Critical Study of Language*. Abingdon-upon-Thames: Routledge. doi: 10.4324/9781315834368

Fizek, S. (2018). Interpassivity and the joy of delegated play in idle games. *Trans. Digit. Games Res. Assoc*. 3, 137–163. doi: 10.26503/todigra.v3i3.81

Flusser, V. (1986). The photograph as post-industrial object: an essay on the ontological standing of photographs. *Leonardo* 19, 329–332. doi: 10.2307/1578381

Forlano, L. (2017). Posthumanism and design. *She Ji* 3, 16–29. doi: 10.1016/j.sheji.2017.08.001

Frankjær, R., and Dalsgaard, P. (2018). "Understanding craft-based inquiry in HCI," in *Proceedings of the 2018 Designing Interactive Systems Conference, DIS '18* (New York, NY: ACM), 473–484. doi: 10.1145/3196709.3196750

Galloway, A. R. (2006). *Gaming: Essays on Algorithmic Culture, Vol. 18*. Minneapolis, MN: University of Minnesota Press.

Gaver, W. (2011). "Making spaces: how design workbooks work," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11* (New York, NY: ACM), 1551–1560. doi: 10.1145/1978942.1979169

Giaccardi, E., Cila, N., Speed, C., and Caldwell, M. (2016). "Thing ethnography: doing design research with non-humans," in *Proceedings of the 2016 ACM Conference on Designing Interactive Systems, DIS '16* (New York, NY: ACM), 377–387. doi: 10.1145/2901790.2901905

Giaccardi, E., and Karana, E. (2015). "Foundations of materials experience: an approach for HCI," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15* (New York, NY: ACM), 2447–2456. doi: 10.1145/2702123.2702337

Gualeni, S. (2014). Augmented ontologies or how to philosophize with a digital hammer. *Philos. Technol*. 27, 177–199. doi: 10.1007/s13347-013-0123-x

Guba, E. G., Lincoln, Y. S., et al. (1994). "Competing paradigms in qualitative research," in *Handbook of Qualitative Research*, 2, 105.

Haraway, D. (1988). Situated knowledges: the science question in feminism and the privilege of partial perspective. *Feminist Stud*. 14, 575–599. doi: 10.2307/3178066

Harman, G. (2015). "Object-oriented ontology," in *The Palgrave Handbook of Posthumanism in Film and Television* (London: Palgrave Macmillan), 401–409. doi: 10.1057/9781137430328_40

Harrison, S., Sengers, P., and Tatar, D. (2011). Making epistemological trouble: third-paradigm HCI as successor science. *Interact. Comput*. 23, 385–392. doi: 10.1016/j.intcom.2011.03.005

Hauser, S., Oogjes, D., Wakkary, R., and Verbeek, P.-P. (2018). "An annotated portfolio on doing postphenomenology through research products," in *Proceedings of the 2018 Designing Interactive Systems Conference, DIS '18* (New York, NY: ACM), 459–471. doi: 10.1145/3196709.3196745

Hayles, N. K. (2014). Speculative aesthetics and object-oriented inquiry (ooi). *Speculations* 5, 158–179. doi: 10.21983/P3.0068.1.00

Holtzblatt, K., Wendell, J. B., and Wood, S. (2005). Rapid contextual design: a how-to guide to key techniques for user-centered design. *Ubiquity* 2005:3. doi: 10.1145/1066348.1066325

Hui, Y. (2016). *On the Existence of Digital Objects*. Minneapolis, MN: University of Minnesota Press. doi: 10.5749/minnesota/9780816698905.001.0001

Huvila, I. (2016). Awkwardness of becoming a boundary object: mangle and materialities of reports, documentation data, and the archaeological work. *Inform. Soc*. 32, 280–297. doi: 10.1080/01972243.2016.1177763

Jackson, S. J., and Kang, L. (2014). "Breakdown, obsolescence and reuse: HCI and the art of repair," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '14* (New York, NY: ACM), 449–458. doi: 10.1145/2556288.2557332

Latour, B. (1984). The powers of association. *Sociol. Rev*. 32, 264–280. doi: 10.1111/j.1467-954X.1984.tb00115.x

Latour, B. (1999). *Pandora's Hope: Essays on the Reality of Science Studies*. London: Harvard University Press.

Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network-Theory. Clarendon Lectures in Management Studies*. Oxford; New York, NY: Oxford University Press.

Law, J. (2006). Traduction/Trahison: Notes on ANT. Convergencia 13, 47–72.

Makhaeva, J., Frauenberger, C., and Spiel, K. (2016). "Creating creative spaces for co-designing with autistic children: the concept of A," in *Proceedings of the 14th Participatory Design Conference: Full Papers - Volume 1, PDC'16* (New York, NY: Association for Computing Machinery), 51–60. doi: 10.1145/2940299.2940306

Mancini, C. (2011). Animal-computer interaction: a manifesto. *Interactions* 18, 69–73. doi: 10.1145/1978822.1978836

Martin, J. (2005). "Close reading: functional linguistics as a tool for critical discourse analysis," in *Researching Language in Schools and Communities: Functional Linguistic Perspectives*, ed L. Unsworth (London: Cassell), 275–302.

McVeigh-Schultz, J., Stein, J., Watson, J., and Fisher, S. (2012). "Extending the lifelog to non-human subjects: ambient storytelling for human-object relationships," in *Proceedings of the 20th ACM International Conference on Multimedia, MM '12* (New York, NY: ACM), 1205–1208. doi: 10.1145/2393347.2396419

Medvidovic, N., Rosenblum, D. S., Redmiles, D. F., and Robbins, J. E. (2002). Modeling software architectures in the unified modeling language. *ACM Trans. Softw. Eng. Methodol*. 11, 2–57. doi: 10.1145/504087.504088

Meissner, J. L., Vines, J., McLaughlin, J., Nappey, T., Maksimova, J., and Wright, P. (2017). "Do-it-yourself empowerment as experienced by novice makers with disabilities," in *Proceedings of the 2017 Conference on Designing Interactive Systems, DIS '17* (New York, NY: ACM), 1053–1065. doi: 10.1145/3064663.3064674

Müller, B. C. N., Chen, S., Nijssen, S. R. R., and Kühn, S. (2018). How (not) to increase older adults'tendency to anthropomorphise in serious games. *PLoS ONE* 13:e199948. doi: 10.1371/journal.pone.0199948

Murer, M. (2018). "Making things apart: gaining material understanding," in *Proceedings of the 2018 Designing Interactive Systems Conference, DIS '18* (New York, NY: ACM), 497–509. doi: 10.1145/3196709.3196806

Nagel, T. (1974). What is it like to be a bat? *Philos. Rev*. 83, 435–450. doi: 10.2307/2183914

Obrenović, v. (2014). The hawthorne studies and their relevance to HCI research. *Interactions* 21, 46–51. doi: 10.1145/2674966

Passmore, C. J., Birk, M. V., and Mandryk, R. L. (2018). "The privilege of immersion: Racial and ethnic experiences, perceptions, and beliefs in digital gaming," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI' 18* (New York, NY: Association for Computing Machinery). doi: 10.1145/3173574.3173957

Preece, J., Rogers, Y., and Sharp, H. (2015). *Interaction Design: Beyond Human-Computer Interaction, 4th Edn*. Hoboken, NJ: Wiley.

Sengers, P., and Gaver, B. (2006). "Staying open to interpretation: engaging multiple meanings in design and evaluation," in *Proceedings of the 6th Conference on Designing Interactive Systems, DIS '06* (New York, NY: ACM), 99–108. doi: 10.1145/1142405.1142422

Spiel, K., Alharthi, S. A., Cen, A. J.-l., Hammer, J., Nacke, L. E., Toups, Z. O., et al. (2019). "It started as a joke": on the design of idle games," in *Proceedings of the Annual Symposium on Computer-Human Interaction in Play, CHI*

PLAY'19 (New York, NY: Association for Computing Machinery), 495–508. doi: 10.1145/3311350.3347180

Spiel, K., Frauenberger, C., Hornecker, E., and Fitzpatrick, G. (2017). "When empathy is not enough: assessing the experiences of autistic children with technologies," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17* (New York, NY: ACM), 2853–2864. doi: 10.1145/3025453.3025785

Strauss, A., and Corbin, J. M. (1990). *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. London: Sage Publications, Inc.

Suchman, L. A. (1987). *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge University Press.

Tanenbaum, T. J. (2014). Design fictional interactions: why HCI should care about stories. *Interactions* 21, 22–23. doi: 10.1145/2648414

Tanenbaum, T. J., Tanenbaum, K., and Antle, A. (2010). "The reading glove: designing interactions for object-based tangible storytelling," in *Proceedings of the 1st Augmented Human International Conference, AH '10* (New York, NY: ACM), 19:1–19:9. doi: 10.1145/1785455.1785474

Tanenbaum, T. J. G., Williams, A. M., Desjardins, A., and Tanenbaum, K. (2013). "Democratizing technology: Pleasure, utility and expressiveness in diy and maker practice," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13* (New York, NY: ACM), 2603–2612. doi: 10.1145/2470654.2481360

Toombs, A. L., Bardzell, S., and Bardzell, J. (2015). "The proper care and feeding of hackerspaces: care ethics and cultures of making," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15* (New York, NY: ACM), 629–638. doi: 10.1145/2702123.2702522

Wakkary, R., Odom, W., Hauser, S., Hertz, G., and Lin, H. (2015). "Material speculation: actual artifacts for critical inquiry," in *Proceedings of The Fifth Decennial Aarhus Conference on Critical Alternatives, AA '15*, (Aarhus: Aarhus University Press), 97–108. doi: 10.7146/aahcc.v1i1.21299

Wakkary, R., Oogjes, D., Hauser, S., Lin, H., Cao, C., Ma, L., et al. (2017). "Morse things: a design inquiry into the gap between things and us," in *Proceedings of the 2017 Conference on Designing Interactive Systems, DIS '17* (New York, NY: ACM), 503–514. doi: 10.1145/3064663.3064734

Wakkary, R., Oogjes, D., Lin, H. W. J., and Hauser, S. (2018). "Philosophers living with the tilting bowl," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18* (New York, NY: ACM), 94:1–94:12. doi: 10.1145/3173574.3173668

Zook, A., Magerko, B., and Riedl, M. (2011). "Formally modeling pretend object play," in *Proceedings of the 8th ACM Conference on Creativity and Cognition, C&C '11* (New York, NY: ACM), 147–156. doi: 10.1145/2069618.2069644

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

BRIEF RESEARCH REPORT
published: 15 June 2020
doi: 10.3389/fpsyg.2020.01252

Moderators of Social Facilitation Effect in Virtual Reality: Co-presence and Realism of Virtual Agents

Paweł M. Strojny[1,2]*, *Natalia Dużmańska-Misiarczyk*[1,2], *Natalia Lipp*[1,2] and *Agnieszka Strojny*[1,2]

[1] Faculty of Management and Social Communication, Institute of Applied Psychology, Jagiellonian University, Kraków, Poland,
[2] R&D Unit, Nano Games sp. z o.o., Kraków, Poland

Social facilitation has been researched for decades, but in the face of the development of virtual reality technology, new questions arise regarding the possibility of its occurrence in this environment — in the presence of computer-generated agents. Past research provided inconclusive answers: several experiments confirmed this possibility, but several others disagreed. On the other hand, previous studies have shown the important role of VR characteristics, such as realism or co-presence, in evoking other psychological phenomena. However, no study has investigated the interplay between the presence of computer-generated agents and perceived social realism in evoking social facilitation in virtual reality. To this end, the present randomized control study was conducted. The sample consisted of professional firefighters ($N = 48$), divided into an experimental group with virtual bystanders and a control group without them. Subjects were instructed to perform a rescue procedure in a virtual reality headset. The performance of participants was logged and they completed questionnaires regarding sense of presence in the virtual environment, perceived realism of the environment and perceived co-presence of virtual agents. The obtained results confirmed the role of social realism as a moderator of the occurrence of social facilitation in the presence of computer-generated agents. At the same time, the main effect of facilitation was not confirmed. These results support predictions that the subjective feeling of being in a realistic company of others may be more important in evoking social facilitation than objective facts. Furthermore, the results contribute to the debate regarding the mechanism of social facilitation, suggesting that simple augmentation of the environment with social distractors is not always enough, thus questioning the attentional explanation of the effect. Taken together, our results extend previous findings on social facilitation and open up new possibilities for designing effective virtual environments.

Keywords: social facilitation, co-presence, social presence, virtual reality, audience effect

OPEN ACCESS

Edited by:
Kathrin Gerling,
University of Lincoln, United Kingdom

Reviewed by:
Arcadio Reyes-Lecuona,
University of Malaga, Spain
Sandra Poeschl,
Technische Universität Ilmenau,
Germany

***Correspondence:**
Paweł M. Strojny
p.strojny@uj.edu.pl

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Psychology

Received: *07 October 2019*
Accepted: *13 May 2020*
Published: *15 June 2020*

Citation:
Strojny PM, Dużmańska-Misiarczyk N, Lipp N and Strojny A (2020) Moderators of Social Facilitation Effect in Virtual Reality: Co-presence and Realism of Virtual Agents. Front. Psychol. 11:1252. doi: 10.3389/fpsyg.2020.01252

1. INTRODUCTION

The influence of other people on individuals performing a task is a common problem in real life. For example, bystanders are often present at various accident sites, possibly influencing performance of the rescuers. It is important to try to understand this influence and studying it in terms of the social facilitation effect appears to be a promising direction. Moreover, if such influence can be replicated in a virtual environment, rescuers (or other people exposed to the influence of bystanders during their work) could be trained in conditions similar to those which are present in real life. In the present paper, results of a study on social facilitation in virtual reality, specifically—in a rescue context, will be described.

Social facilitation occurs "when one animal increases or decreases its behavior in the presence of another animal which does not otherwise interact with it" (Guerin, 2010). Performance is improved for easy tasks (facilitation) and deteriorated for difficult ones (inhibition). For more detailed description of social facilitation see former works (Zajonc, 1965; Bond and Titus, 1983; Baron, 1986). Although this phenomenon has been known for a long time in psychology (Triplett, 1898), researchers are still far from full understanding of it (Cottrell, 1972; Baron, 1986; Huguet et al., 1999). One of several theoretical controversies particularly relevant to current study regards the issue of the *mere presence* vs. *audience* to trigger the social facilitation effect. According to early definition, the sufficient condition of social facilitation occurrence is the presence of others, even if the actor is not an object of their interest (Zajonc, 1965). On the other hand, further studies showed that it is not enough—others have to be focused on the actor (Cottrell et al., 1968). Because the current study was set in a VR depicting specific task (rescue action), the presence of victims (and no bystanders) was necessary in both conditions due to the ecological validity. Being aware of the controversy mentioned earlier, we decided to manipulate with the presence of others being able to observe an actor (bystanders) assuming the presence of victims will not cause the studied effect since they are preoccupied thus unable to observe the actor.

With the development of virtual reality (VR) technologies, researchers have begun to explore the impact of computer-generated agents in virtual environments (VEs) on users in terms of the social facilitation effect. Several studies examining this phenomenon in VR have been published, but according to the recent review their results are not consistent (Sterna et al., 2019). To the best of our knowledge, the full social facilitation and inhibition effect in easy and difficult tasks respectively has been shown only once in VR (Park and Catrambone, 2007). The possibility of its occurrence is supported by the results of other studies in which only social facilitation took place (Pan and Hamilton, 2015; Murray et al., 2016). In several other studies social inhibition was observed (Hoyt et al., 2003; Zanbaka et al., 2007; Emmerich and Masuch, 2016). However, other studies report a null effect (Hayes et al., 2010; Baldwin et al., 2015; Pan and Hamilton, 2015). This discrepancy may stem from methodological shortcomings, but it is possible that other unrevealed variables moderate the relationship. The moderator may affect the direction and/or strength of the relation between dependent and independent variables. We believe that Co-presence, Sense of Presence and some aspects of Realism may play a role here since they are related to subjective impression of being among others in virtual reality. Because the feeling of being in the company of others plays a pivotal role in social facilitation effect, we believe that these variables are able to affect the strength (but not the direction) of this relationship (not necessarily lowering it to zero given the subjective nature of moderators being proposed).

Previous studies have proved the importance of perceived presence (*sense of presence*, defined as the "sense of being there" in a virtual environment, or a human reaction to the experiences the technology delivers; Slater, 2003) in evoking desired reactions to VR (Poeschl and Doering, 2014; Riva et al., 2014). However, in the light of a recent meta-analysis it is also possible that integrating different factors of the multidimensional construct of sense of presence into a single score may be unable to capture the key characteristics responsible for evoking these reactions (Ling et al., 2014). Moreover, realism, defined as the fidelity of simulation—how accurate is the replication of the real environment and objects in virtual reality (Bowman and McMahan, 2007; Poeschl and Doering, 2013) may also play a role in one's responses to a virtual environment. Perhaps the social aspects of realism (understood as impression of fidelity of agents located in VR) play a key role here, particularly in case of phenomena closely related to social interactions, such as social facilitation. Another variable which might be of interest when trying to understand human reactions in a virtual environment, is co-presence—the impression of being in the environment with others, even when they are not physically present and even when they are not humans, but computer-generated agents (Youngblut, 2003). It has been proved that it may be crucial for evoking desired reactions to socially interactive VR (Poeschl, 2017; Felnhofer et al., 2019). Based on the past results, one might expect realism (in the social aspects in particular), sense of presence and co-presence to affect the occurrence of the social facilitation effect, but researchers have not yet controlled these variables.

Summarizing, to shed more light on the relationship between subjectively assessed social characteristics of VR (co-presence, sense of presence, realism) and social facilitation, we conducted a study in which for the first time to our knowledge the level of these characteristics was controlled. It was done in order to determine if they moderate the occurrence of the social facilitation effect evoked by computer-generated agents in VR. The presented study was preceded by an exploratory one, in which we found an interactive influence of spectators' presence and realism on subjectively assessed performance. To capture the social facilitation effect in terms of objective performance, we conducted the study described herein. Since the task we used was well-known by the participants (emergency procedure, professional fire-fighters), we expected the social facilitation (not inhibition) effect. We hypothesized that co-presence (Hypothesis 1), sense of presence (Hypothesis 2), and realism (Hypothesis 3) would moderate the relationship between audience presence and performance: high level of co-presence, sense of presence, and realism separately would be a condition of occurrence of social facilitation, while low level would not.

2. MATERIALS AND METHODS

2.1. Participants

Participants were recruited at the College of the State Fire Service and firefighting units in Cracow (Poland); all of them had undergone at least one year of training and had participated in real-life rescue operations. This research was accepted by the Ethical Committee at Jagiellonian University Institute of Applied Psychology. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The participants received a T-shirt after participating in the study. There were no defined exclusion criteria.

In total, 48 men ($M_{age} = 22.52$, $SD_{age} = 4.55$) participated in the described part of the study[1]. The lack of female participants is a consequence of the gender structure of the firefighting profession. Only a small number of women were enrolled in the aforementioned school and worked in the firefighting units. They were not drawn for the study described in the present paper.

The participants were randomly assigned to one of the conditions (23 in the experimental condition with virtual bystanders and 25 in the control condition without such bystanders). There was no age difference [$t_{(46)} = 1.08$, $p = 0.284$] between conditions. None of participants reported problems with perception of VR, and all of them had previously learned to control the simulator (they participated in two previous iterations of the study where the same simulator was used and they were also instructed about all possible actions and commands shortly before the experimental task).

2.2. Procedure

Firstly, participants were briefly interviewed and equipped with apparatus for measuring physiological variables (ECG, ICG, EDA)[2]. Directly before the task started, the participants were informed that they would be asked to perform the Medical Rescue Sequence detailed in the National Firefighting Rescue System documentation[3] For the full description of the procedure (see **Figure 1**). The task had a fixed 5 min duration. The experimenter received a confirmation of knowledge of the procedure from each participant.

Then, the VR simulation took place. Participants wore a HTC Vive head-mounted display (HMD) with hand-held controllers and headphones. The HMD was connected to a PC with a 3.40 GHz Intel Core i7 processor, 16 GB of RAM and a NVIDIA GeForce GTX 1080 graphics card. The simulation was developed with use of the Unity engine and depicted a collision between a car and a group of six pedestrians on an intersection in a small town (see **Figure 2A**). The possible interactions with victims were: conducting the SAMPLE interview[4], checking several physical parameters (pulse, pain reaction, breathing, airways, and capillary recurrence), covering the person with a blanket, dressing the wounds, performing resuscitation. Moreover, passive oxygen therapy could have been performed with the equipment from the medical bag. All actions were controlled with text commands in a context menu. The menu consists of a list of possible actions, which could be accessed when pointing with a hand-held controller at a specific virtual agent and pushing one of the buttons. To choose between the available actions, the participant had to scroll on a trackpad of the controller. For an example of the menu interface for a victim (see **Figure 3A**). For photos of the experimental setup see the **Supplementary Materials**, and for a video with an example of actions conducted in the simulator used in the study see: https://www.youtube.com/watch?v=Bn_E4wX1RPE&feature=emb_logo.

The experimental condition contained an additional group of three bystanders located next to each victim (18 virtual agents in total, see **Figure 2B**). The bystanders were animated: they performed simple gestures at random moments, followed the participant with their eyes and some of them recorded the event with smartphones. They could be asked (through the list of actions in the menu) whether they were a doctor (and always responded "No") and they could be told to move away (what they always did when asked to). For an example of the menu interface for a bystander (see **Figure 3B**).

For safety reasons, the experimenter was present in the room during the simulation, but she remained silent and could not be seen by participants. After the experimental task, the participants completed questionnaires administered using a PsychoPy script (Peirce, 2007, 2009).

2.3. Measures

The questionnaires used in this study were completed in polish language. The Polish versions were created on basis of the back translation procedure (Brislin, 1970). First of all, two independent German-speaking professional translators translated all of the items from German to Polish. In the next step, two different translators translated back into German. Then, we compared the original versions with those obtained during translation procedure. There were no major discrepancies between them.

2.3.1. Manipulation Check

To assess whether participants noticed the bystanders, we asked them whether they perceived the following elements of the environment: a dog, a drone, policemen, a toy, and bystanders (critical question). Some of them were stimuli in other experimental conditions and some of them were masking items.

2.3.2. Co-presence

The Polish version of The Co-Presence and Social Presence in Virtual Environments Scale (C-PS, Poeschl and Doering, 2015) was administered after the VR session to measure co-presence. It consists of four factors: Reaction to Virtual Agents (4 items), Perceived Virtual Agents' Reaction (4 items), Impression of Interaction Possibilities (4 items), (Co-)Presence of Other People (3 items). The items are rated on a 5-point Likert scale from −2 to 2. We evaluated internal consistency using the reliability analysis. The obtained Cronbach's coefficient was high ($\alpha = 0.89$).

[1]The current study is a part of longitudinal study with three experimental and one control group. For the purpose of this report, we present data from one of the experimental conditions and the control condition from the third iteration of the study. The other conditions were unrelated to social facilitation.

[2]Data gathered with this equipment is irrelevant to social facilitation and is discussed in another paper.

[3]Available in Polish at https://www.straz.gov.pl/download/1854.

[4]SAMPLE is an acronym for six basic questions in a medical assessment: symptoms, allergies, medications, past medical history, last oral intake, and events leading up to present injury.

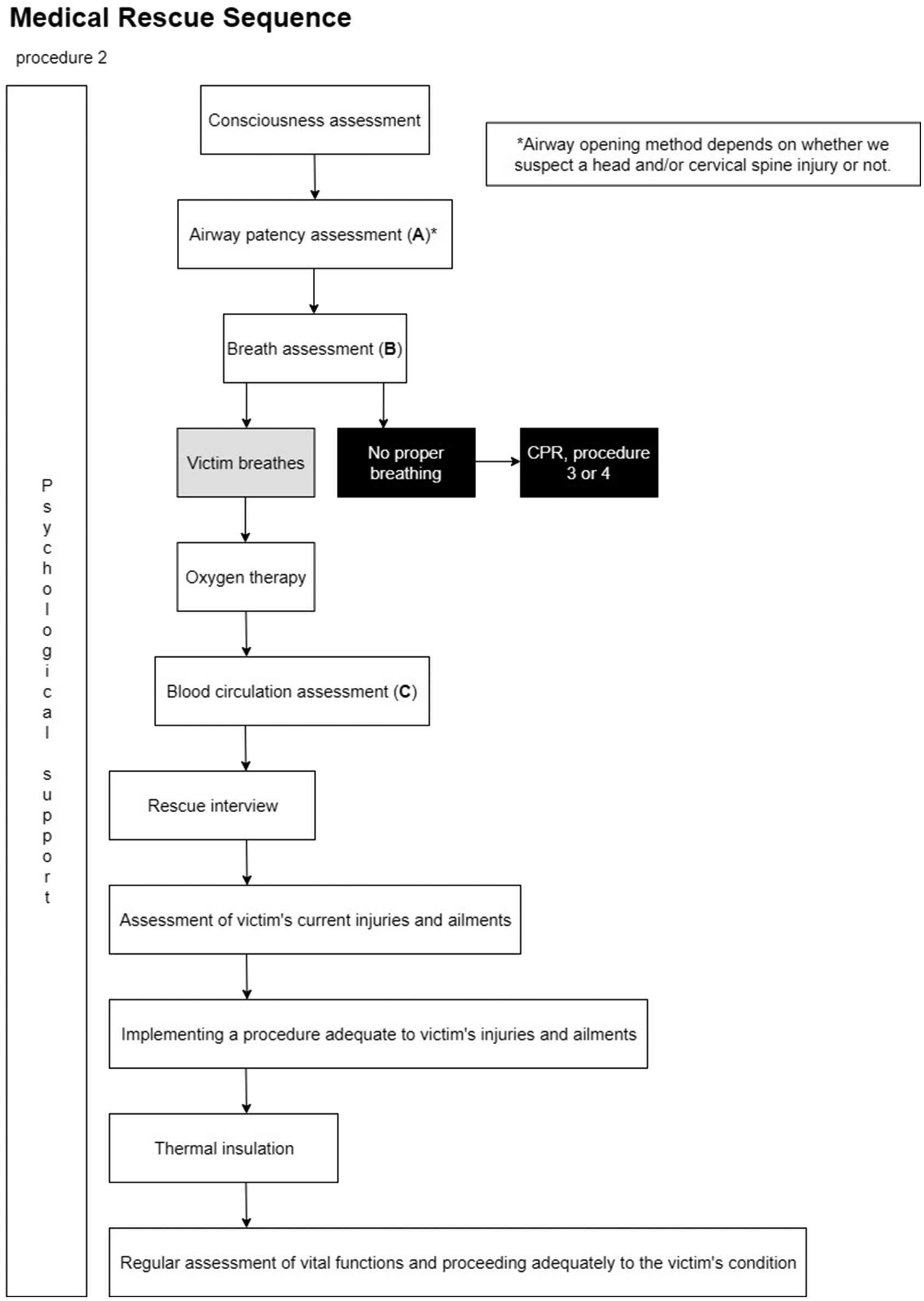


FIGURE 1 | The medical rescue sequence used for the experimental task in the study.

2.3.3. Sense of Presence

The Polish version of 14-item iGroup Presence Questionnaire (IPQ, Schubert et al., 2001) was used to assess sense of presence. It contains 14 items on three subscales: (1) Spatial Presence (6 items), (2) Involvement (4 items), and (3) Realism (4 items). All of the items are rated on 7-point Likert scale from −3 to 3; overall

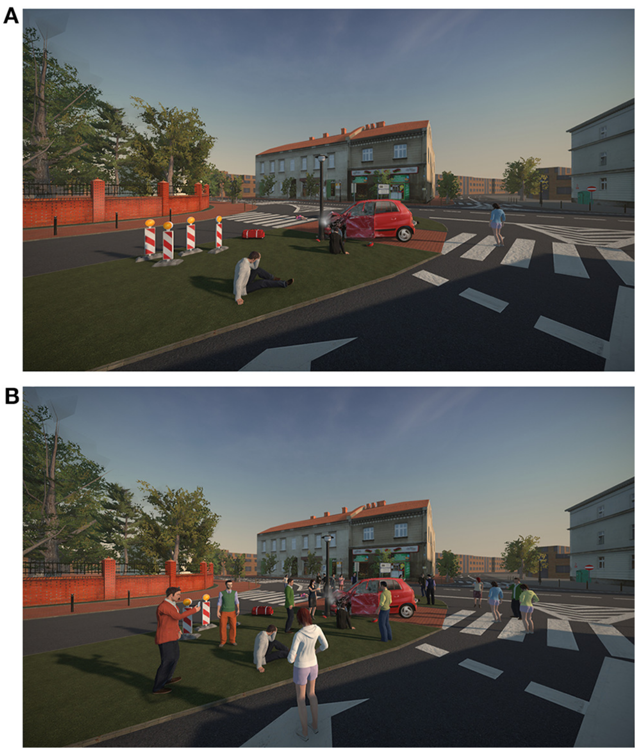


FIGURE 2 | Screenshots showing the virtual environment used in the control **(A)** and experimental condition **(B)**. This point of view was available for the participants.

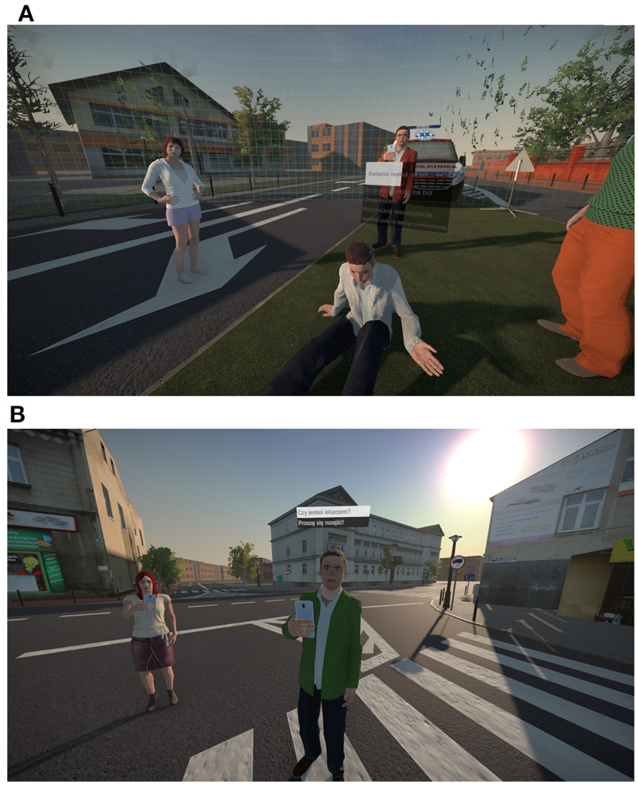


FIGURE 3 | Screenshots showing the context menu in use: for a victim, during an action **(A)** and for a bystander, before choosing an action **(B)**.

score ranges from −42 to 42. We evaluated internal consistency using the reliability analysis. The obtained Cronbach's coefficient was satisfying ($\alpha = 0.77$).

2.3.4. Realism

The German VR Realism Scale in Polish version (VRRS, Poeschl and Doering, 2013) was used to assess perceived realism of simulation. In total it consists 14 items rated on 5-point Likert scale from −2 to 2; overall score ranges from −28 to 28. Thirteen items are divided on three subscales: Scene Realism (5 items), Audience Behavior (4 items), and Audience Appearance (4 items), remaining one item regards sound realism. We evaluated internal consistency using the reliability analysis. The obtained Cronbach's coefficient was very high ($\alpha = 0.92$).

2.3.5. Performance and Activity

To quantify performance, we developed a script automatically logging the correctness of actions taken during the exercise on a basis of National Firefighting Rescue System documentation. Since the rescue procedure we used is defined in the form of an algorithm, we could precisely determine the correctness of the participants' actions. In order for each individual action to be considered "correct," it had to be taken exactly when the rescue procedure foresees it. Otherwise (e.g., performing an unforeseen action or confusing the order), the single action was considered an "error." We counted all actions taken in the wrong order ("errors"). No feedback on the performance was given during the session. An erroneous action could not be corrected, but further actions were calculated according to the rule presented above—it was feasible to avoid further mistakes simply by performing subsequent actions in the correct order resulting from previous decisions.

Moreover, it was possible to count the total sum of taken actions, regardless of their correctness. Therefore, such index was calculated in order to test the possible impact of audience presence on activity.

2.3.6. Other Measures

Participants completed the Polish versions of several other questionnaires at the end of the study: Self-Assessment Manikin (SAM, Bradley and Lang, 1994), The Scale of Emotions (Wojciszke and Baryła, 2005), The Stress Appraisal Questionnaire (SAQ, Włodarczyk and Wrześniewski, 2010), Simulator Sickness Questionnaire (SSQ, Kennedy et al., 1993), The Scale of Aesthetics (Chevalier et al., 2014), NASA Task Load Index (NASA-TLX, Hart and Staveland, 1988; Zieliński and Biernacki, 2010). These tools are not of interest to the hypotheses formulated in the present paper, therefore the analyses concerning the aforementioned variables will not be reported herein.

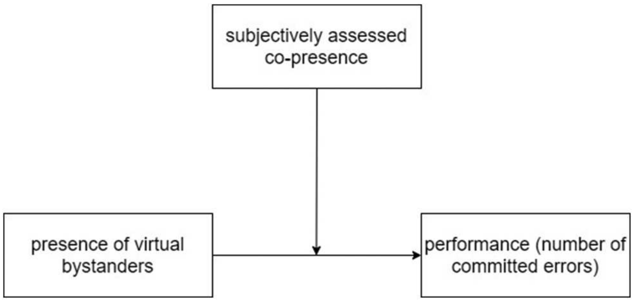


FIGURE 4 | Graphical representation of the moderation analysis (Hayes, 2013). Other tested moderators (sense of presence and perceived realism) could be put in place of co-presence in this visualization.

2.4. Data Analysis

2.4.1. Null Hypothesis Significance Testing and Equivalence Testing

For a proper interpretation of the results it was decided to firstly exclude that the variability in data stems from sources other than the social facilitation effect. Performance could vary between groups not only because of the social facilitation effect, but also because of differences in terms of the mere number of actions conducted by the participants. Therefore, it was checked whether the number of actions in the groups is statistically equivalent. In such cases equivalence testing (two one-sided *t*-tests—TOST) is used (Limentani et al., 2005; Lakens et al., 2018, 2020). For other hypotheses, null hypothesis significance testing (NHST) was used.

2.4.2. Moderation Analysis

For the verification of research hypotheses, moderation analysis was chosen to be used. Such analysis tests the influence of a third variable (moderator) on the relationship between independent and dependent variables. Moderation analysis is used to answer the question which conditions have to be met for an effect to occur (see **Figure 4** for an example of conceptualization of moderation). It is calculated based on a regression model. In the case of the present analysis, the model is as follows:

$$performance = b_0 + b_{1condition} + b_{2moderator} + b_{3(condition*moderator)} \quad (1)$$

Moderator can enhance, reduce or change the influence of predictor on the outcome variable. In moderation analysis with a single moderator (as in the case of the reported study), three main effects are calculated: the separate influences of the predictor and moderator and the interaction of these variables (Fairchild and MacKinnon, 2009).

If the interaction effect is significant, simple main effects are calculated. Such effects indicate on which level of the moderator the influence exists. In the PROCESS macro (Hayes, 2013), subsamples for simple main effects can be chosen with thresholds of +/− 1 SD or 16th, 50th, 84th percentiles on the moderator. The latter technique was used to differentiate between high and low levels of considered moderators in the reported analyses. Additionally, the Johnson-Neyman technique can be used in order to determine the region of significance. This technique is also useful for preparing data for visualization (Johnson and Fay, 1950; D'Alonzo, 2004). Both these techniques will be used in the present paper.

Standardized effect sizes for the moderation analysis were calculated according to Bodner's (2017) guidelines. Thanks to such approach, effect sizes in separate analyses can be compared in terms of strength. Standardized effect sizes higher than 0.4 and lower than 1.0 are considered *small*, higher than 1.0 and lower than 1.6 are *medium* and higher than 1.6—*large*.

3. RESULTS

Data were analyzed with Imago Pro 5.0, the PROCESS (Hayes, 2013) macro and the R environment—the TOSTER package (Lakens, 2017). Data from four participants were excluded due to technical problems with performance logging. Finally, data from 44 participants (22 in each group) were analyzed. For the analyses, the control group was dummy coded as 1 and the experimental group was dummy coded as 2.

3.1. Manipulation Check

The manipulation was successful, only 5 out of 22 participants in the experimental condition reported that they did not notice the bystanders, and 5 out of 22 participants in the control condition reported they noticed bystanders when they could in fact not see them.

3.2. Audience Impact on Activity

To check the possibility that audience presence affected participants' activity (the number of actions), we compared both conditions in such terms. We found that the difference in number of actions in both conditions was not statistically significant and slightly above the "medium" threshold in terms of effect size. ($M_{control} = 25.32$, $SD_{control} = 8.29$, $M_{audience} = 21.27$, $SD_{audience} = 6.73$, $t(42) = 1.78$, $p = 0.08$, $d = 0.53$).

Because the *t*-test yielded insignificant results, we applied the TOST procedure (Limentani et al., 2005; Lakens et al., 2018, 2020). We determined the smallest effect size of interest (SESOI) on the basis of results obtained in the previous iteration in the longitudinal study. Using the study's alpha level and sample size, we calculated the critical effect size (Cohen's *d*, Cohen, 1992). The equivalence test was non-significant, $t_{(42)} = 0.320$, $p = 0.625$, given equivalence bounds of −0.44 and 0.44 and an alpha of 0.05. Based on both the equivalence test and the null-hypothesis test, we may conclude that the observed effect is statistically not different from zero and statistically not equivalent to zero.

3.3. Audience Impact on Performance

In order to test the main effect of audience presence on performance, we separately compared both conditions in terms of errors made. We found main effect of bystanders presence on performance to be statistically not significant and small in terms of effect size ($M_{control} = 9.73$, $SD_{control} = 3.71$, $M_{audience} = 8.27$, $SD_{audience} = 3.27$, $t_{(42)} = 1.38$, $p = 0.175$, $d = 0.42$). Thus, we did not observe the main effect of social facilitation, it is possible

TABLE 1 | Simple linear regression results for considered moderators of performance (social facilitation).

Predictor	β	p	95 % CI[a]		R^2
Co-Presence	−0.006	0.968	−1.76	1.69	< 0.01*
C-PS: Reaction to virtual agents	−0.012	0.939	−1.17	1.08	<.01*
C-PS: Perceived virtual agents' reaction	0.197	0.200	−0.49	2.26	0.04
C-PS: Impression of interaction possibilities	−0.041	0.791	−1.48	1.13	< 0.01*
C-PS: co-presence of other people	−0.221	0.148	−2.49	0.39	0.05
Sense of presence	−0.065	0.674	−0.50	0.33	< 0.01*
IPQ: Spatial presence	−0.068	0.662	−1.31	0.84	< 0.01*
IPQ: Involvement	0.084	0.589	−1.22	0.70	< 0.01*
IPQ: Realism	−0.007	0.964	−0.95	0.91	< 0.01*
Realism	−0.041	0.791	−1.75	1.34	< 0.01*
VRRS: Scene realism	0.068	0.663	−1.19	1.85	< 0.01*
VRRS: Audience behavior	−0.059	0.704	−1.51	1.03	< 0.01*
VRRS: Audience appearance	−0.118	0.446	−1.66	0.74	0.01
VRRS: Sound realism	−0.022	0.889	−1.29	1.12	< 0.01*

*R^2 < 0.01—predictor explains <1% of variances.
[a] 95% Confidence interval.

TABLE 2 | *T*-test results comparing the experimental and control group on considered moderators.

Predictor	Experimental		Control		*t*-test			
	M	*SD*	*M*	*SD*	*t*	*df*	*p*	*d*
Co-presence	−0.52	0.55	−0.29	0.71	1.22	42	0.230	0.37
Sense of presence	−0.28	0.90	−0.18	0.88	0.38	42	0.702	0.12
Realism	0.14	0.68	0.08	0.75	0.29	42	0.777	−0.09

the effect was to small to met conventional criterion of alpha 0.05 with the sample size we used.

3.4. Interaction of Perceived VR Characteristics and the Presence of Agents in the Performance

We conducted moderation analysis according to the steps described in *Data Analysis* section. None of the considered moderators affected the performance on its own (see **Table 1** for main effects of hypothesized moderators on performance). Also, *t*-test was conducted to evaluate the influence of audience presence on the considered moderators—co-presence, sense of presence and realism. The analysis did not reveal the significant effects (see **Table 2**).

3.4.1. Co-presence as a Moderator of Audience Presence and Performance Relationship

According to our first hypothesis, we analyzed the interactional influence between audience presence and subjectively assessed co-presence on performance. We found a statistically significant interaction ($\beta = -2.48$, $b = -3.88$, $SE = 1.68$ $p = 0.026$, $r^2_{increase} = 0.11$). Calculated simple main effects indicated that only in the case of high assessment of co-presence did the audience cause social facilitation and the effect was of moderate strength ($\beta = -4.08$, $SE = 1.52$, $p = 0.010$, *95% CI:* −7.15, −1.02, $\delta = -1.22$). For co-presence assessed as low, the relationship was insignificant and the effect size was very small ($\beta = 0.42$, $b = 0.42$, $SE = 1.33$, $p = 0.75$, *95% CI:* −2.26, 3.10, $\delta = 0.13$). This dependency is shown in **Figure 5A**. Additionally, we used the Johnson-Neyman technique to determine the specific values of co-presence at which the moderation occurred. We found that relatively high (higher than −0.269) co-presence levels resulted in social facilitation in the presence of bystanders. Above that threshold scored 19 participants (11 in control group and 8 in experimental group).

Moreover, we decided to check whether specific subscales are moderators of described dependency. We found two subscales of Co-Presence (Perceived Virtual Agents' Reaction and Impression of Interaction Possibilities) to be statistically significant moderators (see **Table 3** for all interaction results). Accordingly to Hypothesis 1, we found an interactional effect of audience presence and subjectively assessed co-presence on performance (social facilitation).

3.4.2. Sense of Presence as a Moderator of Audience Presence and Performance Relationship

According to our second hypothesis, we analyzed the interactional influence between audience existence in the scene and subjectively assessed sense of presence on performance. The analysis ruled out the role of this variable as a moderator, due to

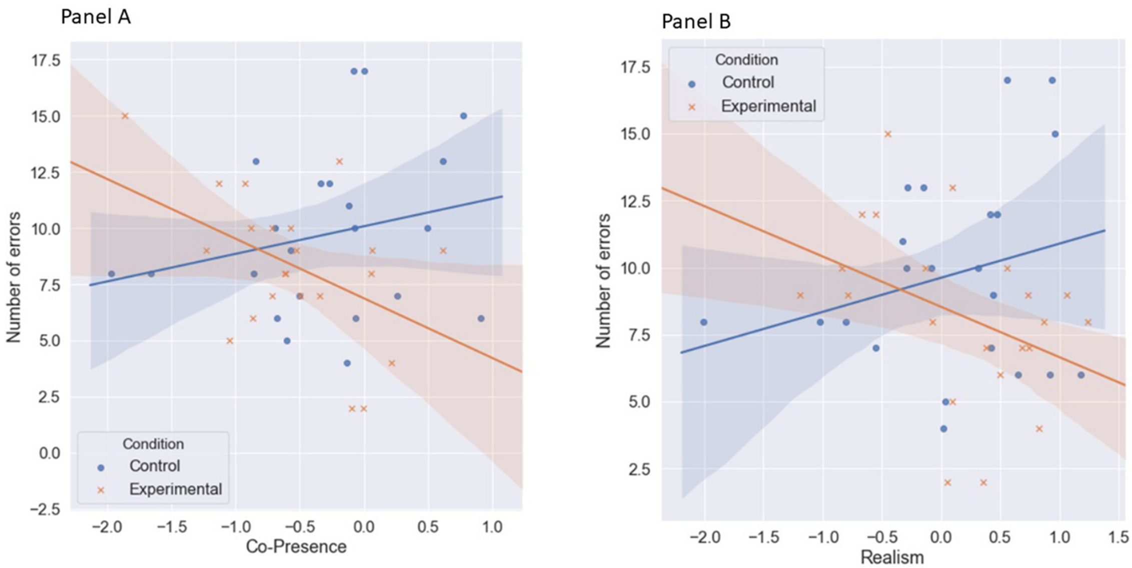


FIGURE 5 | Graphical representation of the interaction—the number of errors committed in VR simulation depending on the interaction of the presence of audience and the subjectively assessed VR characteristics: co-presence **(A)** and sense of realism **(B)**.

TABLE 3 | Interaction between subjectively assessed VR characteristics and the presence of agents on performance.

Predictor	β	b	SE	t	p	95%CI	
Co-Presence x Condition	−2.48	−3.88	1.68	−2.31	0.026*	−7.28	−0.49
Reaction to virtual agents × Condition	−1.02	−1.04	1.17	−0.89	0.378	−3.41	1.33
Perceived virtual agents' reaction × Condition	−2.24	−2.85	1.39	−2.05	0.047*	−5.66	−0.04
Impression of interaction possibilities × Condition	−2.23	−2.66	1.31	−2.02	0.049*	−5.32	−0.001
(Co-)Presence of other people × Condition	−1.80	−2.42	1.49	−1.62	0.113	−5.44	0.60
Sense of presence × Condition	−0.64	−0.24	0.41	−0.59	.556	−1.07	0.59
Spatial presence × Condition	0.14	0.12	0.97	0.13	0.900	−1.84	2.09
Involvement × Condition	−0.58	−0.57	1.09	−0.53	0.601	−2.77	1.62
Realism × Condition	−1.28	−1.08	0.95	−1.15	.259	−3.00	0.83
Realism × Condition	−2.24	−3.15	1.46	−2.15	0.037*	−6.11	−0.19
Scene realism × Condition	−1.36	−1.88	1.55	−1.22	0.230	−5.01	1.24
Audience appearance × Condition	−2.17	−2.38	1.15	−2.07	0.044*	−4.71	−0.06
Audience behavior × Condition	−2.23	−2.57	1.99	−2.14	0.038*	−4.99	−0.15
Sound realism × Condition	−1.48	−1.62	1.22	−1.33	0.189	−4.08	0.84

*$p < 0.05$.

the statistical insignificance of the results ($\beta = -0.64$, $b = -0.24$, $SE = 0.41$ $p = 0.556$, $r^2_{increase} = 0.01$).

3.4.3. Sense of Realism as a Moderator of Audience Presence and Performance Relationship

According to our third hypothesis, we analyzed the interactional influence between audience existence in the scene and subjectively assessed sense of presence on performance. We found a statistically significant interaction ($\beta = -2.24$, $b = -3.15$, $SE = 1.46$ $p = 0.037$, $r^2_{increase} = 0.10$). Calculated simple main effects indicated that high level of realism resulted in statistically significant, positive relationship between virtual agents' presence and subjects' performance—social facilitation, with a moderately strong effect size ($\beta = -3.79$, $SE = 1.49$, $p = 0.015$, *95% CI:* -6.80, -0.77, $\delta = -1.12$). For low perceived realism, the relationship was insignificant and the effect size was very small ($\beta = 0.90$, $SE = 1.49$, $p = 0.55$, *95% CI:* -2.12, 3.92, $\delta = 0.27$). This dependency is shown in **Figure 5B**. Then, we used the Johnson-Neyman technique in order to determine the specific values of realism at which the moderation occurred. We found that relatively high and positive (higher than 0.347) realism levels resulted in social facilitation in the presence of virtual bystanders. Above that threshold scored 19 participants (10 in control group and 9 in experimental group).

Moreover, we decided to check whether specific subscales are moderators of the described dependency. We found two subscales: Audience Appearance and Audience Behavior to be statistically significant moderators (see **Table 3** for interaction analysis results).

4. DISCUSSION

The current results provide support for the moderating role of co-presence (Hypothesis 1) and realism (Hypothesis 3) on social facilitation in VR. On the other hand, sense of presence did not play the same role (Hypothesis 2). In order to successfully evoke the social facilitation effect in the presence of computer-generated bystanders, a certain level of (subjective) co-presence in the VE must be achieved. Realism plays an analogous role. In this study, mere presence of virtual agents on the simulated accident site was enough to improve the trainees' performance (reduce the number of erroneous actions), but only for those who evaluated co-presence and realism as relatively high.

We found no main effect of bystanders' presence on performance, which is in line with some previous studies (Hayes et al., 2010; Baldwin et al., 2015; Pan and Hamilton, 2015). Although the main effect of social facilitation seemed to be statistically insignificant, the effect sizes are of medium strength. Perhaps, a non-moderated social facilitation occurs in VR but it is more difficult to detect than in the real world. In this regard, our results may shed new light on the inconclusiveness of previous studies in which social facilitation did not always occur — it is possible that social realism (co-presence and realism) was not taken into account there. We can only speculate that, in some cases, social facilitation might have taken place, but only among participants who experienced high social realism in the VE. Not only the results for high levels of moderators were statistically significant, but the observed effects were of medium strength (−1.22 for co-presence and −1.12 for sense of realism). At the same time, the effects for low levels of moderators were insignificant and very weak (0.13 for co-presence and 0.27 for sense of realism). This can be interpreted as consistent with the general theory regarding social facilitation, which assumes that mere presence of real actors is enough for evoking the effect. Since the vast majority of previous experiments was conducted in the presence of real observers, it can be assumed that they were perceived as real and interactive. Moreover, some theorists emphasized the importance of establishing a basic psychological relationship between the actor and the observer, claiming that mere physical presence is insufficient also in the real world (Cottrell et al., 1968; Cottrell, 1972). Moreover, no physical presence (even symbolic) is needed to produce the effect: for example, social facilitation takes place in online auctions and the level of symbolic presence plays a role in the absence of a physical presence in this case (Rafaeli and Noy, 2002).

The above considerations are consistent with the results we obtained in exploratory analyses described earlier. Since we inspected the role of individual subscales, we were able to identify those which played the crucial role in the interactions we predicted — note the Realism scale we used includes four aspects of realism—scene, sound, audience appearance and audience behavior. As it turned out, only the last two of them, exactly those related directly to social context, played a role in the moderation we found. Findings regarding Co-Presence are also consistent. In this case two subscales directly related to social interactions played a similar role. The whole picture seems to be complemented by the fact that the third scale (Presence, regarding the subjective sense of presence), which did not turn out to moderate social facilitation, does not include any subscale related to social interactions. Keeping in mind the exploratory nature of these results, further detailed research is needed, but these results clearly suggest the crucial role of social realism in evoking phenomena based in social interactions in VR.

Our results should be also analyzed from the perspective of the new terminology of phenomena regarding illusion of VR realism proposed by Slater (2009). The idea of two orthogonal components of realistic response to VR seems to fit well with our results. Slater proposed the term "place illusion" (PI) as the name of qualia of "being there" which, accordingly to Slater's conclusions, is rather of perceptional than cognitive nature, it is often called "Presence." The second dimension called "plausibility illusion" (Psi) refers to the illusion that events being depicted are actually occurring. It is more difficult to achieve and more susceptible to being broken. Both components are needed to evoke reactions similar to expected in reality called "response-as-if-real" (RAIR). In the case of the present study social facilitation was the RAIR. From this perspective, one could notice that variables we tested as moderators may be assigned to PI or Psi. Presence undoubtedly belongs to PI. Also two aspects of Realism (scene and sound) are rather of perceptual nature while audience behavior should rather be seen as belonging to the Psi domain. Only audience appearance may seem to be difficult to classify, but looking carefully at items (e.g., "Virtual humans in their entirety seemed to be authentic for this occasion."), one can see that they place a lot of emphasis not so much on the appearance itself as on the adequacy of the appearance to the situation, which may suggest assigning this subscale to Psi also. Attempt to classify Co-presence subscales could also be made—inspection of items may suggest that three of four would rather belong to Psi domain (Perceived Virtual Agents' Reaction, Impression of Interaction Possibilities, and Reaction to Virtual Agents) and the last one would be difficult to classify (Co-Presence of Other People). Remembering that the above reasoning was made post-factum, it is easy to see that the role of moderators of the relationship between the actual presence of observers on the stage and performance was played only by variables assigned to Psi. In other words RAIR depended on the sufficiently high level of Psi-related variables. In this light our findings may suggest that Psi may be particularly important in evoking social reactions involving high-order cognitive processes.

Our results may contribute to the theoretical dispute about the mechanism of the social facilitation effect. According to one of the most popular explanations, attentional conflict caused by the physical presence of other people is the basis of the social facilitation phenomenon (Baron, 1986). If this is true in the case of VR, one could expect social facilitation to occur in VR training simply due to the existence of (social) distractors. Thus, the effect should occur regardless of the level of the subjective social realism, although it was not observed in our study. On the contrary, we found this effect to be dependent on the perceived level of co-presence and realism of the virtual agents. Therefore, it seems that Baron's explanation does not entirely match our results, according to which the distractors do not only need to *exist*, but they have to be perceived as *realistic* and *(co-)present* as well for the social facilitation effect to occur.

All of the aforementioned observations lead to a conclusion that it is useful to implement social stimuli in VR training simulators, especially in the case when real people are present during actual implementation of the task being trained. It would be beneficial to implement virtual agents in the simulation, as it was confirmed that such agents can, to some extent, evoke effects similar to those observed in the real world. Bystanders are very often present at accident sites. They influence the rescuers' emotions and may sometimes actively hinder the operation (as reported by the firefighters themselves; Strojny et al., 2018), therefore they should also be included in training procedures somehow. Moreover, while designing training simulators, attention should be paid to increasing the experience of social realism, e.g., through implementing realistic animations or a possibility to interact with the virtual agents. Moreover, social realism could be further increased by creating more diverse groups of bystanders. Such variations should also be tested in further studies.

There are several limitations to our study. We tested only the positive side of the effect (social facilitation in contrast to social inhibition). Since we conducted our study on specific participants, we could not find the inhibition effect during a task which was easy for them. However, using an existing procedure instead of an abstract task may increase the ecological validity of the study. Subsequent research should address this issue either by recruiting participants from a general population (not familiar with rescue procedures) or by manipulating the task in order to transform a well-practiced procedure into a counter-intuitive one. In both cases we would expect the social inhibition effect. Moreover, the characteristics of the population from which the participants were recruited led to the lack of representation of women in the study, which also may be viewed as a limitation of the study. Therefore, further studies with a more diverse group of participants should be conducted in order to improve generalizability of the results.

The hypotheses we formulated regarded all of three potential moderators separately. However, it is plausible that the moderators are interrelated—we did not take it into consideration during experiment preparation. A model that includes three moderators at the same time would require a much larger sample to draw conclusions from it. We see our results as a first step in analysing the phenomenon of social facilitation in VR in context of subjective perception of the virtual environment (i.e., co-presence, sense of presence, or realism). Testing whether these variables are related to each other should be considered in further studies on this issue.

Besides, we operationalized VR characteristics as self-reported. Further studies should use experimental manipulation of these variables. Moreover, we used a simple method of assessing the participants' performance — namely, the number of committed errors. In further studies it could be useful to develop a more sophisticated performance measure (e.g., not only the correctness but also the speed). Moreover, we used a well-documented, but still highly specific activity. In future studies more general tasks should be used to strengthen the external validity of the results.

Lastly, it could be viewed as a big limitation of the study that virtual agents—the victims—were present on the scene in both conditions. The understanding of social facilitation effect as the influence of *the mere presence* of other people, as it was firstly defined by Zajonc (1965) and showed in some studies (e.g., Markus, 1978; Platania and Moran, 2001) may lead to expectation of social facilitation occurrence in both conditions. In this case, due to design of the scene in our study we cannot draw conclusions about social facilitation. *Some* virtual people were *merely present* on the scene regardless of the condition. Adding the virtual bystanders would be viewed in this case as a change in *quantity*, not *quality* of the stimuli. However, mere presence might not be enough for evoking the effect, as it was proposed by Cottrell et al. (1968). In his study, mere presence of other people (not interested in the person performing the experimental task, not looking at them etc.) did not evoke the effect, while presence of *audience* (people observing the participant and overly interested in them). Therefore, it may not be the fact that social facilitation is evoked by other people being there, but by the roles they have, the affective states they evoke in participants (i.e., anticipation of evaluation by the audience). This interpretation matches our results—the virtual bystanders were in fact similar to Cottrell et al's (1968) audience—their purpose was to observe the participant and even to "record" their actions with smartphones. Therefore, we decided to go with Cottrell's argumentation. However, more studies on this issue should be conducted, with focus on distinguishing between effects of mere presence and (evaluative) audience presence.

In sum, co-presence and realism seem to play an important moderating role in the relationship between the presence of computer-generated agents and social facilitation in VR. This finding is crucial considering the increasing use of similar tools to teach complex skills in a social context.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethical Committee at Jagiellonian University Institute of Applied Psychology. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

PS created the main conception and study design, analyzed majority of the data, interpreted the results, drafted and prepared the final version of the paper. NL and ND-M performed the experiment, gathered, and partially analyzed data. AS provided substantial contribution to the conception of the experiment and took part in data interpretation. ND-M wrote a minor part of the paper. NL visualized data. All authors contributed to manuscript revisions, read and approved the submitted version.

FUNDING

This work was supported by the Polish National Centre for Research and Development under the grant Widespread Disaster Simulator—research and preparation for implementation (project number POIR.01.01.01-00.0042/16; the Smart Growth Operational Programme, sub-measure 1.1.1. Industrial research and development work implemented by enterprises) received by Nano Games sp. z o.o.

ACKNOWLEDGMENTS

The authors would like to thank Radosław Sterna, Krzysztof Rębilas, and Bartosz Wojciechowski for their help with the data acquisition, Gabriela Czarnek for sharing her methodological knowledge in the process of designing the procedure, Konrad Klocek and the whole Nano Games development team for creating the VR simulator used in the study and Barbara Giżycka for proofreading the article.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.01252/full#supplementary-material

REFERENCES

Baldwin, N., Branyon, J., Sethumadhavan, A., and Pak, R. (2015). "In search of virtual social facilitation effects," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 59* (Sage, CA; Los Angeles, CA: SAGE Publications), 90–94.

Baron, R. S. (1986). "Distraction-conflict theory: progress and problems," in *Advances in Experimental Social Psychology, Vol. 19*, ed M. P. Zanna (Waterloo, ON: Elsevier), 1–40.

Bodner, T. E. (2017). Standardized effect sizes for moderated conditional fixed effects with continuous moderator variables. *Front. Psychol.* 8:562. doi: 10.3389/fpsyg.2017.00562

Bond, C. F., and Titus, L. J. (1983). Social facilitation: a meta-analysis of 241 studies. *Psychol. Bull.* 94:265. doi: 10.1037/0033-2909.94.2.265

Bowman, D. A., and McMahan, R. P. (2007). Virtual reality: how much immersion is enough? *Computer* 40, 36–43. doi: 10.1109/MC.2007.257

Bradley, M. M., and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *J. Behav. Therapy Exp. Psychiatry* 25, 49–59. doi: 10.1016/0005-7916(94)90063-9

Brislin, R. W. (1970). Back-translation for cross-cultural research. *J. Cross Cult. Psychol.* 1, 185–216. doi: 10.1177/135910457000100301

Chevalier, A., Maury, A.-C., and Fouquereau, N. (2014). The influence of the search complexity and the familiarity with the website on the subjective appraisal of aesthetics, mental effort and usability. *Behav. Inform. Technol.* 33, 117–132. doi: 10.1080/0144929X.2013.819936

Cohen, J. (1992). A power primer. *Psychol. Bull.* 112:155. doi: 10.1037/0033-2909.112.1.155

Cottrell, N. B. (1972). *Social Facilitation. Experimental Social Psychology*. New York, NY: Holt; Rhinehart & Wilson.

Cottrell, N. B., Wack, D. L., Sekerak, G. J., and Rittle, R. H. (1968). Social facilitation of dominant responses by the presence of an audience and the mere presence of others. *J. Pers. Soc. Psychol.* 9:245. doi: 10.1037/h0025902

D'Alonzo, K. T. (2004). The johnson-neyman procedure as an alternative to ancova. *Western J. Nurs. Res.* 26, 804–812. doi: 10.1177/0193945904266733

Emmerich, K., and Masuch, M. (2016). "The influence of virtual agents on player experience and performance," in *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (Austin, TX: ACM), 10–21.

Fairchild, A. J., and MacKinnon, D. P. (2009). A general model for testing mediation and moderation effects. *Prevent. Sci.* 10, 87–99. doi: 10.1007/s11121-008-0109-6

Felnhofer, A., Hlavacs, H., Beutl, L., Kryspin-Exner, I., and Kothgassner, O. D. (2019). Physical presence, social presence, and anxiety in participants with social anxiety disorder during virtual cue exposure. *Cyberpsychol. Behav. Soc. Netw.* 22, 46–50. doi: 10.1089/cyber.2018.0221

Guerin, B. (2010). "Social facilitation," in *The Corsini Encyclopedia of Psychology*, eds I. B. Weiner and W. E. Craighead (Hoboken, NJ: John Wiley & Sons, Inc.), 1–2.

Hart, S. G., and Staveland, L. E. (1988). "Development of nasa-tlx (task load index): results of empirical and theoretical research," in *Advances in Psychology, Vol. 52* (North-Holland: Elsevier), 139–183.

Hayes, A. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. New York, NY: Guilford Press.

Hayes, A. L., Ulinski, A. C., and Hodges, L. F. (2010). "That avatar is looking at me! social inhibition in virtual worlds," in *International Conference on Intelligent Virtual Agents* (Berlin; Heidelberg: Springer), 454–467.

Hoyt, C. L., Blascovich, J., and Swinth, K. R. (2003). Social inhibition in immersive virtual environments. *Presence* 12, 183–195. doi: 10.1162/105474603321640932

Huguet, P., Galvaing, M. P., Monteil, J. M., and Dumas, F. (1999). Social presence effects in the stroop task: further evidence for an attentional view of social facilitation. *J. Pers. Soc. Psychol.* 77:1011. doi: 10.1037/0022-3514.77.5.1011

Johnson, P. O., and Fay, L. C. (1950). The johnson-neyman technique, its theory and application. *Psychometrika* 15, 349–367. doi: 10.1007/BF02288864

Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. (1993). Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *Int. J. Aviat. Psychol.* 3, 203–220. doi: 10.1207/s15327108ijap0303_3

Lakens, D. (2017). Equivalence tests: a practical primer for t-tests, correlations, and meta-analyses. *Soc. Psychol. Pers. Sci.* 1, 1–8. doi: 10.1177/1948550617697177

Lakens, D., McLatchie, N., Isager, P. M., Scheel, A. M., and Dienes, Z. (2020). Improving inferences about null effects with bayes factors and equivalence tests. *J. Gerontol. Ser. B* 75, 45–57. doi: 10.1093/geronb/gby065

Lakens, D., Scheel, A. M., and Isager, P. M. (2018). Equivalence testing for psychological research: a tutorial. *Adv. Methods Pract. Psychol. Sci.* 1, 259–269. doi: 10.1177/2515245918770963

Limentani, G. B., Ringo, M. C., Ye, F., Bergquist, M. L., and McSorley, E. O. (2005). Beyond the t-test: statistical equivalence testing. *Anal. Chem.* 77, 221A–226A. doi: 10.1021/ac053390m

Ling, Y., Nefs, H. T., Morina, N., Heynderickx, I., and Brinkman, W.-P. (2014). A meta-analysis on the relationship between self-reported presence and anxiety in virtual reality exposure therapy for anxiety disorders. *PLoS ONE* 9:e96144. doi: 10.1371/journal.pone.0096144

Markus, H. (1978). The effect of mere presence on social facilitation: an unobtrusive test. *J. Exp. Soc. Psychol.* 14, 389–397. doi: 10.1016/0022-1031(78)90034-3

Murray, E. G., Neumann, D. L., Moffitt, R. L., and Thomas, P. R. (2016). The effects of the presence of others during a rowing exercise in a virtual reality environment. *Psychol. Sport Exerc.* 22, 328–336. doi: 10.1016/j.psychsport.2015.09.007

Pan, X., and Hamilton, A. F. d. C. (2015). Automatic imitation in a rich social context with virtual characters. *Front. Psychol.* 6:790. doi: 10.3389/fpsyg.2015.00790

Park, S., and Catrambone, R. (2007). Social facilitation effects of virtual humans. *Hum. Factors* 49, 1054–1060. doi: 10.1518/001872007X249910

Peirce, J. W. (2007). Psychopy-psychophysics software in python. *J. Neurosci. Methods* 162, 8–13. doi: 10.1016/j.jneumeth.2006.11.017

Peirce, J. W. (2009). Generating stimuli for neuroscience using psychopy. *Front. Neuroinformatics* 2:10. doi: 10.3389/neuro.11.010.2008

Platania, J., and Moran, G. P. (2001). Social facilitation as a function of the mere presence of others. *J. Soc. Psychol.* 141, 190–197. doi: 10.1080/00224540109600546

Poeschl, S. (2017). Virtual reality training for public speaking–a quest-vr framework validation. *Front. ICT* 4:13. doi: 10.3389/fict.2017.00013

Poeschl, S., and Doering, N. (2013). The german VR simulation realism scale-psychometric construction for virtual reality applications with virtual humans. *Annu. Rev. Cyberther. Telemed.* 11, 33–37. doi: 10.3233/978-1-61499-282-0-33

Poeschl, S., and Doering, N. (2014). Effects of simulation fidelity on user experience in virtual fear of public speaking training–an experimental study. *Annu. Rev. CyberTher. Telemed.* 12, 66–70. doi: 10.1109/VR.2014.6802073

Poeschl, S., and Doering, N. (2015). "Measuring co-presence and social presence in virtual environments–psychometric construction of a german scale for a fear of public speaking scenario," *Annual Review of Cybertherapy and Telemedicine 2015*, eds B. K. Wiederhold, G. Riva, and M. D. Wiederhold (San Diego, CA: IOS Press), 58–63.

Rafaeli, S., and Noy, A. (2002). Online auctions, messaging, communication and social facilitation: a simulation and experimental evidence. *Eur. J. Inform. Syst.* 11, 196–207. doi: 10.1057/palgrave.ejis.3000434

Riva, G., Waterworth, J., and Murray, D. (2014). *Interacting With Presence: HCI and the Sense of Presence in Computer-Mediated Environments*. Walter de Gruyter GmbH & Co KG.

Schubert, T., Friedmann, F., and Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence* 10, 266–281. doi: 10.1162/105474601300343603

Slater, M. (2003). A note on presence terminology. *Presence Connect* 3, 1–5.

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 364, 3549–3557. doi: 10.1098/rstb.2009.0138

Sterna, R., Strojny, P., and Rebilas, K. (2019). Can virtual observers affect our behavior? Social facilitation in virtual environments: mini-review. *Soc. Psychol. Bull.* 14:e30091. doi: 10.32872/spb.v14i3.30091

Strojny, P., Strojny, A., Kałwak, W., and Bańbura, A. (2018). Take your eyes off me. the effect of the presence of witnesses on the conduct of rescue operations. *Bezpieczeństwo i Technika Pożarnicza*, 49, 14–22. doi: 10.12845/bitp.49.1.2018.1

Triplett, N. (1898). The dynamogenic factors in pacemaking and competition. *Am. J. Psychol.* 9, 507–533. doi: 10.2307/1412188

Włodarczyk, D., and Wrześniewski, K. (2010). Kwestionariusz oceny stresu (kos). *Przeglad Psychol.* 53, 479–496.

Wojciszke, B., and Baryła, W. (2005). Skale do pomiaru nastroju i sześciu emocji. *Czasopismo Psychol. Pychol. J.* 11, 31–47.

Youngblut, C. (2003). *Experience of Presence in Virtual Environments*. Technical report, Institute for Defense Analyses, Alexandria, VA.

Zajonc, R. (1965). Social facilitation. *Science* 149, 269–274.

Zanbaka, C. A., Ulinski, A. C., Goolkasian, P., and Hodges, L. F. (2007). "Social responses to virtual humans: implications for future interface design," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, CA: ACM), 1561–1570.

Zieliński, P., and Biernacki, M. (2010). Analiza psychometryczna polskiego przekładu narzędzia do subiektywnej oceny obciążenia zadaniowego nasa-tlx. *Polski Przeglad Medycyny Lotniczej* 3, 219–239.

Conflict of Interest: The authors were employed by company Nano Games sp. z o.o.

frontiers in Psychology

ORIGINAL RESEARCH
published: 22 July 2020
doi: 10.3389/fpsyg.2020.01307

Motivational Profiling of League of Legends Players

Florian Brühlmann[1]*, Philipp Baumgartner[1], Günter Wallner[2,3,4], Simone Kriglstein[5,6] and Elisa D. Mekler[7]

[1] Human-Computer Interaction Research Group, Department of Psychology, Center for General Psychology and Methodology, University of Basel, Basel, Switzerland, [2] Department of Industrial Design, Eindhoven University of Technology, Eindhoven, Netherlands, [3] Faculty of Business and IT, Ontario Tech University, Oshawa, ON, Canada, [4] Institute of Art and Technology, University of Applied Arts Vienna, Vienna, Austria, [5] AIT Austrian Institute of Technology GmbH, Vienna, Austria, [6] Faculty of Computer Science, University of Vienna, Vienna, Austria, [7] Department of Computer Science, Aalto University, Espoo, Finland

Player motivation is a key research area within games research, with the aim of understanding how the motivation of players is related to their experience and behavior in the game. We present the results of a cross-sectional study with data from 750 players of *League of Legends*, a popular Multiplayer Online Battle Arena game. Based on the motivational regulations posited by Self-Determination Theory and Latent Profile Analysis, we identify four distinct motivational profiles, which differ with regards to player experience and, to a lesser extent, in-game behavior. While the more self-determined profiles "Intrinsic" and "Autonomous" report mainly positive experience-related outcomes, a considerable part of the player base does not. Players of the "Amotivated" and "External" profile derive less enjoyment, experience more negative affect and tension, and score lower on vitality, indicating game engagement that is potentially detrimental to players' well-being. With regards to game metrics, minor differences in the rate of assists in unranked matches and performance indicators were observed between profiles. This strengthens the notion that differences in experiences are not necessarily reflected in differences in behavioral game metrics. Our findings provide insights into the interplay of player motivation, experience, and in-game behavior, contributing to a more nuanced understanding of player-computer interaction.

Keywords: motivation, MOBA, game analytics, self-determination theory, latent profile analysis

OPEN ACCESS

Edited by:
Z. O. Toups,
New Mexico State University,
United States

Reviewed by:
Magy Seif El-Nasr,
Northeastern University, United States
Yubo Kou,
Pennsylvania State University (PSU),
United States

***Correspondence:**
Florian Brühlmann
florian.bruehlmann@unibas.ch

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Psychology

Received: *30 January 2020*
Accepted: *18 May 2020*
Published: *22 July 2020*

Citation:
Brühlmann F, Baumgartner P, Wallner G, Kriglstein S and Mekler ED (2020) Motivational Profiling of League of Legends Players. Front. Psychol. 11:1307. doi: 10.3389/fpsyg.2020.01307

1. INTRODUCTION

For many people, playing games is one of the most rewarding and motivating activities. In turn, people's motivation for playing games shapes their player experience and in-game behavior (e.g., Yee et al., 2012; Canossa et al., 2013; Schaekermann et al., 2017; Melhart et al., 2019), as well as their well-being (Przybylski et al., 2009; Vella et al., 2013; Perry et al., 2018). However, while concepts from motivational psychology, particularly Self-Determination Theory (SDT, Deci and Ryan, 2000), commonly inform research on player experience (Tyack and Mekler, 2020) and game analytics (e.g., Canossa et al., 2013; Melhart et al., 2019), the notion of motivational regulation (Deci and Ryan, 2000)has received limited attention in the context of games (Tyack and Mekler, 2020). This is an unfortunate gap in our understanding of the player-computer interaction, as motivational regulations have been found to determine to what extent people experience positive emotions and need satisfaction, as well as how persistently they engage in a behavior (Neys et al., 2014). Motivational regulations describe an underlying regulatory process of people's motivation, which determine the quality of their behavior, the extent of need satisfaction they experience, and the impact of these behaviors on their well-being (Deci and Ryan, 2000). Multiplayer Online Battle Arena (MOBA) games pose a particularly intriguing case. They enjoy enduring popularity, with a player base ranging in the millions, despite often affording a range of negative experiences (Johnson et al., 2015; Tyack et al., 2016). Specifically, MOBA players report decreased autonomy and increased frustration (Johnson et al., 2015), counter to SDT-based notions of positive player experience. Moreover, they afford complex, sometimes uncomfortable, social interactions amidst a highly competitive gaming environment. Considering players' underlying motivational regulations may hence provide a better understanding of the interplay of player experience and in-game behavior in MOBA games.

Identifying motivational profiles may enable us to study similarities between players and to highlight differences in experience, well-being, and behavior between these profiles. In that sense, this paper provides researchers and game designers with enhanced knowledge to better discern differing motivations and with it, experiences of their player basis. Building upon previous workon player profiling (e.g., Drachen et al., 2014; Chen et al., 2017; Nascimento Junior et al., 2017; Schaekermann et al., 2017), we present the results of a cross-sectional study with self-report and behavioral data from 750 players of *League of Legends* (*LoL*, Riot Games, 2009), a popular MOBA game. Drawing from work on SDT-based motivational profiling (Pastor et al., 2007; Howard et al., 2016; Wang et al., 2017; Gustafsson et al., 2018), we identify four distinct motivational player profiles (i.e., Amotivated, External, Intrinsic, and Autonomous) and compare these in terms of player experience and in-game behavior. We provide empirical evidence of the relation between motivational regulation and player experience. Specifically, we show that despite overall high intrinsic motivation, players can be categorized into distinct motivational profiles, which affect their quality of experience. Intrinsically and Autonomously motivated player profiles report consistently more positive player experiences, as evidenced by high scores on enjoyment, need satisfaction, and harmonious passion. In contrast, already slight increases in amotivation and external motivation were related to reduced enjoyment, more tension, and less harmonious passion, indicating game engagement that is potentially less conducive to players' well-being (Vella et al., 2013; Johnson et al., 2016). These findings extend our understanding of the role of motivation for the player-computer interaction, as well as provide context for conflicting results regarding the player experience of MOBA games (Johnson et al., 2015; Tyack et al., 2016). Second, we investigate how player motivation relates to in-game behavior, where we observe only a few clear-cut differences between motivational profiles. As such, our findings showcase that even when little to no behavioral differences are apparent, motivational regulations clearly color the quality of player experience.

Abbreviations: MOBA, Multiplayer Online Battle Arena; LoL, League of Legends; SDT, Self-Determination Theory; OIT, Organismic Integration Theory; EXT, external regulation; INT, introjected regulation; IDE, identified regulation; INT, integrated regulation; UMI, User Motivation Inventory; IMI, Intrinsic Motivation Inventory; PENS, Player Experience Need Satisfaction; PANAS, Positive and Negative Affect Schedule.

2. RELATED WORK

In the following section, we first review research around the interplay of player motivation, experience, and in-game behavior, after which we outline key motivational regulations posited by SDT and research on motivational profiling. Finally, we focus on the unique properties of MOBA games.

2.1. Player Motivation

Player motivation is a central research area in player-computer interaction, where the goal is to gain a better understanding of how motivational factors shape players' experience and behavior.

2.1.1. Motivation and Player Experience

Motivation is widely considered a key determinant of players' gaming experiences and preferences. Early works primarily linked motivation to typologies of player preferences and were not grounded in any established psychological frameworks or theories of human motivation. Bartle (1996), for instance, identified four distinct player "types" with varying play preferences in Multi-User Dungeon games. Similarly, Yee (2006) identified achievement, immersion, and social aspects of gameplay as key motivators for why people find playing online games appealing.

More recently, a growing body of player motivation research has emerged around Self-Determination Theory (SDT), a major psychological theory of human motivation (Deci and Ryan, 2000; Ryan and Deci, 2000). Notably, Ryan et al. (2006) criticized Yee's player motivation typology for focusing only on game content, rather than considering universal personal factors that generalize across a variety of players and game genres. Instead, they demonstrated in a series of studies that satisfaction of innate psychological needs for autonomy, competence, and relatedness, predict game enjoyment and future play across a variety of game genres. Indeed, this relation between psychological need satisfaction and positive player experience has been repeatedly demonstrated across several studies (e.g., Vella et al., 2013; Neys et al., 2014; Johnson et al., 2015, see also Tyack and Mekler, 2020 for a recent overview). Moreover, need satisfaction has also been linked to increased time spent playing (Johnson et al., 2016).

2.1.2. Motivation and In-Game Behavior

Digital games motivate a variety of goal-directed behaviors (Przybylski et al., 2010), which may be reflected in players' in-game behavior (Schaekermann et al., 2017). As such, a growing body of research has emerged around detecting player motivation profiles from game metrics. Specifically, game analytics provide detailed and granular insights into players' in-game behavior to identify hot spots or problem areas (e.g., Drachen and Canossa, 2009; Wallner et al., 2014). Bauckhage et al. (2012), for example, investigated behavioral telemetry data from five different games to understand how players engaged with these games over a

Type of Motivation	Amotivation	Extrinsic Motivation				Intrinsic Motivation
Type of Regulation	Non-regulation (AMO)	External Regulation (EXT)	Introjected Regulation (INJ)	Identified Regulation (IDE)	Integrated Regulation (INT)	Intrinsic Regulation (IMO)

FIGURE 1 | The six types of motivational regulation as posited by Self-Determination Theory. Ranging from the least self-determined (amotivation) to the most self-determined regulation (intrinsic motivation). Figure adapted from Deci and Ryan (2002), p. 16.

longer time period. Similarly, Harpstead et al. (2015) presented an approach for creating engagement profiles of game players. In the context of massively multiplayer online role-playing games, Feng et al. (2007) analyzed long-term player workloads and behavior in *EVE Online* (CCP, 2003). Suznjevic et al. (2011) identified categories of player actions in *World of Warcraft* (Blizzard Entertainment, 2004), which formed the basis for creating a player behavior model and combined it with network traffic models of the action categories.

However, while game analytics provide insight into players' in-game behavior, that is, *what* they are doing when playing, consideration of motivational frameworks may help contextualize *why* players behave in such a way (Hazan, 2013). Other works therefore attempted to link pre-defined motivational categories to in-game behavior. Yee et al. (2012), for instance, found that players' in-game behavior in *World of Warcraft* (Blizzard Entertainment, 2004) was to some extent predictive of their motivation (i.e., the aforementioned motives for immersion, achievement, and social interaction, Yee, 2006). Players motivated by achievement, for example, were more likely to engage in dungeoneering and Player vs. Player battles. In another study, Schaekermann et al. (2017) correlated self-reported player curiosity scores with in-game behavioral metrics in *Destiny* (Bungie, Inc., 2014), with curiosity considered a motivational driver for playing games. Among their results, they found that social curiosity was positively correlated to players' tendency toward exploratory behavior. Finally, some studies applied combined motivational psychology, data analysis, and machine learning techniques to better predict player engagement. Canossa et al. (2013), for example, investigated bivariate correlations and applied multiple supervised learning methods to identify relationships between in-game behavior in *Minecraft* (Mojang, 2011) and motivational factors, as measured by the Reiss Motivation Profiler (Reiss and Havercamp, 1998). Melhart et al. (2019), in contrast, employed support vector machines to predict motivation in *Tom Clancy's: The Division* (Massive Entertainment, 2016) based on game metrics. They found that both linear and non-linear models successfully predicted motivation with an average accuracy of 65.89 and 75.62% respectively. Notably, motivation was measured by the Ubisoft Perceived Experience Questionnaire (Azadvar and Canossa, 2018), a proxy for psychological need satisfaction in games, as posited by SDT (Ryan et al., 2006). However, correlations between the self-reported measures and game metrics remained weak.

2.2. Motivational Regulation

Organismic Integration Theory (OIT), a mini–theory of SDT, differentiates six types of motivational regulations (Deci and Ryan, 2000; Ryan and Deci, 2000). According to OIT, the underlying regulation of people's motivation determine the quality of their behavior, the extent of need satisfaction they experience and the consequences of these behaviors for their well-being (Deci and Ryan, 2000).

As depicted in **Figure 1**, these motivational regulations range on a spectrum from non-self-determined (amotivation) to fully self-determined (intrinsic motivation). Set in context, need satisfaction is an outcome of pursuing an activity (Deci and Ryan, 2000), while the degree to which an activity (e.g., playing a game) supports need satisfaction is determined by the underlying motivational regulation (e.g., *why* an activity is being pursued). Consequences (e.g., decreased need satisfaction) are more negative, the less self-determined the motivation for pursuing that activity is (Deci and Ryan, 2002). Specifically, OIT distinguishes three types of motivation: (1) Amotivation describes a lack or absence of motivation, hence being the least self-determined form of motivational regulation. (2) Extrinsic motivation refers to activity pursued for a separable outcome. More precisely, SDT distinguishes different types of extrinsic motivation comprised of four types of regulations: *external regulation* (EXT), *introjected regulation* (INJ), *identified regulation* (IDE), and *integrated regulation* (INT). EXT is the least self-determined form of extrinsic motivation and occurs in situations where people act to obtain a reward or avoid punishment (e.g., other players would pressure me if I perform badly at *League of Legends*). INJ regulation has been partially internalized, but not truly accepted as one's own. Such behaviors are pursued to avoid guilt or shame or to achieve feelings of self-worth or approval. IDE follows from the conscious valuing of an activity as personally important, rendering the pursuit of such an activity more self-determined. INT results when an activity is congruent with personally endorsed values and goals, and thus forms the most self-determined regulation among extrinsic motivations. Finally, (3) intrinsic motivation refers to an activity being pursued for its own sake, because it is experienced as enjoyable and interesting (Deci and Ryan, 2000).

2.2.1. Motivational Regulation in Human-Computer Interaction and Games

Motivational regulations, as posited by OIT, have also been explored within Human-Computer Interaction (HCI) and games research. In the context of general technology use, for instance, Brühlmann et al. (2018) developed and validated the User Motivation Inventory (UMI), an instrument that covers the whole spectrum of motivational regulation. Specifically, Brühlmann et al. (2018) found that respondents who reported higher levels of amotivation and scored lower on more self-determined regulations (IDE, INT) and intrinsic motivation, were more likely to consider to stop using a device. In contrast, participants scoring high on more self-determined and autonomous motivations reported more positive user experiences. Similarly, Peters et al. (2018) applied OIT to create a model that describes and predicts the impact of technologies on technology adoption, engagement and well-being. Hence, a better understanding of users' motivational regulations may help detect and prevent user churn, as well as identify potential negative effects of technology use on well-being.

The notions of need satisfaction and intrinsic motivation are also prevalent in player-computer interaction research (Tyack and Mekler, 2020). However, OIT has received relatively little attention (Tyack and Mekler, 2020). A few works have employed the Situational Motivation Scale (SIMS, Guay et al., 2000), but report no results (Alexandrovsky et al., 2019; Johanson et al., 2019). Birk and Mandryk (2018) used the SIMS to assess whether customization affected participants' motivation and behavior in a game-like self-improvement program taking place over 3 weeks. Curiously, they found that while customization resulted in significantly less attrition and more login counts, participants' self-reported motivation remained unaffected. Finally, Lafrenière et al. (2012) developed the Gaming Motivation Scale (GAMS), a questionnaire that assesses all six motivational regulations, specifically in the context of gaming.

Of particular interest to the present work, OIT has also been applied to study the player experience and gaming persistence of hardcore, heavy, and more casual players (Neys et al., 2014). Self-identified hardcore gamers reported the highest degree of intrinsic motivation and identified regulation, but also slightly elevated levels of external regulation, compared to heavy and casual gamers. Curiously, while also scoring high on intrinsic motivation and identified regulation, casual gamers scored highest on amotivation. With regards to playing persistence, immediate enjoyment was most predictive, but intrinsic motivation and external regulation were also significantly associated with increased persistence.

2.2.2. Motivational Regulation Profiles

More recently, works have drawn from OIT and attempted to profile people according to their motivational regulations. Gustafsson et al. (2018) explored the link between elite athletes' motivational profiles and burnout. Using Latent Profile Analysis (LPA), they identified five profiles with distinct patterns of motivational regulations. Athletes with high levels of amotivation as well as moderately controlled regulation showed higher burnout risk when compared to other profiles from the LPA. The quality of athletes' motivations might therefore be an important factor in protecting them from negative outcomes related to their health, performance and well-being. In the workplace setting, Howard et al. (2016) identified four motivational profiles of two samples of employees from different countries. They found that autonomous forms of motivation support positive workplace-related outcomes, such as performance and well-being. In another study, Wang et al. (2017) used LPA to identify four motivational profiles in secondary school students. Results showed that students in the highly self-determined motivational profile reported more effort, higher competence, value, and time spent on math beyond homework, when compared to the other profiles. In Pastor et al. (2007), LPA was used to classify college students into different goal orientation profiles using 2-, 3-, and 4-factor conceptualizations of goal orientation. The main goal was to show the advantages of LPA over other clustering methods. By using LPA, they were able to apply stricter criteria when deciding upon the final cluster solutions, represent students' cluster membership partially, and classify students from a different sample into clusters. This would not have been possible to the same extent with multiple regression or cluster analysis. Therefore, a person-centered approach to the study of motivational regulations seems promising.

2.3. MOBA Games

Multiplayer online battle arena (MOBA) games have been extremely popular throughout the years and are among the most profitable games on the market[1]. Thus, it comes to little surprise that a growing body of research has emerged around MOBA players' experience and behavior to better understand what keeps them motivated to play (see Mora-Cantallops and Sicilia, 2018, for a recent overview).

Johnson et al. (2015), for instance, found that compared to other genres, MOBA players report increased frustration and a reduced sense of autonomy. The authors hypothesize that this may be due to the intense competition with others. Relatedly, Kou et al. (2018) identified streakiness, i.e., whether players had winning or losing streaks—as crucial to player retention and experience of League of Legends, potentially because it impacts players' sense of competence Kou et al. (2018). Indeed, a common reason to quit playing MOBAs is that players simply do not experience them as fun anymore (Tyack et al., 2016).

Besides their competitive nature, MOBAs are also known for the complex social interactions they afford, with toxic player behavior among the major sources of negative experiences (Kwak and Blackburn, 2014; Kwak et al., 2015; Tyack et al., 2016). Tyack et al. (2016), for instance, identified deviant behavior from teammates as a reason to abandon playing MOBA games, although most players ultimately quit due to reasons unrelated to the game. In contrast, the opportunity to play with friends is a key motivator to start and keep playing MOBAs. However, despite this growing body of work around player churn and retention, none of the aforementioned studies have examined how players' experience relate to their in-game behavior.

With regards to players' in-game behavior, works have attempted to detect patterns in combat tactics of winning teams (Yang et al., 2014) based on the game data from *Dota 2* (Valve Corporation, 2013), analyzed professional and public

[1] https://www.statista.com/statistics/505613/leading-digital-pc-games-by-global-revenue/ (viewed: 28. Jan 2020).

matches for classifying playstyles (Gao et al., 2013), as well as classified player behavior in order to identify roles within player teams (Eggert et al., 2015). However, none of these works have considered players' motivation to engage with MOBAs. A notable exception is the work by Kahn et al. (2015), who developed a typology of player motives, similar to the work by Yee (2006), Yee et al. (2012). They validated their typology on a sample of over 18,000 League of Legends players and correlated the questionnaire with various game metrics. The motive to socialize was correlated with the average percentage of teammates that players already knew, whereas the completionist motivation was correlated with the number of different champions played. Finally, competitiveness was positively correlated with the number of kills and killing sprees. However, Kahn et al. (2015) did not explore how these motives relate to players' experience, nor is their typology grounded in any established framework of human motivation.

3. METHODS

The aim of this study was to explore how players' underlying motivational regulations relate to their experience and in-game behavior in a MOBA game. In contrast to previous research on predicting motivation from in-game metrics (Melhart et al., 2019), we present a novel, theory-driven approach for detecting motivational profiles, and compare these in terms of player experience and in-game behavior.

3.1. League of Legends

League of Legends (*LoL*) (Riot Games, 2009) is a MOBA game where players take on the role of *summoners* that control a single character (i.e., champion). Two teams of usually three or five players compete against each other. The two teams start on opposite sides of a map near a main building called *Nexus*. The goal of the game is to destroy the enemy's Nexus. The Nexus is defended by the enemy team, computer-controlled units (so-called "minions") and towers. The minions are sent in the direction of the enemy main building and follow certain paths (so-called "lanes") and attack close enemies. By killing minions, monsters, enemy champions, and destroying enemy towers, the player's own champion gains experience, i.e., they reach a higher level where new abilities can be unlocked or improved. These abilities are determined by the respective champion and are not freely selectable. In addition, the player who delivers the final deathblow to an enemy unit will receive a certain amount of gold. This gold can be used to purchase special items for the champion in the base, which improve various attributes (such as attack damage) or otherwise have positive effects. At the time of writing, there were a total of three maps with different game modes available. Among others, *LoL* offers the game modes "ranked" and "unranked" matches. Ranked matches are recorded in a central ranking system. Upon winning, players ascend in the rankings, and move down when they lose. Ranked games resemble unranked games but require a summoner level of 30 and a minimum of 20 champions to participate.

We chose to focus on *LoL*, because it is to date one of the most played games in the world[2]. Moreover, LoL is known to afford complex, sometimes negative social interactions (e.g., Kwak and Blackburn, 2014), and is among the most studied games in the MOBA research literature (Mora-Cantallops and Sicilia, 2018). Because of this complexity and the large player base, we expected that a variety of motivational regulations were present. Another advantage of *LoL* is the availability of a public Application Programming Interface (API), which allowed us to collect activity data to investigate player in-game behavior.

3.2. Participants

The survey was advertised on the *League of Legends* subreddit on the American social news aggregation website reddit.com. A total of 2,056 people started the survey, of which 877 completed the survey. Forty-four participants were excluded for not passing the instructed response item (*This is a verification Item. Please choose "Strongly disagree"*) (Brühlmann et al., 2020). We also conducted a longstring analysis to detect repeated answering schemes among the User Motivation Inventory (UMI) items (as in Brühlmann et al., 2018). However, no additional cases were flagged for exclusion through this procedure. Of the remaining 833 participants, 83 did not provide valid summoner names or showed incomplete data sets and were subsequently removed. After data cleaning, 750 participants were included in the analysis. Forty-five participants were women (6 percent), nine participants identified as non-binary and 12 preferred not to specify their gender. Participants' age ranged between 18 and 65 years (M = 21.5 years, SD = 4.05 years). In total, participants had played between seven and 5,012 matches (M = 1577.3 matches, SD = 860.5), with summoner levels ranging from 30 to 234 (M = 90.8, SD = 31.3).

3.3. Procedure

Upon clicking the survey link, participants were introduced to the study. After providing consent, participants were asked to provide basic demographic information (gender, age, experience with MOBAs, experience with playing *LoL*), their summoner name (i.e., the name the player is known in the game) and player region. The latter two were used to collect in–game data through the API made available by Riot Games (Riot Games, 2018). Participants then rated their motivation for playing *LoL* and answered a variety of player experience measures (see section 3.4). The individual measures were presented in a constant sequence, but the order of items was randomized for each measure. Finally, participants were given the option to comment on the survey and asked to indicate whether they had answered questions conscientiously. Participants did not receive any compensation for completing the survey, but were presented with a *LoL* "Player-Style" badge as a reward, similar to how previous work (Schaekermann et al., 2017) provided Brainhex (Nacke et al., 2014) badges upon survey completion. On average, the survey took 12 min to complete.

[2]100 million monthly active users in 2016 https://www.statista.com/statistics/317099/number-lol-registered-users-worldwide/ (viewed: 28. Jan 2020) and one of the free-to-play games that generated the most revenue in 2019 https://www.statista.com/statistics/346515/leading-f2p-mmo-games/ (viewed: 28. Jan 2020).

TABLE 1 | Means (M), standard deviations (SD), medians (Mdn), Cronbach's α, and hierarchical omega (ω) for all self-report measures over all participants ($N = 750$) and for each profile.

	M (SD)	Mdn	α	ω	Amotivated () n = 220	External () n = 329	Intrinsic () n = 90	Autonomous () n = 111	No. of items
UMI									**18**
IMO	5.31 (1.39)	5.67	0.86	0.88	5.11 (1.40)	4.97 (1.51)	6.34 (0.55)	5.92 (0.70)	3
INT	3.08 (1.48)	3.00	0.80	0.80	2.84 (1.47)	3.19 (1.49)	2.98 (1.59)	3.29 (1.28)	3
IDE	3.38 (1.39)	3.33	0.70	0.71	3.15 (1.40)	3.50 (1.42)	3.23 (1.44)	3.57 (1.16)	3
INJ	2.35 (1.55)	1.67	0.81	0.81	1.70 (0.80)	3.33 (1.75)	1.03 (0.10)	1.78 (0.64)	3
EXT	1.88 (1.22)	1.33	0.79	0.79	1.00 (0.00)	2.73 (1.37)	1.03 (0.09)	1.84 (0.47)	3
AMO	3.37 (1.91)	3.00	0.90	0.90	3.71 (1.81)	4.11 (1.86)	1.19 (0.33)	2.26 (1.02)	3
IMI									**12**
ENJ	5.23 (1.16)	5.43	0.86	0.87	5.01 (1.20)	4.99 (1.22)	6.07 (0.62)	5.72 (0.70)	7
TENS	3.65 (1.40)	3.60	0.81	0.82	3.48 (1.38)	4.10 (1.35)	2.81 (1.21)	3.31 (1.26)	5
PENS									**10**
REL	4.19 (1.63)	4.33	0.78	0.82	3.75 (1.60)	4.31 (1.69)	4.23 (1.58)	4.68 (1.38)	3
COM	5.05 (1.27)	5.00	0.79	0.80	4.98 (1.28)	4.91 (1.36)	5.37 (1.12)	5.34 (0.96)	3
AUT	4.96 (1.26)	5.00	0.75	0.76	4.72 (1.34)	4.78 (1.29)	5.67 (0.96)	5.38 (0.85)	4
ACH_GOAL									**11**
PerfAp	5.24 (1.58)	5.67	0.86	0.86	5.14 (1.66)	5.45 (1.52)	4.87 (1.61)	5.12 (1.47)	3
PerfAv [3]	4.24 (1.71)	4.25	0.65	0.65	4.05 (1.79)	4.68 (1.62)	3.39 (1.62)	3.99 (1.54)	2
MastAp	4.86 (1.61)	5.00	0.82	0.82	4.65 (1.77)	4.98 (1.56)	4.85 (1.53)	4.91 (1.43)	3
MastAv	3.70 (1.78)	3.67	0.85	0.86	3.50 (1.78)	4.19 (1.77)	2.76 (1.59)	3.37 (1.5)	3
Passion									**10**
HP	4.06 (1.34)	4.20	0.79	0.79	3.83 (1.36)	3.98 (1.39)	4.38 (1.33)	4.46 (1.03)	5
OP	2.47 (1.42)	2.20	0.87	0.87	2.32 (1.44)	2.92 (1.48)	1.50 (0.75)	2.21 (1.04)	5
PANAS									**20**
PA	35.68 (7.17)	36	0.84	0.84	34.58 (7.69)	35.43 (7.13)	37.50 (6.85)	37.16 (5.96)	10
NA	22.14 (7.27)	21	0.81	0.81	21.51 (6.31)	24.98 (7.50)	16.38 (5.44)	19.68 (5.48)	10
VITALITY	3.59 (1.16)	3.57	0.78	0.89	3.37 (1.19)	3.55 (1.15)	3.87 (1.28)	3.92 (0.91)	**7**

3.4. Measures

We collected subjective self-report measures and behavioral game metrics. All self-report measures consisted of 7-point Likert scales ranging from strongly disagree (1) to strongly agree (7), unless noted otherwise. Descriptive statistics and reliability scores (Cronbach's α and hierarchical ω) for each measure are depicted in **Table 1**.

3.4.1. User Motivation Inventory (UMI)

To measure the six motivational regulations outlined in section subsection 2.2, we employed the User Motivation Inventory (UMI, Brühlmann et al., 2018). The UMI is a validated 18-item questionnaire, which distinguishes amotivation, external regulation, introjected regulation, identified regulation, integrated regulation, and intrinsic motivation in the context of technology use. While all based on SDT, we chose the UMI over the SIMS (Guay et al., 2000) and ACTA (Peters et al., 2018), as they do not assess introjected and integrated regulation or amotivation, respectively. We also considered the UMI more suitable than the GAMS (Lafrenière et al., 2012). While it specifically measures motivational regulations in the context of gaming, it does not account for social aspects of (external) motivational regulation (Lafrenière et al., 2012), which we expected to be particularly pertinent to the experience of playing *LoL* with others (Tyack et al., 2016; Mora-Cantallops and Sicilia, 2018).

3.4.2. Player Experience Need Satisfaction (PENS)

Psychological need satisfaction is a core concept in SDT (Deci and Ryan, 2000; Ryan and Deci, 2000), and motivational regulation is known to shape the extent to which experiences satisfy people's psychological needs of autonomy, competence, and relatedness. Need satisfaction is also prevalent in player-computer interaction research (Tyack and Mekler, 2020), where it has been consistently linked to positive player experience across a variety of genres (Ryan et al., 2006; Johnson et al., 2015, 2016) and playing persistence (Neys et al., 2014). However, with regards to MOBA games, players have reported less autonomy satisfaction, as well as increased frustration (Johnson et al., 2015), hinting at a possible relation to competence. For these reasons, we included the Player Experience Need Satisfaction scale (PENS, Ryan et al., 2006) to assess players' perceptions of autonomy, competence, and relatedness when playing *LoL*.

3.4.3. Interest-Enjoyment and Pressure-Tension (IMI)

Intrinsically motivated behavior is characterized by the experience of interest and enjoyment (Deci and Ryan, 2000;

Ryan and Deci, 2000). Hence, we employed the dimension interest-enjoyment of the Intrinsic Motivation Inventory (IMI, Ryan et al., 1983; McAuley et al., 1989) to assess self-reported intrinsic motivation. The IMI is commonly employed in player-computer interaction as a proxy for game enjoyment and positive player experience (Tyack and Mekler, 2020). We also included the pressure-tension dimension of the IMI, because it is a negative predictor of intrinsic motivation (Deci and Ryan, 2000; Ryan and Deci, 2000), and because it commonly characterizes the experiences of MOBA players (Johnson et al., 2015; Tyack et al., 2016).

3.4.4. Positive and Negative Affect (PANAS)

Players of MOBA games, such as *LoL*, often experience pronounced positive and negative affect (Johnson et al., 2015; Tyack et al., 2016). Hence, we employed the PANAS by Watson et al. (1988) to assess positive affect (PA) and negative affect (NA). Items were rated on a 5-point Likert-type scale.

3.4.5. Vitality

Mora-Cantallops and Sicilia (2018) called for more research into the impact of MOBA play on player well-being. Hence, we measured vitality, an established well-being index in SDT-based research (Ryan and Frederick, 1997). Specifically, people's experience of vitality varies as a function of both contextual and psychological factors, for instance, to the degree that one is unburdened by external pressures. We employed the vitality scale developed by Ryan and Frederick (1997). Item wording was adapted to fit the survey context, for instance, "When I play *LoL* I feel alive and vital."

3.4.6. Harmonious and Obsessive Passion

As we decided to advertise the survey on the *League of Legends* subreddit, we expected that most participants would be very passionate players of the game. However, passion to play can be harmonious or obsessive (Przybylski et al., 2009; Puerta-Cortés et al., 2017; Schaekermann et al., 2017; Perry et al., 2018). Hence, we included measures of harmonious and obsessive passion (Vallerand et al., 2003). Specifically, *harmonious passion* describes the autonomous and self-determined internalization of an activity into one's identity (Vallerand et al., 2003), whereby the activity is aligned with different areas of a person's life (i.e., they have freely chosen to play *LoL* and the activity "harmonizes" with other areas of their life, and does not interfere with their work or social life). In contrast, *obsessive passion* refers to non-self-determined internalization of an activity due to external or internal pressure (i.e., the person feels compelled to play *LoL*, for example, because of other players or personal dependencies; Vallerand et al., 2003). As such, harmonious and obsessive passion are closely linked to motivational regulation and have also been found to impact the amount of play, game enjoyment, and tension following play (Przybylski et al., 2009).

We employed an adapted version of the Harmonious and Obsessive Passion for Gambling scale (Vallerand et al., 2003; Przybylski et al., 2009). To match the context of the study, items were re-worded by replacing "this activity" with "*LoL*".

3.4.7. Achievement Goals

The gameplay of *LoL* is performative and often highly competitive in nature. Therefore, we measured players' achievement goals orientation. While not per se based on SDT, achievement goals orientation refers to how people approach competence-relevant behavior, such as studying or training (Elliot and McGregor, 2001), where different achievement goals have been found to impact intrinsic motivation to varying degrees (Chen et al., 2019). Specifically, Elliot and McGregor (2001) distinguish four related, albeit distinct achievement goals. *Mastery approach* goal orientation refers to a focus on mastering an activity and developing skills, whereas *mastery avoidance* focuses on not losing previously acquired knowledge or skills. Mastery approach, in particular, has been linked to intrinsic motivation and is associated with a wide range of positive effects in educational settings (Elliot and McGregor, 2001). In contrast, people oriented toward *performance avoidance*[3] strive not to underperform relative to normative standards or peers, while *performance approach* is oriented toward performing better than peers or externally imposed standards. Such a performance orientation has been linked to extrinsic motivation and reduced intrinsic motivation. To measure these four orientations, we employed the achievement goal questionnaire developed by Elliot and McGregor (2001).

3.4.8. Behavioral Game Metrics

Using the summoner name and region provided by participants, match histories and behavioral in-game data up until August, 16, 2018 were obtained from the API using *Riot-Watcher* (Przybylski et al., 2018)—a Python wrapper for the Riot Games API. For some matches, detailed data was not available or was incomplete. These matches were excluded from subsequent processing. We chose to focus on more recent matches played during Season 7, as well as—at the time of data sampling—ongoing Season 8 (including its preseason), i.e., matches played between January 30, 2017 and August 16, 2018. This procedure resulted in a total of 1,179,828 matches. During this period, three game maps with fundamentally different types of gameplay, strategy, match length, and team size were available (Summoner's Rift, The Twisted Treeline, and Howling Abyss). To exclude possible variability in the data due to these differences, the analysis was focused on the most popular game map, Summoner's Rift (973,564 [82.5%] of all matches). Two participants had to be excluded from the analysis because no data was available for this map.

In-game metrics for individual players derived from these matches were aggregated separately for ranked and unranked matches and, when appropriate, normalized to account for different numbers of matches.

Measures that were considered relevant for ranked and unranked matches separately include *time played*, *win rate*, *deaths*, *kills*, *assists* (i.e., helping an ally to kill an opponent), *kda* (describing the ratio of kills, deaths and assists), *killing sprees* (requiring a player to kill a certain amount of enemies

[3]*Note.* Due to an error in the survey, the item "I just want to avoid doing poorly in LoL" had to be excluded from analysis.

TABLE 2 | Description of in-game measures.

Feature	Description
COMBINED	
totalMatches	Total number of matches (ranked and unranked)
level	The level of the summoner level
RANKED AND UNRANKED	
timePlayed	Total time spent in matches [in hours]
winrate	Won matches/total matches [in %]
kda	($\sum$ kills + $\sum$ deaths)/ $\sum$ assists
deaths	Avg. number of deaths per match
kills	Avg. number of kills per match
assists	Avg. number of assists per match
killingSprees	Avg. number of killing sprees per match
totalDamageDealt	Avg. total damage dealt per match
totalHeal	Avg. total heal per match
goldEarned	Avg. gold earned per match
goldSpent	Avg. gold spent per match
championsPlayed	Number of different champions played

without dying), *total damage dealt*, *total heal* (restoring one's own or an ally's health), *gold earned* (gold as in–game currency can be earned either passively (i.e., automatically without player interaction) or by actively performing certain actions, such as killing units), *gold spent* (gold can be spent on items which provide further benefits to the player) and *champions played* (the number of different champions played). Moreover, players' *level* (as a measure of experience) and *total number of matches played* represent aggregated measures over ranked and unranked matches. In total, these measures account for broad information on time, performance, and economy related in-game behavior. Note that the level of a summoner is roughly indicative of how much time a player spent playing a game and determines whether they can access some features of the game. Most prominently, a summoner level of 30 or higher is required to play ranked games. The maximum summoner's level cap was changed in the end of 2017 from 30 to limitless. The constraint of level 30 to play ranked games remained unchanged. See **Table 2** for a description and **Table 3** for descriptive statistics of each metric.

4. RESULTS

The results are structured as follows: First, we report correlations between self-report player experience measures and in-game metrics. Second, we test the measurement model of the UMI and use the factor scores to identify distinct motivational profiles. Third, the different motivational profiles are compared in terms of player experience and in-game behavior. Descriptive statistics for all self-report measures are presented in **Table 1** and for all behavioral metrics in **Table 3**.

4.1. Correlation Analysis

To assess to what extent motivational regulation was related to participants' in-game behavior, we calculated a series of Pearson correlations. Overall, several significant correlations emerged between the different motivational regulations and in-game behavior, ranging from small to moderate. For the sake of brevity, only significant correlations with $r \geq |0.1|$ (Pearson correlation, bootstrapped p-values with 1,000 iterations) are reported here. Individual p-values and the complete correlation matrix are included as **Supplementary Material**.

Amotivation correlated negatively with *assists* in unranked ($r = -0.13$) and in ranked matches ($r = -0.11$) and positively with *goldSpent* in ranked matches ($r = 0.10$). Put differently, more amotivated players were less likely to assist other players in kills but spent more gold in ranked matches.

External regulation was only correlated positively with *totalHeal* unranked ($r = 0.13$). Introjected regulation, however, correlated positively with *totalMatches* ($r = 0.10$), *level* ($r = 0.11$), *timePlayed* ranked ($r = 0.11$), *winrate* ranked ($r = 0.10$), and *championsPlayed* ranked ($r = 0.10$). This suggests that players were more motivated to avoid feelings of guilt or failure, spent more time playing *LoL*, especially ranked matches. Moreover, introjected regulation was also positively correlated with *killingSprees* ($r = 0.11$), as well as ranked ($r = 0.11$) and unranked *kills* ($r = 0.11$).

For identified regulation, only two noteworthy correlations were observed: Players who considered playing *LoL* important, had played more *totalMatches* ($r = 0.11$) and achieved a higher *level* ($r = 0.15$). Similar correlational patterns emerged for integrated regulation ($r = 0.11$ and $r = 0.14$, respectively). Additionally, integrated regulation correlated positively with *timePlayed* ranked ($r = 0.14$) and *championsPlayed* ranked ($r = 0.13$).

Finally, intrinsic motivation correlated positively with achieved *level* ($r = 0.12$) and *assists* in ranked matches ($r = 0.11$). Intrinsic motivation was also negatively correlated with *kills* unranked ($r = -0.10$), *killingSprees* unranked ($r = -0.11$), *totalDamage* ranked ($r = -0.10$) *goldEarned* unranked ($r = -0.11$), *goldSpent* unranked ($r = -0.12$), *goldEarned* ranked ($r = -0.11$), and *goldSpent* unranked ($r = -0.10$). This suggests that intrinsically motivated players scored fewer kills in unranked matches, dealt less damage in ranked matches, as well as earned and spent less gold overall.

Note that correlation analysis offers only variable-centered insights into relationships between particular motivational regulations and individual metrics. Recall that SDT instead conceptualizes motivation as a multi–dimensional construct, spanning a continuum of self-determination (Deci and Ryan, 2000). Hence, it is more insightful to study how combinations of motivation variables relate to experiential and behavioral variables, rather than individual (cor)relations. Moreover, our goal was to go beyond variable-centered approaches and apply a person-centered method to identify qualitatively different motivational profiles of *LoL* players.

4.2. Motivational Profile Analysis

4.2.1. Confirmatory Factor Analysis (CFA)

To test the measurement model of the UMI, a six-factor confirmatory factor analysis (CFA) was conducted. All items were specified to load on their designated factor, and the loading of the first item was constrained to one. Multivariate normality was not

TABLE 3 | Means (M), standard deviations (SD), medians (Mdn) for behavioral metrics from the game map *Summoner's Rift* over all participants and for each profile (N = 748).

	M (SD)	Mdn	Amotivated (■) *n* = 220	External (■) *n* = 329	Intrinsic (■) *n* = 89	Autonomous (■) *n* = 110
COMBINED						
totalMatches	1577.31 (860.47)	1455.50	1509.59 (808.3)	1596.26 (908.54)	1560.15 (828.52)	1669.96 (839.12)
level	90.75 (31.34)	88	86.31 (29.41)	92.62 (33.46)	90.38 (30.44)	94.33 (28.54)
RANKED						
timePlayed	292.12 (291.92)	217.50	315.39 (293.98)	283.11 (302.02)	288.79 (289.26)	275.25 (258.32)
winrate	51.32 (7.93)	51.29	51.04 (6.56)	51.54 (8.49)	51.07 (6.2)	51.45 (9.79)
kda	2.78 (0.9)	2.70	2.68 (0.62)	2.74 (0.58)	2.82 (0.63)	3.04 (1.81)
deaths	5.28 (0.99)	5.22	5.38 (0.98)	5.3 (0.98)	5.2 (1.02)	5.09 (1)
kills	5.26 (1.88)	5.53	5.31 (1.77)	5.36 (1.91)	4.99 (1.85)	5.12 (2.03)
assists	8.83 (2.37)	8.39	8.6 (2.21)	8.78 (2.42)	9.16 (2.45)	9.18 (2.39)
killingSprees	1.15 (0.46)	1.22	1.17 (0.43)	1.16 (0.46)	1.09 (0.44)	1.11 (0.5)
totalDamageDealt	108736.02 (37386.71)	118687.60	110965.31 (36541.62)	110363.24 (36371.49)	104970.8 (40501.7)	102456.99 (39010.61)
totalHeal	5577.06 (2313.61)	5133.46	5602.27 (2212.22)	5622.33 (2390.04)	5418.86 (2293.13)	5519.25 (2320.58)
goldEarned	11079.72 (1199.34)	11276.85	11136.71 (1210.19)	11137.29 (1176.31)	10970.12 (1232.92)	10882.27 (1207.63)
goldSpent	10043.92 (1167.41)	10237.11	10115.51 (1175.32)	10099.83 (1141.06)	9917.16 (1194.8)	9836.08 (1190.54)
championsPlayed	45.82 (31.52)	39.50	49.82 (34.96)	44.18 (28.46)	44.2 (30.7)	44.02 (33.29)
UNRANKED						
timePlayed	332.4 (281.74)	253.50	298.31 (281.84)	349.28 (283.64)	343.71 (281.26)	340.97 (273.95)
winrate	54.39 (7.32)	52.96	55.23 (9.37)	54.14 (6.31)	52.68 (4.97)	54.82 (6.87)
kda	2.69 (1.2)	2.47	2.76 (1.84)	2.65 (0.77)	2.55 (0.55)	2.78 (1.01)
deaths	6.14 (1.39)	6.07	6.21 (1.55)	6.14 (1.33)	6.17 (1.19)	5.98 (1.36)
kills	7.26 (2.24)	7.11	7.44 (2.3)	7.33 (2.25)	6.79 (2.19)	7.04 (2.07)
assists	8.25 (1.55)	8.13	8.04 (1.58)	8.19 (1.49)	8.47 (1.53)	8.64 (1.61)
killingSprees	1.54 (0.49)	1.52	1.58 (0.5)	1.55 (0.48)	1.47 (0.49)	1.51 (0.48)
totalDamageDealt	111332.13 (26068.91)	113456.90	111426.6 (26398.6)	112084.53 (24939.86)	110957.54 (29379.94)	109195.91 (26118.44)
totalHeal	5056.64 (1262.85)	4963.70	4914.68 (1225.82)	5123.63 (1234.24)	5023.89 (1276.16)	5166.66 (1394.35)
goldEarned	12075.91 (1396.49)	12053.52	12165.74 (1490.8)	12028.94 (1375.63)	12065.53 (1419.59)	12045.12 (1246.69)
goldSpent	10944.39 (1373.6)	10917.23	11042.02 (1486.24)	10895.64 (1350.68)	10912.57 (1319.88)	10920.7 (1252.36)
championsPlayed	85.65 (33.3)	88.50	84.15 (34.28)	86.36 (33.46)	84.88 (32.6)	87.13 (31.7)

given (Mardia tests: $\chi^2 = 4644.83$, $p < 0.001$, $Z_k = 52.98$, $p < 0.001$), hence a robust Maximum Likelihood Estimation method with Huber-White standard errors and a Yuan-Bentler based scaled test statistic was used[4]. Results of the CFA suggested that the six factor model fits the data well [$\chi^2 = 257.21$, $p < 0.001$, $\chi^2/df = 2.14$, $CFI = 0.972$, SRMR = 0.050, RMSEA = 0.039, PCLOSE = 0.999].

4.2.2. Latent Profile Analysis (LPA)

Latent Profile Analysis (LPA) is a latent variable modeling technique that detects clusters of observations with similar values on cluster indicators (Pastor et al., 2007). In other words, it can be used to identify combinations of motivation variables, which can then be related to other variables, such as player experience and in-game behavior, while circumventing the aforementioned issues around correlation analysis. Although a relatively novel technique, it has previously been applied in SDT research to study motivation in educational (Pastor et al., 2007; Wang et al., 2017), work (Howard et al., 2016) and athletic settings (Gustafsson et al., 2018).

To assess whether the data exhibited distinct motivational profiles, we conducted an LPA using factor scores retained from the CFA six factor model. Conducting an LPA with factor scores instead of scale scores allows for partial control of measurement errors by giving more weight to items (Howard et al., 2016; Kam et al., 2016). When determining the optimal number of profiles, it is key to consider not only the statistical adequacy of the found solution, but also the theoretical conformity of the profiles (Morin and Marsh, 2015; Howard et al., 2016). In deciding upon our final model, information-based methods like the Bayesian Information Criterion (BIC) and Integrated Complete-data Likelihood (ICL), as well as resampling methods, such as the Bootstrap Likelihood Ratio Test (BLRT), were considered for each solution (Scrucca et al., 2016). Other indices, such as entropy, AIC, LMR, ALMR are not recommended for selecting

[4] *Note.* We also conducted an Exploratory Factor Analysis (EFA) and parallel analysis, which proposed a five-factor model instead. However, a subsequent CFA indicated that the five-factor model had a significantly worse fit (χ^2 diff. = 84.96, $p < 0.001$).

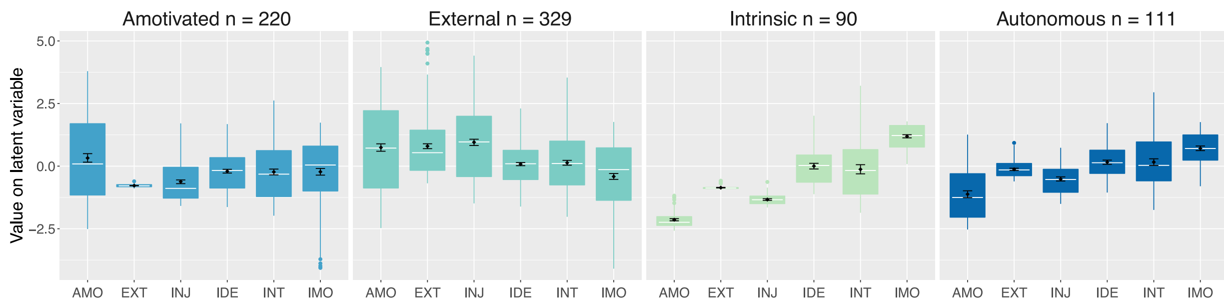


FIGURE 2 | Motivational pattern of the four profiles identified in the sample. The white lines in the boxplot indicate the median and the black rhombi indicate the mean with bootstraped 95% confidence intervals (1,000 iterations).

the optimal number of profiles (Tofighi and Enders, 2008; Diallo et al., 2017).

The estimated fit indices proposed a divergent optimal number of profiles. The BIC, ICL, and investigation of the Elbow plots indicated that four profiles were most appropriate and parsimonious (BIC (VVV), five groups: −10652.0, ICL (VVV), four groups: −10771.4). Visual interpretation of the elbow plot for the BIC criterion also revealed four groups to be most appropriate. In contrast, the BLRT found the optimal group size to be seven, reflecting the data (Likelihood Ratio Test 7 vs. 8 groups: −165.92, $p = 0.996$). After considering the theoretical conformity of the profiles (i.e., resulting group sizes, group specific motivational profiles), we deemed four profiles to be optimal.

Figure 2 shows the distribution of scores for all six motivational regulations for each of the four profiles, where 0.0 depicts the overall mean score for each latent variable (i.e., M = 5.31 for intrinsic motivation; M = 3.08 for integrated regulation, etc.). As listed in **Table 1**, participants overall reported high levels of intrinsic motivation ($M = 5.31$, $SD = 1.39$) and low scores on the remaining regulations, especially introjected ($M = 2.35$, SD = 1.55) and external regulation ($M = 1.88$, $SD = 1.22$).

- Profile 1 ($n = 220$) was characterized by above average amotivation. Compared to other players, participants in this profile also reported below average intrinsic motivation and external regulation, while the other motivational regulations scored close to 0.0 (i.e., average). This does not mean that this player profile lacked in intrinsic motivation. In fact, players in this profile reported considerable intrinsic motivation (M = 5.11, see **Table 1**). However, participants' rather elevated amotivation ratings ($M = 3.71$, **Table 1**) were what primarily differentiated Profile 1 from the other profiles. Based on the motivational spectrum posited by SDT (see **Figure 1**), we hence refer to Profile 1 as "Amotivated."
- Profile 2 ($n = 329$) featured markedly above average scores on amotivation, external and introjected regulation, as well as slightly above average scores on identified regulation and integrated regulation. While still considerable ($M = 4.97$), intrinsic motivation scores were below average, compared to the overall sample. Similar to the "Amotivated" profile, players in this profile reported considerable amotivation ($M = 4.11$). However, what distinguishes Profile 2 from the other profiles, are the comparably higher scores on external and introjected regulation ($M = 2.73$ and $M = 3.33$, respectively). Hence, we dubbed this the "External" profile.
- Profile 3 ($n = 90$) scored above average on intrinsic motivation, whereas the other motivational regulations were at average or below average levels. In other words, players in this profile were predominantly intrinsically motivated, and accordingly scored high on intrinsic motivation ($M = 6.34$). Hence, we refer to this as the "Intrinsic" profile.
- Profile 4 ($n = 111$) scored above average on intrinsic motivation ($M = 5.92$), but less so than the "Intrinsic" profile. Moreover, it featured slightly above average levels on identified and integrated regulation, as well as average levels of external regulation. In contrast to the "Intrinsic" profile, players in this profile were most characterized by a blend of intrinsic motivation and slightly higher scores on the other motivational regulations. Nevertheless, as the "autonomous" regulations (i.e., intrinsic motivation, identified and integrated regulation, Deci and Ryan, 2000; Ryan and Deci, 2000) were more salient, we refer to this as the "Autonomous" profile.

4.3. Player Experience

Kruskal-Wallis rank sum tests were conducted to test whether the four motivational profiles differed significantly with regards to the self-report player experience measures. Statistically significant differences were found for every measure at an alpha-level of .001. However, due to the exploratory nature of this study and the large number of variables, the results are interpreted based on descriptive statistics (means, medians, and distributions). Note also that statistical significance testing between each pair of profiles for all measures would greatly increase the likelihood of type 1 errors (i.e., false positives). Therefore, **Figures 2**, **3** include a bootstrapped (1,000 iterations) 95% confidence interval of the mean. If the proportion of overlap of 95% confidence intervals of two means is 0.5 or less, they indicate statistical significance at an alpha-level of 5% (Cumming and Finch, 2005).

As pictured in **Figure 3** (see also **Table 1**), all profiles reported high enjoyment, especially the Intrinsic player profile ($M = 6.07$,

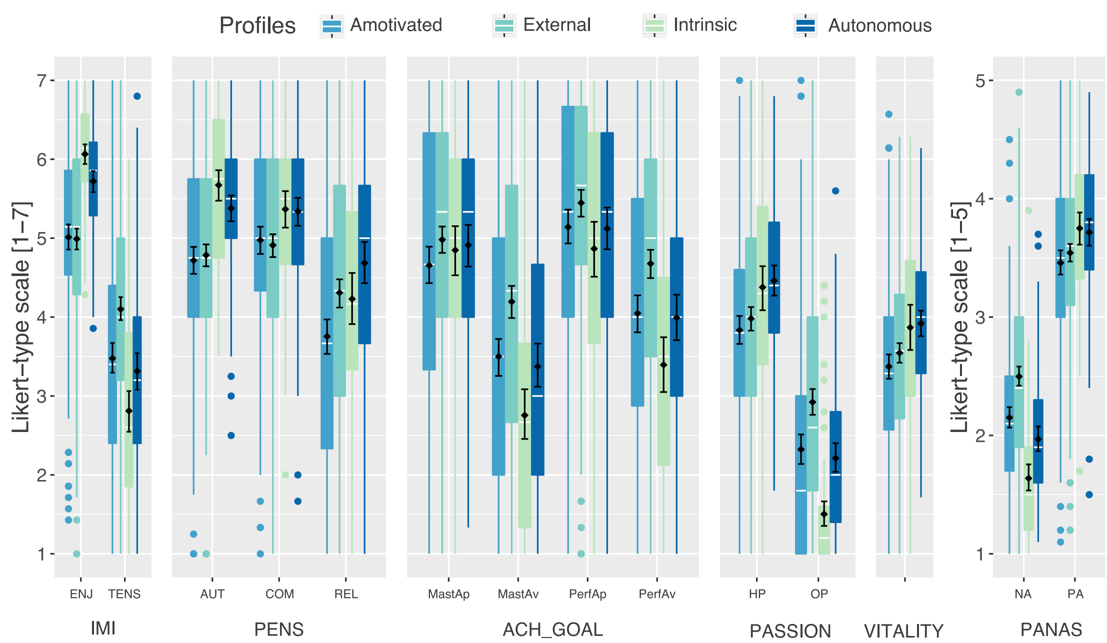


FIGURE 3 | Distribution of the values on the different player experience measures. The white lines in the box plot indicate the median and the black rhombi indicate the mean with bootstraped 95% confidence intervals (1,000 iterations).

SD = 0.62). In contrast, the External profile scored highest on tension. Moreover, all motivational profiles scored relatively high on relatedness, autonomy, and competence need satisfaction, with relatedness being least salient. However, the Intrinsic and Autonomous player profiles reported the highest levels of need satisfaction for all three needs, where the latter scored highest on relatedness.

With regards to achievement goals, participants overall scored highest on performance approach, followed by mastery approach and performance avoidance. Looking at the individual profiles, the External player profile reported the highest levels of performance approach and avoidance, as well as mastery avoidance. In contrast, the Intrinsic profile scored lowest on avoidance for both performance and mastery. Mastery approach was comparable between profiles, but lowest for Amotivated players.

In general, participants scored low on obsessive passion and around midpoint ($M = 4.06$) on harmonious passion. The Autonomous and Intrinsic player profiles reported the highest levels of harmonious passion, with the Intrinsic profile scoring particularly low on obsessive passion. In contrast, External players reported markedly higher levels of obsessive passion compared to the other profiles.

Overall, vitality after playing *LoL* was slightly below midpoint ($M = 3.59$, $SD = 1.16$), where the Autonomous and Intrinsic profiles experienced more vitality than the Amotivated and External players.

Finally, with regards to affect, the Amotivated and especially the External profiles reported markedly increased levels of negative affect compared to the other profiles. Positive affect was rather pronounced for all profiles, but more so for the Autonomous and Intrinsic player profiles.

4.4. Behavioral Game Metrics

An overview of all behavioral metrics is presented in **Table 3**, and **Figure 4** includes confidence intervals for the means. Overall, participants had played almost 1,600 matches on average between January 30, 2017, and August 16, 2018. More time was spent playing unranked than ranked matches. In the following, each metric will be compared between the four profiles. A series of Kruskal-Wallis rank sum tests was conducted to test whether there were overall significant differences in the behavioral data. Results showed that *winrate* unranked, $\chi^2(3) = 9.68$ $p < 0.05$, *kda* ranked, $\chi^2(3) = 10.9$ $p < 0.05$, and *assists* unranked, $\chi^2(3) = 14.64$ $p < 0.05$, showed significant differences between profiles.

4.4.1. Number of Matches, Level, and Playtime

For the total amount of matches and the average level of the players, a slight increase from the Amotivated toward the Autonomous player profile is visible. Amotivated players spent the most time playing ranked matches and the least amount of time in unranked matches. These players seem to be more ranked games oriented. However, they were on average on a lower in-game level, whereas the Autonomous profile featured more higher-level players.

4.4.2. Performance Measures

With players being keen on improving their performance, as shown by the high scores on performance approach orientation,

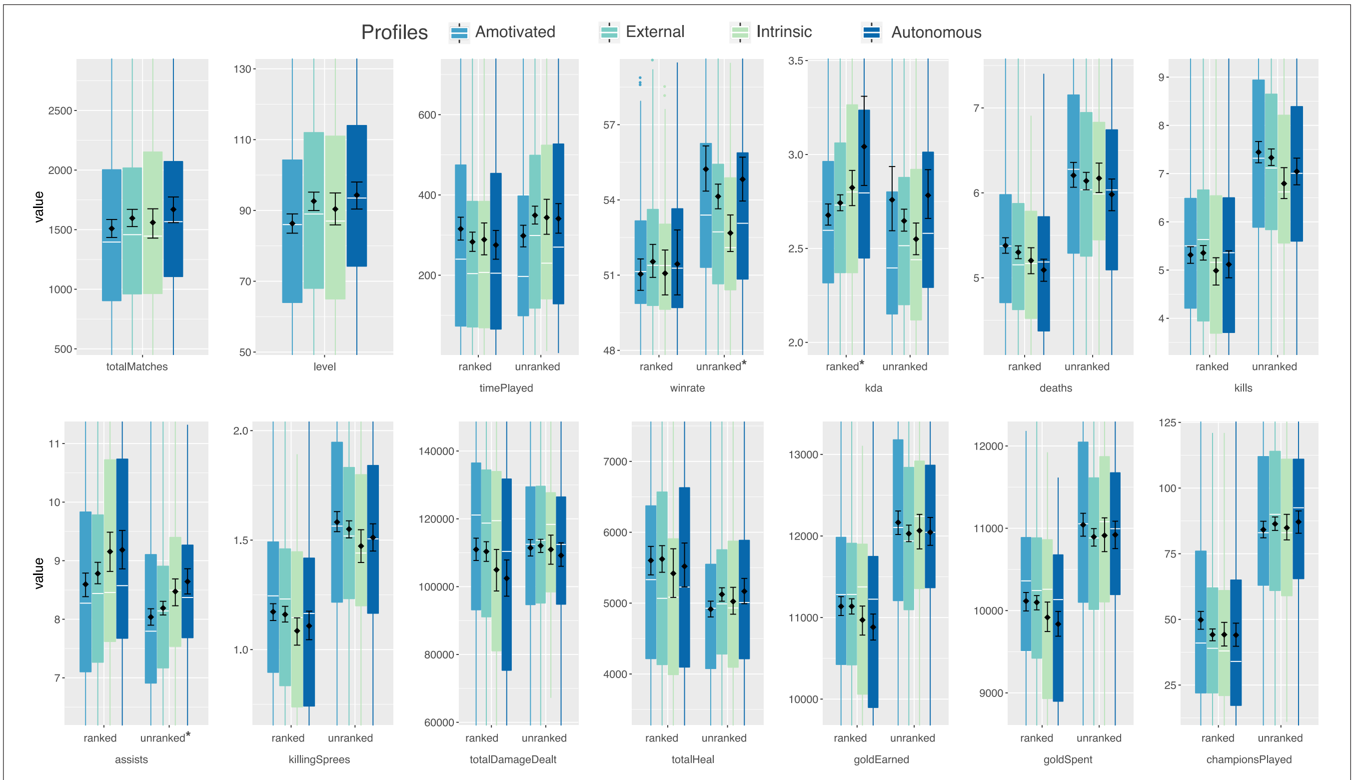


FIGURE 4 | Comparison of behavioral metrics between the four profiles. The white lines in the box plot indicate the median and the black rhombi indicate the mean with bootstraped 95% confidence intervals (1,000 iterations). Asterisks highlight statistically significant differences with Kruskal-Wallis rank sum tests and $\alpha = 0.05$.

we were interested in exploring the relations between wins and losses, as well as kills, deaths, and assists.

For unranked matches, Amotivated players showed a significantly higher *winrate* than the Intrinsic player profile ($Z = 2.923$, $p < 0.05$, Dunn's multiple comparison with *p*-values adjusted with the Holm method), while the Autonomous and External profiles are in-between. In ranked matches, a comparison of the *winrate* reveals very similar means for all profiles, slightly above 50% each, confirming the effectiveness of the *LoL* match-making mechanism.

However, in terms of the number of deaths in ranked matches, the more self-determined profiles "Intrinsic" and "Autonomous" show lower values, but they also score less kills in both ranked and unranked matches. Intrinsic and Autonomous players scored more assists in ranked and unranked matches. For unranked matches, *post-hoc* comparisons showed that Autonomous and Intrinsic player profiles performed statistically significant more assists than Amotivated profile ($Z = 3.224$, $p < 0.05$; $Z = 2.794$, $p < 0.05$).

The kill-death-assist ratio (*kda*) in ranked matches suggests that Autonomous players were the highest-performing profile, whereas the Amotivated profile performed worst ($Z = 2.922$, $p < 0.05$). Descriptively, the pattern is less clear for unranked matches where intrinsically motivated players have the lowest average value and amotivated and autonomous players are on par. However, the differences between the mean and median values is relatively large, suggesting that there are a outliers present who have very high *kda* values in unranked matches.

Taken together, the Amotivated profile's champions die the most, but they also kill more opponents compared to both Intrinsic and Autonomous player profiles. This may suggest that Amotivated players exhibit a more "reckless" playstyle compared to other profiles. However, this behavior appears less successful in ranked matches than in non-ranked ones, as indicated by the *kda* ratio and the *winrate*.

4.4.3. Economy Related Behaviors

Across all profiles the amount of gold earned and spent in both ranked and unranked matches is very similar, with only ranked matches showing slight differences. With multiple sources and ways to acquire gold, it is however difficult to determine how the motivational profiles relate to gold earned.

5. DISCUSSION

Playing games is commonly considered an enjoyable and intrinsically motivating activity (Ryan et al., 2006; Przybylski et al., 2010). League of Legends and other MOBA games, however, are massively popular, despite players reporting comparatively subpar experiences relative to other game genres (Johnson et al., 2015). The present study shows that people's underlying motivational regulations for playing *LoL* may play a crucial role therein. Based on Organismic Integration Theory, a mini-theory of Self-Determination Theory (Deci and Ryan, 2000), we identified four distinct motivational profiles, which differed markedly in their player experience and, to a lesser extent, in their in-game behavior. The Intrinsic player profile reported overall the most positive experience. Contrary to previous findings that MOBA games afford less autonomy and more frustration than other game genres (Johnson et al., 2015), players in this profile experienced a considerable sense of autonomy and competence when playing LoL, as well as reported low levels of tension and negative affect. In contrast, the Amotivated and External player profiles seem to derive markedly less enjoyment from their playing experience, as well as reported more tension and negative affect. They also scored lower on experienced autonomy and competence need satisfaction—with autonomy ratings similar to the ones reported by Johnson et al. (2015) (i.e., below $M = 5.0$).

As such, our findings are in line with OIT and previous research on motivational regulations and technology use. As posited by SDT, more self-determined player profiles (i.e., Intrinsic and Autonomous profiles) reported a more positive experiences (Deci and Ryan, 2000) and more harmonious passion for play (Vallerand et al., 2003), compared to the less self-determined profiles (Amotivated and External profiles). Moreover, recall that previous research found people reporting higher levels of amotivation to be more at risk of burn out (Gustafsson et al., 2018), as well as more likely to consider abandoning a technology (Brühlmann et al., 2018). As such, players in the Amotivated profile might be more inclined to quit playing LoL. While participants in our sample may be considered dedicated players, as evidenced by their being active in the *LoL* subreddit, the Amotivated and External profiles enjoyed playing substantially less. Indeed, lack of fun is one of the reasons players stop engaging with MOBAs (Tyack et al., 2016).

Our results also support existing findings on motivation and achievement goal orientation (Elliot and McGregor, 2001; Chen et al., 2019). Compared to the other profiles, the External profile scored higher on performance approach and performance avoidance orientation. Recall that this profile is more motivated by external pressure and avoiding feelings of guilt. These players may therefore feel particularly driven to perform well in *LoL* relative to their peers. However, performance and mastery approach orientation was rather high across all profiles, which is not surprising, considering the highly competitive nature of *LoL*, where players strive to improve their skills and perform well in front of their teammates (Johnson et al., 2015; Kahn et al., 2015; Tyack et al., 2016; Mora-Cantallops and Sicilia, 2018).

Next, the four profiles differ considerably in group size. Many more players fell into the Amotivated (29.3%) and External (43.9%) profiles than the Intrinsic (12%) or Autonomous profiles (14.8%). As such, it seems that a majority of players have a less positive experience when playing *LoL* and are not purely driven by intrinsic motivation. While Johnson et al. (2015) did not recruit participants over Reddit, it could be that the majority of MOBA players in their sample also fell into the Amotivated or External profiles, which might explain their more negative player experience ratings. What is less obvious is why these players reported less self-determined motivations. As of now, it is unclear if these players were already more amotivated and/or externally motivated when they started playing *LoL*—perhaps not to let a friend down Tyack et al. (2016),—or whether their motivation shifted over time.

Notably, our group size numbers are inconsistent with previous work on motivational regulation profiles. In their study of elite athletes, Gustafsson et al. (2018) found that only 22% of participants fell into the amotivated and moderately controlled profile (i.e., they reported more external and introjected regulation), with even fewer falling into the predominantly amotivated profile (6.9%). Similarly, in a study on work motivation (Howard et al., 2016), between 13.1 and 27.6% of participants were classified into the amotivated profile. With regards to the Intrinsic and Autonomous profiles, our findings are more comparable. The autonomous profiles in the aforementioned studies (Howard et al., 2016; Gustafsson et al., 2018) encompassed 15.9–25.6% of all participants.

Importantly, our study showcases that participants' motivations for playing *LoL* are not mutually exclusive. While some motivations were more salient for certain profiles (e.g., the Intrinsic profile), most profiles can be considered a motivational blend, where intrinsic motivation was reported along amotivation and other motivational regulations. Indeed, profiles share some considerable overlap, as intrinsic motivation was rather high across all player profiles. This is not surprising, as intrinsic motivation (i.e., seeking enjoyment in an activity) and the experience of enjoyment are key motivators for play for casual, heavy, and hardcore gamers (Neys et al., 2014).

In contrast to previous work on motivational profiles (Howard et al., 2016; Gustafsson et al., 2018), we observed no "high" motivation profile, i.e., where people score high on all motivational regulations, except amotivation. At least with regards to highly involved *LoL* players (i.e., active on the subreddit), it seems that certain motivational regulations are more salient (e.g., amotivation, intrinsic motivation). Nevertheless, our findings suggest that even small increments in amotivation, external and identified regulation are already associated with a less positive experience (operationalized as increased enjoyment, positive affect and need satisfaction, as well as lower levels of tension and negative affect).

5.1. Motivation and In-Game Behavior

Results indicate that motivational regulations shape patterns of need satisfaction and player experience. However, the four player profiles exhibited fairly similar in-game behavior overall. We found several statistically significant, albeit small to moderate correlations between game metrics and self-report measures. These results indicate a slight linear relationships between certain behaviors and motivational regulation. This is not surprising, as previous research examining game metrics and self-reported experience measures also reported low to medium correlations (Canossa et al., 2013; Schaekermann et al., 2017; Melhart et al., 2019). Among the 14 metrics we studied, the four motivational profiles varied significantly in terms of their *winrate* and assists in unranked matches, and kill-death-assist ratios in ranked matches (see also **Table 3** and **Figure 4**).

For kill-death-assist ratios in ranked matches and assists in unranked matches, the Amotivated profile showed the lowest median sore, while the more self-determined player profiles show slightly higher performance, especially in unranked matches.

In unranked matches, the Intrinsic player profile was characterized by an increased number of assists and a low rate of won games on the Summoner's Rift map. However, this profile did not report higher levels of relatedness. Hence, it would be misleading to claim that this profile featured more social or supportive players. Rather they seem to perceive their game play as highly autonomous and experience the most enjoyment of all profiles. Thus, they may simply enjoy the game and care less about winning than the other players, as reflected by the lower scores on performance approach and avoidance goal orientations.

However, behavioral metrics collected in this study are on a relatively high level of analysis (i.e., aggregated over all matches of a player) and findings need to be taken with a grain of salt. Consider that the behavioral metrics in our sample constitute of data aggregated over a longer period of time, whereas the self-report survey only covers a single measuring point. We examined metrics of *LoL* which reflect performance (e.g., winrate, kill-death-assist ratio), playstyle (e.g., killingSprees, totalHeal, championsPlayed), and engagement (e.g., totalMatches, timePlayed) aggregated over a period of about 18 months. If the effects of motivational regulations change over time, behavioral differences between the four motivational profiles may be only observable with detailed trend analyses. Further, the interplay of experience and behavior may be highly game-specific; there may be only a limited number of ways a game can be played. However, the few observed behavioral differences between the profiles show that similar behavior—with different underlying motivational regulations—can lead to very different experiences.

5.2. Limitations and Future Work

The present study is the first to apply OIT to better understand the interplay of player motivation, experience and in-game behavior in League of Legends. Specifically, we employed Latent Profile Analysis, a novel approach to profile players according to their motivational regulations. That said, our study comes with several caveats and limitations. First, note that due to the LPA approach, differences between profiles are relative. For example, while participants in the External player profile reported higher tension ($M = 4.10$), this is only slightly above the scale midpoint (3.5). Similarly, in terms of obsessive passion and negative affect, all profiles scored below the scale midpoint on average (3.5 and 2.5, respectively). Overall, participants in our sample did not report negative experiences when playing League of Legends. Nevertheless, it seems that minor fluctuations in motivational regulations may already shape the player experience toward more adverse or more positive outcomes.

That said, the exploratory nature of this study does not allow for causal inferences. Although in line with SDT propositions, it is unclear whether motivational regulations shape experiential outcomes and in-game behavior, whether players' experiences and behaviors impact their motivation, or—most likely—whether there are bidirectional effects. Repeated data collection of self-reported and logged behavioral data may provide more insights into how different motivational regulations affect experience and changes in behavior. It may also help mitigate certain limitations inherent to retrospective self-reports (i.e., recall bias) (Solhan et al., 2009).

Second, due to the cross-sectional design of the study (i.e., only one measuring point for self-reported motivation and experience), the present work cannot make any statements about potential changes in motivation over time. Longitudinal studies are necessary to assess whether motivational player profiles remain fairly stable, or fluctuate when players start playing, have been playing for a long time already, or decide to stop playing (Tyack et al., 2016). As such, future work should consider how long players have already engaged with *LoL* or other MOBAs.

Another promising avenue for studying motivational shifts over time is to consider the notion of internalization. Recall that SDT posits motivational regulations may shift through the process of internalization, along the controlled-to-autonomous continuum (Deci and Ryan, 2000, see also **Figure 2**, from left to right). When people take up values, attitudes, or regulatory structures, initially externally regulated behaviors may become internalized and then no longer require the presence of rewards or pressure (Deci and Ryan, 2000). For instance, it could be that certain players are initially both intrinsically and externally motivated. That is, they might choose to play *LoL* to experience enjoyment, but also due to perceived pressure from friends and teammates (Tyack et al., 2016). Over time, and over repeatedly experiencing a sense of autonomy, competence and relatedness, players might shift toward the Autonomous player profile, because playing *LoL* becomes personally meaningful to them. Or they might perhaps shift to the predominantly Intrinsic player profile, as they no longer feel pressured from others or themselves to play.

Longitudinal studies on players' motivational regulations could also provide insights into other aspects of MOBA play. For instance, whether professional esports athletes go through different motivational shifts than more casual players, due to experiencing more pressure to play or succeed (Deterding, 2016; Peters et al., 2018). Or whether the experience of toxic social interactions (Kwak and Blackburn, 2014; Shores et al., 2014) result in initially intrinsically motivated players shifting toward external regulation or even amotivation. Identifying such contributing factors could facilitate the design of interventions to counteract negative effects early on, as well as inform game design to promote mastery over performance orientation in players.

Third, note that the motivational profiles outlined in the present study only represent a momentary snapshot, whereas the processed behavioral data extend over a period of about 18 months—over which League of Legends has undergone several patches and changes. As such, the collected data operate on two different levels of analysis. While rather challenging and time-consuming, it would be useful to collect self-reports of motivational regulation and player experience after each season or pre-season, or better yet, after individual matches. This would allow for a more tightly coupled and granular analysis of the interplay of motivational regulation and in-game behavior, as well as help control for various changes due to patch updates and the introduction of new champions (Mora-Cantallops and Sicilia, 2018). It would also be interesting to classify players based on their in-game behavior (e.g., as in Melhart et al., 2019), and then compare them in terms of their motivational regulations.

Fourth, a sample selection bias toward highly engaged players is likely, as participants were recruited from the LoL subreddit. As such, participants were not only eager LoL players, but clearly also invested in the metagame (Donaldson, 2017), e.g., they read patch notes or discuss strategies with other players. Future studies should therefore take into account whether participants identify as hardcore or more casual *LoL* players (Neys et al., 2014), as well as how they perceive their reputation within the player community, which may affect their motivational regulation (and vice versa). Conversely, novice players might be more oriented toward mastering the game mechanics, especially when playing with friends (Tyack et al., 2016), and may not yet be as performance oriented (Elliot and McGregor, 2001).

Moreover, our sample is biased toward men, with only slightly over 6% of participants identifying as women or non-binary, slightly less than the expected 10%[5]. As gender stereotypes are known to affect the in-game character design, players' perception of abilities, and social conventions in LoL (Gao et al., 2017), future studies should be mindful of the experiences and motivations of female, non-binary and trans players.

It remains to be seen whether the present findings generalize to other MOBAs or game genres. According to SDT, the negative effects of less self-determined motivational regulations and amotivation on well-being are largely context-independent (Deci and Ryan, 2000). Hence, we expect that similar player profiles broadly manifest for other MOBAs and genres, and that motivational regulations may similarly shape players' experience—although the number and specific patterns of motivational profiles may vary to some extent.

Lastly, it would be interesting to combine OIT with other motivational frameworks or personality models. Indeed, recent works successfully combined game analytics and self-report questionnaires of player typologies to profile players and identify game design improvements (Yee et al., 2012; Canossa et al., 2013; Kahn et al., 2015; Schaekermann et al., 2017). According to SDT (Deci and Ryan, 2000), all of these motivational typologies describe "what" activity players seek to pursue (e.g., curiosity, competition, socializing, etc.), whereas the motivational regulations posited by OIT refer to the underlying reasons "why" these activities are being pursued. Similarly, according to causality orientation—another SDT mini-theory—people differ in the extent to which they generally experience their actions as self-determined (Deci and Ryan, 2000). As such, it is possible that some participants in our sample were broadly more Autonomy or Control oriented (i.e., more inclined toward autonomous or external regulations, respectively), or tended toward amotivation, regardless of any situational factors. Finally, OIT could be combined with other personality models, such as the Big Five model (Sheldon and Prentice, 2019), which has already been successfully combined with game analytics (Canossa et al., 2015).

[5]https://www.statista.com/statistics/694381/gamer-share-world-genre-and-gender/ (viewed: 28 Jan 2020).

6. CONCLUSION

We present findings from a theory-driven exploratory approach toward understanding player motivation and experiences in League of Legends. Combining Self-Determination Theory, Latent Profile Analysis and game analytics, we identified four motivational profiles, which differ with regards to player experience and, to a lesser extent, player behavior. In particular, our findings highlight the importance of considering amotivation and extrinsic regulation types, which hitherto have received only scant attention in player experience research. As such, this paper provides researchers and game designers with a novel and theoretically grounded perspective on player motivation.

DATA AVAILABILITY STATEMENT

The anonymized survey data can be accessed on the Open Science Framework https://osf.io/ue82s/. Anonymized aggregated behavioral metrics are available upon request.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

FB, GW, SK, and EM contributed the conception and design of the study. PB prepared the survey. EM distributed and collected the survey data. GW matched and prepared the behavioral data. FB and PB performed the statistical analysis. FB, PB, and EM wrote the first draft of the manuscript. FB, PB, GW, SK, and EM wrote the sections of the manuscript. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.01307/full#supplementary-material

REFERENCES

Alexandrovsky, D., Friehs, M. A., Birk, M. V., Yates, R. K., and Mandryk, R. L. (2019). "Game dynamics that support snacking, not feasting," in *Proceedings of the Annual Symposium on Computer-Human Interaction in Play, CHI PLAY '19* (New York, NY: Association for Computing Machinery), 573–588. doi: 10.1145/3311350.3347151

Azadvar, A., and Canossa, A. (2018). "UPEQ: ubisoft perceived experience questionnaire: a self-determination evaluation tool for video games," in *Proceedings of the 13th International Conference on the Foundations of Digital Games* (New York, NY: ACM), 5. doi: 10.1145/3235765.3235780

Bartle, R. (1996). Hearts, clubs, diamonds, spades: players who suit MUDs. *J. Virt. Environ.* 1:19.

Bauckhage, C., Kersting, K., Sifa, R., Thurau, C., Drachen, A., and Canossa, A. (2012). "How players lose interest in playing a game: an empirical study based on distributions of total playing times," in *Proceedings of the IEEE Conference on Computational Intelligence and Games, CIG* (Piscataway, NJ: IEEE), 139–146. doi: 10.1109/CIG.2012.6374148

Birk, M. V., and Mandryk, R. L. (2018). "Combating attrition in digital self-improvement programs using avatar customization," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18* (New York, NY: Association for Computing Machinery). doi: 10.1145/3173574.3174234

Blizzard Entertainment (2004). *World of Warcraft*. Game [PC]. Irvine, CA: Blizzard Entertainment.

Brühlmann, F., Petralito, S., Aeschbach, L. F., and Opwis, K. (2020). The quality of data collected online: an investigation of careless responding in a crowdsourced sample. *Methods in Psychol.* 2:100022. doi: 10.1016/j.metip.2020.100022

Brühlmann, F., Vollenwyder, B., Opwis, K., and Mekler, E. D. (2018). "Measuring the "why" of interaction: development and validation of the user motivation inventory (UMI)," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18* (New York, NY: ACM), 106:1–106:13. doi: 10.1145/3173574.3173680

Bungie Inc. (2014). *Destiny*. Game [PS4]. Santa Monica, CA: Activision.

Canossa, A., Badler, J. B., El-Nasr, M. S., Tignor, S., and Colvin, R. C. (2015). "In your face (t) impact of personality and context on gameplay behavior," in *Proceedings of the 10th International Conference on the Foundations of Digital Games, (FDG 2015)* (Pacific Grove, CA: Society for the Advancement of the Science of Digital Games).

Canossa, A., Martinez, J. B., and Togelius, J. (2013). "Give me a reason to dig Minecraft and psychology of motivation," in *2013 IEEE Conference on Computational Intelligence in Games (CIG)* (Piscataway, NJ: IEEE), 1–8. doi: 10.1109/CIG.2013.6633612

CCP (2003). *EVE Online*. Game [PC]. New York, NY: Simon & Schuster.

Chen, C., Elliot, A. J., and Sheldon, K. M. (2019). Psychological need support as a predictor of intrinsic and external motivation: the mediational role of achievement goals. *Educ. Psychol.* 39, 1090–1113. doi: 10.1080/01443410.2019.1618442

Chen, Z., Sun, Y., El-nasr, M. S., and Nguyen, T.-H. D. (2017). Player skill decomposition in multiplayer online battle arenas. *arXiv* arXiv:1702.06253.

Cumming, G., and Finch, S. (2005). Inference by eye: confidence intervals and how to read pictures of data. *Am. Psychol.* 60, 170–180. doi: 10.1037/0003-066X.60.2.170

Deci, E. L., and Ryan, R. M. (2000). The "what" and "why" of goal pursuits: human needs and the self-determination of behavior. *Psychol. Inquiry* 11, 227–268. doi: 10.1207/S15327965PLI1104_01

Deci, E. L., and Ryan, R. M. (2002). *Handbook of Self-Determination Research*. Rochester, NY: University Rochester Press.

Deterding, S. (2016). "Contextual autonomy support in video game play: a grounded theory," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (New York, NY: ACM), 3931–3943. doi: 10.1145/2858036.2858395

Diallo, T. M., Morin, A. J., and Lu, H. (2017). The impact of total and partial inclusion or exclusion of active and inactive time invariant covariates in growth mixture models. *Psychol. Methods* 22, 166–190. doi: 10.1037/met0000084

Donaldson, S. (2017). Mechanics and metagame: exploring binary expertise in league of legends. *Games Culture* 12, 426–444. doi: 10.1177/1555412015590063

Drachen, A., and Canossa, A. (2009). "Analyzing spatial user behavior in computer games using geographic information systems," in *Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era, MindTrek '09* (New York, NY: ACM), 182–189. doi: 10.1145/1621841.1621875

Drachen, A., Yancey, M., Maguire, J., Chu, D., Wang, I. Y., Mahlmann, T., Schubert, M., and Klabajan, D. (2014). "Skill-based differences in spatio-temporal team behaviour in defence of the ancients 2 (dota 2)," in *2014 IEEE Games Media Entertainment* (Toronto, ON), 1–8. doi: 10.1109/GEM.2014.7048109

Eggert, C., Herrlich, M., Smeddinck, J., and Malaka, R. (2015). "Classification of player roles in the team-based multi-player game Dota 2," in *Proceedings of the 14th International Conference on Entertainment Computing, ICEC* (Cham: Springer International Publishing), 112–125. doi: 10.1007/978-3-319-24589-8_9

Elliot, A. J., and McGregor, H. A. (2001). A 2 × 2 achievement goal framework. *J. Pers. Soc. Psychol.* 80, 501–519. doi: 10.1037/0022-3514.80.3.501

Feng, W., Brandt, D., and Saha, D. (2007). "A long-term study of a popular MMORPG.," in *Proceedings of the 6th ACM SIGCOMM Workshop on Network and System Support for Games, NetGames '07* (New York, NY: ACM), 19–24. doi: 10.1145/1326257.1326261

Gao, G., Min, A., and Shih, P. C. (2017). "Gendered design bias: gender differences of in-game character choice and playing style in league of legends," in *Proceedings of the 29th Australian Conference on Computer-Human Interaction, OZCHI '17* (New York, NY: Association for Computing Machinery), 307–317. doi: 10.1145/3152771.3152804

Gao, L., Judd, J., Wong, D., and Lowder, J. (2013). *Classifying Dota 2 Hero Characters Based on Play Style and Performance*. Salt Lake City, UT: University of Utah Course on ML.

Guay, F., Vallerand, R. J., and Blanchard, C. (2000). On the assessment of situational intrinsic and extrinsic motivation: the situational motivation scale (SIMS). *Motiv. Emot.* 24, 175–213. doi: 10.1023/A:1005614228250

Gustafsson, H., Carlin, M., Podlog, L., Stenling, A., and Lindwall, M. (2018). Motivational profiles and burnout in elite athletes: a person-centered approach. *Psychol. Sport Exerc.* 35, 118–125. doi: 10.1016/j.psychsport.2017.11.009

Harpstead, E., Zimmermann, T., Nagapan, N., Guajardo, J. J., Cooper, R., Solberg, T., and Greenawalt, D. (2015). "What drives people: creating engagement profiles of players from game log data," in *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play, CHI PLAY '15* (New York, NY: ACM), 369–379. doi: 10.1145/2793107.2793114

Hazan E. (2013) "Contextualizing Data," in *Game Analytics*, eds M. Seif El-Nasr, A. Drachen, and A. Canossa, (London: Springer). doi: 10.1007/978-1-4471-4769-5_21

Howard, J., Gagné, M., Morin, A. J., and Van den Broeck, A. (2016). Motivation profiles at work: a self-determination theory approach. *J. Vocat. Behav.* 95, 74–89. doi: 10.1016/j.jvb.2016.07.004

Johanson, C., Gutwin, C., Bowey, J. T., and Mandryk, R. L. (2019). "Press pause when you play: comparing spaced practice intervals for skill development in games," in *Proceedings of the Annual Symposium on Computer-Human Interaction in Play, CHI PLAY '19* (New York, NY: Association for Computing Machinery), 169–184. doi: 10.1145/3311350.3347195

Johnson, D., Gardner, J., and Sweetser, P. (2016). Motivations for videogame play: predictors of time spent playing. *Comput. Hum. Behav.* 63, 805–812. doi: 10.1016/j.chb.2016.06.028

Johnson, D., Nacke, L. E., and Wyeth, P. (2015). "All about that base: differing player experiences in video game genres and the unique case of MOBA games," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15* (New York, NY: ACM), 2265–2274. doi: 10.1145/2702123.2702447

Kahn, A. S., Shen, C., Lu, L., Ratan, R. A., Coary, S., Hou, J., Meng, J., Osborn, J., and Williams, D. (2015). The trojan player typology: a cross-genre, cross-cultural, behaviorally validated scale of video game play motivations. *Comput. Hum. Behav.* 49, 354–361. doi: 10.1016/j.chb.2015.03.018

Kam, C., Morin, A. J., Meyer, J. P., and Topolnytsky, L. (2016). Are commitment profiles stable and predictable? A latent transition analysis. *J. Manag.* 42, 1462–1490. doi: 10.1177/0149206313503010

Kou, Y., Li, Y., Gui, X., and Suzuki-Gill, E. (2018). "Playing with streakiness in online games: how players perceive and react to winning and losing streaks in League of Legends," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18* (New York, NY: ACM), 578:1–578:14. doi: 10.1145/3173574.3174152

Kwak, H., and Blackburn, J. (2014). "Linguistic analysis of toxic behavior in an online video game," in *International Conference on Social Informatics* (Barcelona: Springer), 209–217. doi: 10.1007/978-3-319-15168-7_26

Kwak, H., Blackburn, J., and Han, S. (2015). "Exploring cyberbullying and other toxic behavior in team competition online games," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15* (New York, NY: Association for Computing Machinery), 3739–3748. doi: 10.1145/2702123.2702529

Lafrenière, M.-A. K., Verner-Filion, J., and Vallerand, R. J. (2012). Development and validation of the gaming motivation scale (GAMS). *Pers. Indiv. Differ*. 53, 827–831. doi: 10.1016/j.paid.2012.06.013

Massive Entertainment (2016). *Tom Clancy's The Division*. Game [PC]. Montreuil: Ubisoft.

McAuley, E., Duncan, T., and Tammen, V. V. (1989). Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: a confirmatory factor analysis. *Res. Quart. Exerc. Sport* 60, 48–58. doi: 10.1080/02701367.1989.10607413

Melhart, D., Azadvar, A., Canossa, A., Liapis, A., and Yannakakis, G. N. (2019). Your gameplay says it all: modelling motivation in Tom Clancy's The Division. *arXiv* 1902.00040. doi: 10.1109/CIG.2019.8848123

Mojang (2011). *Minecraft*. Game [PC]. Redmond, WA: Microsoft Studios.

Mora-Cantallops, M., and Sicilia, M.-Á. (2018). Moba games: a literature review. *Entertain. Comput*. 26, 128–138. doi: 10.1016/j.entcom.2018.02.005

Morin, A. J., and Marsh, H. W. (2015). Disentangling shape from level effects in person-centered analyses: an illustration based on university teachers' multidimensional profiles of effectiveness. *Struct. Equat. Model*. 22, 39–59. doi: 10.1080/10705511.2014.919825

Nacke, L. E., Bateman, C., and Mandryk, R. L. (2014). Brainhex: a neurobiological gamer typology survey. *Entertain. Comput*. 5, 55–62. doi: 10.1016/j.entcom.2013.06.002

Nascimento Junior, F. F. d., Melo, A. S. d. C., da Costa, I. B., and Marinho, L. B. (2017). "Profiling successful team behaviors in league of legends," in *Proceedings of the 23rd Brazillian Symposium on Multimedia and the Web, WebMedia '17* (New York, NY: Association for Computing Machinery), 261–268. doi: 10.1145/3126858.3126886

Neys, J. L., Jansz, J., and Tan, E. S. (2014). Exploring persistence in gaming: the role of self-determination and social identity. *Comput. Hum. Behav*. 37, 196–209. doi: 10.1016/j.chb.2014.04.047

Pastor, D. A., Barron, K. E., Miller, B., and Davis, S. L. (2007). A latent profile analysis of college students' achievement goal orientation. *Contemp. Educ. Psychol*. 32, 8–47. doi: 10.1016/j.cedpsych.2006.10.003

Perry, R., Drachen, A., Kearney, A., Kriglstein, S., Nacke, L. E., Sifa, R., Wallner, G., and Johnson, D. (2018). Online-only friends, real-life friends or strangers? Differential associations with passion and social capital in video game play. *Comput. Hum. Behav*. 79, 202–210. doi: 10.1016/j.chb.2017.10.032

Peters, D., Calvo, R. A., and Ryan, R. M. (2018). Designing for motivation, engagement and wellbeing in digital experience. *Front. Psychol*. 9:797. doi: 10.3389/fpsyg.2018.00797

Przybylski, A. K., Rigby, C. S., and Ryan, R. M. (2009). A motivational model of video game engagement. *Rev. Gen. Psychol*. 14, 154–166. doi: 10.1037/a0019440

Przybylski, A. K., Weinstein, N., Ryan, R. M., and Rigby, C. S. (2018). Having to versus wanting to play: background and consequences of harmonious versus obsessive engagement in video games. *Cyber Psychol. Behav*. 12, 485–492. doi: 10.1089/cpb.2009.0083

Pseudonym117 (2010). *Riot-Watcher*. Available online at: https://github.com/pseudonym117/Riot-Watcher (accessed Jan 28, 2020).

Puerta-Cortés, D. X., Panova, T., Carbonell, X., and Chamarro, A. (2017). How passion and impulsivity influence a player's choice of videogame, intensity of playing and time spent playing. *Comput. Hum. Behav*. 66, 122–128. doi: 10.1016/j.chb.2016.09.029

Reiss, S., and Havercamp, S. M. (1998). Toward a comprehensive assessment of fundamental motivation: factor structure of the Reiss profiles. *Psychol. Assess*. 10:97. doi: 10.1037/1040-3590.10.2.97

Riot Games (2009). *League of Legends*. Game [PC]. Los Angeles, CA: Riot Games.

Riot Games (2018). *Riot Developer Portal*. Available online at: https://developer.riotgames.com/ (accessed Jan 28, 2020).

Ryan, R. M., and Deci, E. L. (2000). Intrinsic and extrinsic motivations: classic definitions and new directions. *Contemp. Educ. Psychol*. 25, 54–67. doi: 10.1006/ceps.1999.1020

Ryan, R. M., and Frederick, C. (1997). On energy, personality, and health: subjective vitality as a dynamic reflection of well-being. *J. Pers*. 65, 529–565. doi: 10.1111/j.1467-6494.1997.tb00326.x

Ryan, R. M., Mims, V., and Koestner, R. (1983). Relation of reward contingency and interpersonal context to intrinsic motivation: a review and test using cognitive evaluation theory. *J. Pers. Soc. Psychol*. 45:736. doi: 10.1037/0022-3514.45.4.736

Ryan, R. M., Rigby, C. S., and Przybylski, A. (2006). The motivational pull of video games: a self-determination theory approach. *Motiv. Emot*. 30, 344–360. doi: 10.1007/s11031-006-9051-8

Schaekermann, M., Ribeiro, G., Wallner, G., Kriglstein, S., Johnson, D., Drachen, A., Sifa, R., and Nacke, L. E. (2017). "Curiously motivated: profiling curiosity with self-reports and behaviour metrics in the game "Destiny"," in *Proceedings of the Annual Symposium on Computer-Human Interaction in Play, CHI PLAY '17* (New York, NY: ACM), 143–156. doi: 10.1145/3116595.3116603

Scrucca, L., Fop, M., Murphy, T. B., and Raftery, A. E. (2016). mclust 5: clustering, classification and density estimation using gaussian finite mixture models. *R J*. 8, 289–317. doi: 10.32614/RJ-2016-021

Sheldon, K. M., and Prentice, M. (2019). Self-determination theory as a foundation for personality researchers. *J. Pers*. 87, 5–14. doi: 10.1111/jopy.12360

Shores, K. B., He, Y., Swanenburg, K. L., Kraut, R., and Riedl, J. (2014). "The identification of deviance and its impact on retention in a multiplayer game," in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Baltimore, MD: ACM), 1356–1365. doi: 10.1145/2531602.2531724

Solhan, M. B., Trull, T. J., Jahng, S., and Wood, P. K. (2009). Clinical assessment of affective instability: comparing EMA indices, questionnaire reports, and retrospective recall. *Psychol. Assess*. 21:425. doi: 10.1037/a0016869

Suznjevic, M., Stupar, I., and Matijasevic, M. (2011). "MMORPG player behavior model based on player action categories," in *Proceedings of the 10th Annual Workshop on Network and Systems Support for Games, NetGames '11* (Piscataway, NJ: IEEE), 6:1–6:6. doi: 10.1109/NetGames.2011.6080982

Tofighi, D., and Enders, C. K. (2008). "Identifying the correct number of classes in growth mixture models," in: *Advances in Latent Variable Mixture Models*, eds G. R. Hancock, and K. M. Samuelson (Charlotte, NC: Information Age).

Tyack, A., and Mekler, E. D. (2020). "Self-determination theory in hci games research: current uses and open questions," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20* (New York, NY: ACM). doi: 10.1145/3313831.3376723

Tyack, A., Wyeth, P., and Johnson, D. (2016). "The appeal of MOBA games: what makes people start, stay, and stop," in *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play, CHI PLAY '16* (New York, NY: ACM), 313–325. doi: 10.1145/2967934.2968098

Vallerand, R. J., Blanchard, C., Mageau, G. A., Koestner, R., Ratelle, C., Léonard, M., Gagné, M., and Marsolais, J. (2003). Les passions de l'ame: on obsessive and harmonious passion. *J. Pers. Soc. Psychol*. 85, 756–767. doi: 10.1037/0022-3514.85.4.756

Valve Corporation (2013). *Dota 2*. Game [PC]. Bellevue, WA: Valve Corporation.

Vella, K., Johnson, D., and Hides, L. (2013). "Positively playful: when videogames lead to player wellbeing," in *Proceedings of the First International Conference on Gameful Design, Research, and Applications, Gamification '13* (New York, NY: ACM), 99–102. doi: 10.1145/2583008.2583024

Wallner, G., Kriglstein, S., Gnadlinger, F., Heiml, M., and Kranzer, J. (2014). "Game user telemetry in practice: a case study," in *Proceedings of the 11th Conference on Advances in Computer Entertainment Technology, ACE '14* (New York, NY: ACM), 45:1–45:4. doi: 10.1145/2663806.2663859

Wang, C. K. J., Liu, W. C., Nie, Y., Chye, Y. L. S., Lim, B. S. C., Liem, G. A., Tay, E. G., Hong, Y.-Y., and Chiu, C.-Y. (2017). Latent profile analysis of students' motivation and outcomes in mathematics: an organismic integration theory perspective. *Heliyon* 3:e00308. doi: 10.1016/j.heliyon.2017.e00308

Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the panas scales. *J. Pers. Soc. Psychol*. 54, 1063–1070. doi: 10.1037/0022-3514.54.6.1063

Yang, P., Harrison, B. E., and Roberts, D. L. (2014). "Identifying patterns in combat that are predictive of success in MOBA games," in *Proceedings of the 9th International Conference on the Foundations of Digital Games, FDG* (Ft. Lauderdale, FL: Society for the Advancement of the Science of Digital Games).

Yee, N. (2006). Motivations for play in online games. *Cyber Psychol. Behav*. 9, 772–775. doi: 10.1089/cpb.2006.9.772

Yee, N., Ducheneaut, N., and Nelson, L. (2012). "Online gaming motivations scale: development and validation," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12* (New York, NY: ACM), 2803–2806. doi: 10.1145/2207676.2208681

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Computer Science

ORIGINAL RESEARCH
published: 06 August 2020
doi: 10.3389/fcomp.2020.00029

Validation of User Preferences and Effects of Personalized Gamification on Task Performance

Gustavo F. Tondello[1,2]* and *Lennart E. Nacke*[2,3]

[1] Cheriton School of Computer Science, University of Waterloo, Waterloo, ON, Canada, [2] HCI Games Group & the Games Institute, University of Waterloo, Waterloo, ON, Canada, [3] Stratford School of Interaction Design and Business & Department of Communication Arts, University of Waterloo, Waterloo, ON, Canada

Personalized gamification is the tailoring of gameful design elements to user preferences to improve engagement. However, studies of user preferences have so far relied on self-reported data only and few studies investigated the effects of personalized gameful systems on task performance. This study shows that personalized gamification works in practice as predicted by survey studies and leads to higher task performance. We asked 252 participants in two studies to interact with a customized (experimental) or a generic (control) online gameful application to classify images. In the customized version, they could select the game elements that they wanted to use for their experience. The results showed significant correlations between participants' choice of gameful design elements and their Hexad user type scores, which partly support existing user preference models based on self-reported preferences. On the other hand, user type scores were not correlated with participants' preferred game elements rated after interacting with the gameful system. These findings demonstrate that the Hexad user types are a viable model to create personalized gameful systems. However, it seems that there are other yet unknown factors that can influence user preferences, which should be considered together with the user type scores. Additionally, participants in the experimental condition classified more images and rated their experience of selecting the game elements they wanted to use higher than in the control, demonstrating that task performance improved with personalization. Nonetheless, other measures of task performance that were not explicitly incentivized by the game elements did not equally improve. This contribution shows that personalized gameful design creates systems that are more successful in helping users achieve their goals than generic systems. However, gameful designers should be aware that they must balance the game elements and how much they incentivize each user behavior, so that the business goals can be successfully promoted. Finally, we analyzed participants' qualitative answers about their experience with the generic and the customized gameful applications, extracting useful lessons for the designers of personalized gameful systems.

Keywords: gamification, gameful design, personalization, adaptation, customization, Hexad user types

OPEN ACCESS

Edited by:
Marc Herrlich, University of Kaiserslautern, Germany

Reviewed by:
Patrick Dickinson, University of Lincoln, United Kingdom
Athanasios Mazarakis, University of Kiel, Germany

****Correspondence:***
Gustavo F. Tondello
gustavo@tondello.com

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Computer Science

Received: *31 January 2020*
Accepted: *29 June 2020*
Published: *06 August 2020*

Citation:
Tondello GF and Nacke LE (2020) Validation of User Preferences and Effects of Personalized Gamification on Task Performance. Front. Comput. Sci. 2:29. doi: 10.3389/fcomp.2020.00029

1. INTRODUCTION

Gamification is now a established design approach in human-computer interaction (HCI) to create engaging gameful systems (Seaborn and Fels, 2015; Landers et al., 2018; Koivisto and Hamari, 2019). Gamification or gameful design is the use of gameful design elements in non-game contexts (Deterding et al., 2011). In the past 5 years, gamification research has been maturing. Recent publications have been developing the theories that inform the gameful design practice and providing detailed empirical evidence of the effects of specific gameful design elements, for specific users, in specific contexts (Nacke and Deterding, 2017; Landers et al., 2018; Rapp et al., 2019).

One of the approaches to improve the design of gameful systems is personalized (or adaptive) gamification, meaning the tailoring of the gameful design elements, the interaction mechanics, the tasks, or the game rules according to the preferences or skills of each user (Lessel et al., 2016; Böckle et al., 2017; Tondello et al., 2017b; Klock et al., 2018; Tondello, 2019). Recent advances in the study of personalized gamification include the development of personalized gameful design methods (see section 2.1), the development of user preferences models and taxonomies of game elements (see section 2.2), and the evaluation of the effects of personalized gameful systems (see section 2.3).

Nonetheless, studies of user preferences have so far mostly relied on self-reported data instead of observation of actual user behavior. In addition, only a few studies investigated the effects of personalized gameful systems in comparison to generic alternatives. In the present work, we contribute to the literature on personalized gamification by observing user interaction with an online gameful system to study their game element preferences and the effects of personalization on their behavior and performance. In two studies, we observed 252 participants who interacted with either a customized (experimental) or a generic (control) version of a gameful image classification platform and reported on their experiences. Participants on the experimental condition were allowed to select the gameful design elements for their interaction with the platform, whereas participants in the control condition had all the gameful design elements available without the possibility of customization. This research answers two questions:

RQ1: If allowed to choose the gameful design elements they prefer, do user choices correspond to the theoretical relationships with user types, personality, gender, and age reported in previous survey-based studies?

RQ2: Are user performance and engagement better for a personalized gameful system than a generic system?

The results show several significant correlations between participants' choices of gameful design elements in the personalized condition with their Hexad user type scores, congruent to the expected relationships between elements and types according to the existing literature (Tondello et al., 2016b, 2017a). However, the results were less conclusive for personality traits, gender, and age. In addition, participants in the experimental condition classified more images and rated the experience of selecting which game elements to use higher than participants in the control condition. This new empirical evidence based on user behavior supports the user preference models previously devised based on Hexad user types and self-reported preferences. It also adds to the growing body of knowledge on personalization in gamification research demonstrating that user performance can be improved with personalized gameful design.

This contribution is important to the HCI and gamification communities because it provides evidence of the validity of personalized gameful design methods based on the selection of gameful design elements considering the different Hexad user types (such as Lessel et al., 2018; Marczewski, 2018; Mora Carreño, 2018; Tondello, 2019). Therefore, gamification designers can use the insights from this and the related works to create personalized gameful systems that are more effective than generic systems in helping users achieve their goals, such as improved learning, engagement, health, or well-being.

2. RELATED WORK

2.1. Methods for Personalized Gameful Design

Personalized gamification (or gameful design) is the tailoring of the gameful design elements, the interaction mechanics, the tasks, or the game rules for each user, according to their preferences. The tailoring is usually based on some knowledge about the users and their preferences and aims to boost the achievement of the goals of the gameful system (Tondello, 2019, chapter 3). Personalization in gamification is inspired by the reported positive results with other digital applications in general (Adomavicius and Tuzhilin, 2005; Sundar and Marathe, 2010), and more specifically in closely related applications such as games (e.g., Bakkes et al., 2012; Orji et al., 2013, 2014) and persuasive technologies (e.g., Nov and Arazy, 2013; Kaptein et al., 2015; Orji and Moffatt, 2018). Personalization can be implemented in two ways (Sundar and Marathe, 2010; Orji et al., 2017; Tondello, 2019):

- as a **customization** (also referred as user-initiated personalization), where the user selects the elements that they wish to use;
- as a (semi-)**automatic adaptation** (also referred as system-initiated personalization), where the system takes the initiative to select the gameful design elements for each user—with or without some user input in the process.

In previous work, we proposed a method for personalized gameful design (Tondello, 2019) based on three steps: (1) classification of user preferences using the Hexad user types (Tondello et al., 2016b, 2019b), (2) classification and selection of gameful design elements, where the user selects what elements they want to use (customization) or the system (semi-)automatically selects elements based on the user's Hexad scores and the classification of gameful design elements (Tondello et al., 2017a), and (3) a heuristic evaluation (Tondello et al., 2016a, 2019a) to verify if all the dimensions of motivational affordances are potentially integrated into the design.

Mora Carreño (2018) employs a similar approach based on the Hexad user types and a selection of gameful design elements for different groups of users. His work is more focused on the design of educational gamification services.

Lessel et al. (2016) also present a similar approach that is based on letting users customize their gameful experience by deciding when to use gamification and what elements to use. However, it is more focused on letting users freely choose from a defined (Lessel et al., 2016) or undefined (Lessel et al., 2018) set of gameful design elements, instead of relying on user types to aid in the selection. They have named this approach "bottom-up gamification."

Böckle et al. (2018) also propose a framework for adaptive gamification. It is based on four main elements, which can be applied to the gameful design process in diverse orders: (1) the purpose of the adaptivity, which consists on defining the goal of the adaptation, such as support of learning of participation, (2) the adaptivity criteria, such as user types or personality traits, which serve as an input for the adaptation, (3) the adaptive game mechanics and dynamics, which is the actual tailoring of game elements to each user, and (4) adaptive interventions, such as suggestions and recommendations, which represent the adaptation in the front-end layer.

In the gamification industry, Marczewski (2018) uses the Hexad user types to select gameful design elements for different users or as design lenses to design for different audiences. Furthermore, Chou (2015) considers different user profiles in one of the levels of the Octalysis Framework. The specific user model to be employed is not specified, with common examples being Bartle's player types (Bartle, 1996) and the Hexad user types.

Looking at these personalized gameful design methods together, there are some commonalities between them. All these methods suggest some means of understanding the user (e.g., user types or personality traits), some means of selecting gameful design elements for different users, and some mechanism to allow users to interact with the adaptation (e.g., customization or recommendation). In the present work, we build upon our previous publications by evaluating the user experience with a gameful application created using our personalized gameful design method (Tondello, 2019) and comparing the results with related works.

2.2. User Preference Models

The Hexad framework (Tondello et al., 2016b, 2019b; Marczewski, 2018) is the most used model of user preferences in gamification (Klock et al., 2018; Bouzidi et al., 2019). Monterrat et al. (2015) also developed a mapping of gamification elements to BrainHex player types (Nacke et al., 2014). However, Hallifax et al. (2019) compared the Hexad user types with the BrainHex and the Big-5 personality traits (Goldberg, 1993; Costa and McCrae, 1998). They concluded that the Hexad is the most appropriate for use in personalized gamification because the results with the Hexad were the most consistent with the definitions of its user types and it had more influence on the perceived user motivation from different gameful design elements than the other two models.

Although there are studies of the relationships between the Hexad user types and different variables in the literature, the relationship with participants' preferred gameful design elements is of particular interest for our study because our personalized gameful application relies on element selection. Publications that provide data about these relationships include the works of Tondello et al. (2016b, 2017a), Marczewski (2018), Orji et al. (2018), Mora et al. (2019), and Hallifax et al. (2019).

Studies that investigate user preferences in gamification by personality traits, gender, and age are also abundant in the literature. Again, we are interested in the publications that establish relationships between these variables and participants' preferred gameful design elements, so we could validate the relationships in the present study. Publications that provide these relationships with personality traits include the works of Butler (2014), Jia et al. (2016), Tondello et al. (2017a), Orji et al. (2017), and Hallifax et al. (2019); relationships with gender are provided by Tondello et al. (2017a) and Codish and Ravid (2017); and relationships with age are provided only by Tondello et al. (2017a).

These findings suggest that if allowed to choose the gameful design elements for their experience, participants' choices would be influenced by their user type scores, personality trait scores, gender, and age. Therefore, our first research question (**RQ1**) aims to validate these relationships.

2.3. Evaluation of Personalized Gameful Systems

We previously conducted a pilot study of personalized gamification (Tondello, 2019, chapter 7) using the same gameful application that we use in this study. We asked 50 participants to select four gameful design elements to customize their experience. The goal of that pilot study was to test the personalized gameful design method and gather participants' impressions regarding how they customize their experience. Progress feedback was the game element that was selected more often by participants: 36 times. It was followed by levels (30), power-ups (30), leaderboards (23), chance (23), badges (20), unlockable content (16), challenges (16), and moderating role (6 times).

The user types and personality trait scores were generally not good predictors of game element selection in the pilot study. However, there were some significant relationships: participants who chose challenges scored lower in conscientiousness; participants who chose unlockable content scored higher in the user type achiever and in emotional stability; participants who chose leaderboards scored lower in conscientiousness; participants who chose levels scored higher in the user type achiever and in openness to experiences; and participants who chose progress feedback scored lower in the user types socialiser and achiever, as well as emotional stability.

In the qualitative analysis, around 80% of participants expressed a positive experience, 10% expressed a negative experience, and 10% were neutral. The answers highlighted how participants enjoyed the variety of elements offered and the perceived control over their own experience. This shows that participants generally appreciated the customization options. Participants who expressed neutral or negative experiences would have preferred no gamification at all, rather than having an issue with the customization. Therefore, we concluded that participants can understand, carry out, and comment on the gamification customization task. Therefore, we suggested that more studies should be carried out with more participants and comparing personalized with non-personalized conditions to better understand the effects of personalized gamification, which is precisely what we do in the present study.

In the educational context, Mora et al. (2018) compared a generic with a personalized gameful learning experience with 81 students of computer network design. The descriptive statistics suggested that personalization seems to better engage students behaviorally and emotionally. However, the characteristics of the sample did not lead to any statistically significant result, suggesting that additional studies would be needed to confirm the preliminary findings. Herbert et al. (2014) observed that learner behavior on a gameful application varied according to their user types. Araújo Paiva et al. (2015) created a pedagogical recommendation system that suggested missions to students according to their most common and least common interactions, to balance their online behavior. Roosta et al. (2016) evaluated a gamified learning management system for a technical English course and demonstrated that student participation increased in a personalized version in comparison with a control version. Barata et al. (2017) conducted an extensive study to classify student behavior with a gameful interactive course. Based on their results, they presented a model that classifies students in four clusters and provided design lessons for personalized gameful education systems.

Evaluating their "bottom-up gamification" approach, Lessel et al. (2017) conducted a study with 106 participants in which they had to complete several image classification, article correction, or article categorization tasks. Several conditions where tested, from a fully generic gameful system (in which all elements were enabled) to a fully customizable system (in which participants could combine the elements in any way), and a control condition with no gamification. Participants who could customize their experience performed significantly better, solving more tasks faster without a decrease in correctness. The authors conclude that "bottom-up gamification" can lead to a higher motivational impact than fixed gamification.

In another study with 77 participants, Lessel et al. (2019) tested the impact of allowing participants to enable or disable gamification for an image tagging task. They found out that the choice did not affect participants who used gamification, but it improved the motivation of participants who were not attracted by the elements when they had the choice. Therefore, allowing users to enable or disable gamification seems to be a simple, but useful customization option when more sophisticated personalization is not available.

Böckle et al. (2018) employed their adaptive gameful design method to gamify an application for knowledge exchange in medical training. They compared application usage in the 6 months directly after introduction of adaptive gamification and in the period preceding it and noted an increase in overall system activity. However, they did not explicitly test if the effect was due to the adaptive nature of the implementation, or just due to the introduction of gamification itself.

Altogether, these related works show promising evidence that personalized gameful systems can be more engaging and lead to better task performance than generic systems with fixed gameful design elements. However, additional studies are required to replicate these initial findings and expand the available evidence to different applications and contexts. In response, we seek to provide additional evidence that personalized gamification increases user engagement and task performance (**RQ2**) in a context that was previously tested before: image classification tasks. Therefore, we provide additional evidence of the benefits of personalized gamification by replicating the positive effects of previous studies in a similar context, but with a different personalized design.

3. METHODS

3.1. Gameful Application

The two studies reported here were carried out using a gameful online application developed by the first author. The platform was designed as a customizable system that uses a variety of gameful design elements implemented around a central task, which was an image classification task for these studies. Thus, each task consisted on listing all the classification tags that the participant could think of for a stock image. Royalty-free stock images were randomly downloaded from Pexels[1]. The gameful design elements can be activated or deactivated by the researcher or the user, allowing experiments to be conducted in which participants interact with different sets of elements.

The use of classification tasks was already reported on previous studies of customizable gamification (Altmeyer et al., 2016; Lessel et al., 2017). Therefore, this is an interesting type of task to allow for comparisons with previous results. Moreover, these tasks are similar to brainstorming tasks, which have also been used in previous empirical studies of gamification (Landers et al., 2017) because they have been found to provide a good opportunity to investigate task performance in relation to goal setting. By combining these two types of tasks in our study, we implemented gameful design elements with the goal of motivating participants on two levels: (1) to complete more tasks and (2) to perform better in each task by listing a higher number of tags.

Following our proposed method for personalized gameful design (Tondello, 2019), we employed gameful design elements that would be appealing to users with different preferences. This design method suggests trying to include at least one or two game elements from each of the eight groups identified by Tondello et al. (2017a). The rationale for the design elements selected from each group for inclusion in the application is as follows:

- **Progression elements:** *Levels* are a common choice of progression element because they are easy to implement and are generally engaging. Therefore, it was our chosen progression element for the application.

[1] https://www.pexels.com/

- **Altruism elements:** This group includes elements that promote social interactions in which one user helps the other. In our application, direct help was not possible because users did not interact with each other directly. Therefore, we chose the element *moderating role*, as we anticipated that by feeling they could help moderate the tags entered into the platform, users could feel they were somehow being helpful.
- **Incentive:** This group includes elements that reward the user for completing tasks. We selected two types of incentives that we could easily implement in the application: *badges* and *unlockable content* (additional avatar choices).
- **Socialization:** Similar to the altruism group, social interaction was limited in the application because users did not have direct contact with each other. Therefore, we decide to implement only a *leaderboard* because it is a social element that works without the need for direct user interaction.
- **Risk/Reward:** This group includes elements that reward the user for taking chances or challenges. Together with elements from the Incentive group, these elements can be very engaging in short-term experiences. Therefore, we selected two elements from this group: *challenges* and earnings moderated by *chance*.
- **Assistance:** This group includes elements that help the user accomplish their goals. We selected *power-ups* as the assistance element for our platform because it is generally easy to implement and well-received by users.
- **Customization:** We chose to let users change their *avatar* in the platform as an element of the customization group.
- **Immersion:** We did not find any suitable immersion element that we could easily implement. The tasks that users had to complete (image tagging) were not very immersive on their own, unless users decided to focus on taking some time to appreciate the images that they were tagging. Other elements that could provide additional immersion, such as a narrative or theme, could not be easily integrated into the application in the available time for development. Therefore, we did not select any element from this group.

The gameful design elements included in the application are listed in **Table 1**. **Figure 1** shows the user interface of the application. In addition to the elements listed in the table, four features were implemented to support the gameful elements: points, progress feedback, avatars, and customization.

Points are used by the following elements: levels, to decide when the user should level up; unlockable content, so users can spend points to unlock additional avatars; leaderboards, which allow users to compare the amount of points they earned with other users; chance, which applies a random modifier to the amount of points earned after each task; and power-ups, which apply a fixed modifier to the amount of points earned. Points are automatically enabled when any of these elements are also enabled, otherwise they are disabled. Users earn 10 points each time they submit tags for an image, with an additional one point per tag provided.

Progress feedback is implemented in form of a progress bar that shows how many of the total available images the user has already completed and how many are left to be completed. It was always enabled. An avatar can be selected by the user to represent them in the system. It is always possible to select an avatar, but the available options are limited unless the game element unlockable content is enabled. Customization allows the user to select what gameful design elements they want to use in the application. In this study, customization was enabled for participants in the customized (experimental) condition and disabled for the generic (control) condition.

3.2. Study Design

3.2.1. Experimental Conditions

Participants were divided into two conditions:

- Participants in the **generic** (control) condition were presented the list of game design elements for information only and all elements were automatically enabled for them. This conditions represents a generic (or one-size-fits-all) system because all participants should have similar experiences as they all have the same game elements in the interface. This mimics the current approach in gamification (without personalization), which consists in including different elements into the system to please different users, but without offering any mechanism for adaptation. We believe that this may overwhelm the user with too many elements to interact with, lead them to just ignore the game elements, or force users to select the elements they want to use just by directing their attention, i.e., by using the desired elements and ignoring the others in the interface.
- Participants in the **customized** (experimental) condition were asked to select as many game elements they wanted to use from the eight available options (see **Table 1**). **Figure 2** shows the user interface for customization, including the description of each game element provided to users before their selection. This is an example of user-initiated personalization (customization). The goal of this customization is to allow the users to improve their experience by removing the elements they do not want from the interface. In other words, the game elements that users do not select will not appear while they are working in the image classification tasks. Therefore, it should be easier for users to interact with the selected elements on a cleaner interface, potentially improving their experience and engagement. While answering our second research question, we will evaluate if these expectations will indeed correspond to the experience reported by the participants.

3.2.2. RQ1: Influence of User Characteristics on Element Selection

Our first research question is "If allowed to choose the gameful design elements they prefer, do user choices correspond to the theoretical relationships with user types, personality, gender, and age reported in previous survey-based studies?" The values for these four demographic variables were obtained from a survey presented to participants at the start of the experiment. We used the 24-item Hexad user types scale from Tondello et al. (2019b) and the 10-item Big-5 personality traits scale from Rammstedt and John (2007). The dependent variables were boolean values representing if the user selected each game element or not when given the choice in the customized condition. Therefore, data from

TABLE 1 | Gameful design elements implemented in the application.

Element (type)	Description
Levels (Progression)	After submitting the tags for each image, users would see a popup dialog informing if they leveled up as they earned points. The current level is also always displayed in the menu bar.
Moderating role (Altruism)	After writing tags for an image, the user can check a list of tags given by other people on a popup dialog and flag the unrelated tags.
Badges (Incentive)	Users earn badges as they complete tasks. When this element is selected, a new menu option appears that allows users to check the acquired and available badges and select one of the acquired badges to display in their profile besides their nickname.
Unlockable content (Incentive)	When this element is selected, additional customization options for the avatar are displayed, which are initially unlocked. Users can spend virtual coins (points) to unlock and use them.
Leaderboards (Socialization)	When this element is selected, a new option appears in the menu. Users can then see how they compare to others (points and level) in the leaderboard.
Challenges (Risk/Reward)	When this element is selected, a new menu option appears that allow users to see the available challenges, such as tagging a certain number of images or writing a certain number of tags for an individual image. Users earn additional points by completing any of the challenges.
Chance (Risk/Reward)	After each completed task, the amount of points received will be decided by luck. When this element is selected, a value between 5 and 1/5 is randomly selected, the earned points are multiplied by this value, and the results are displayed to the user in the popup dialog.
Power-ups (Assistance)	A power-up boosts the number of points received by the user for a few tasks (e.g., double the points earned for the next five images). When this element is selected, users will randomly earn a power-up after submitting the tags for an image. This power-up can be activated at any time in the image classification interface and will apply the boost for the next classified images.

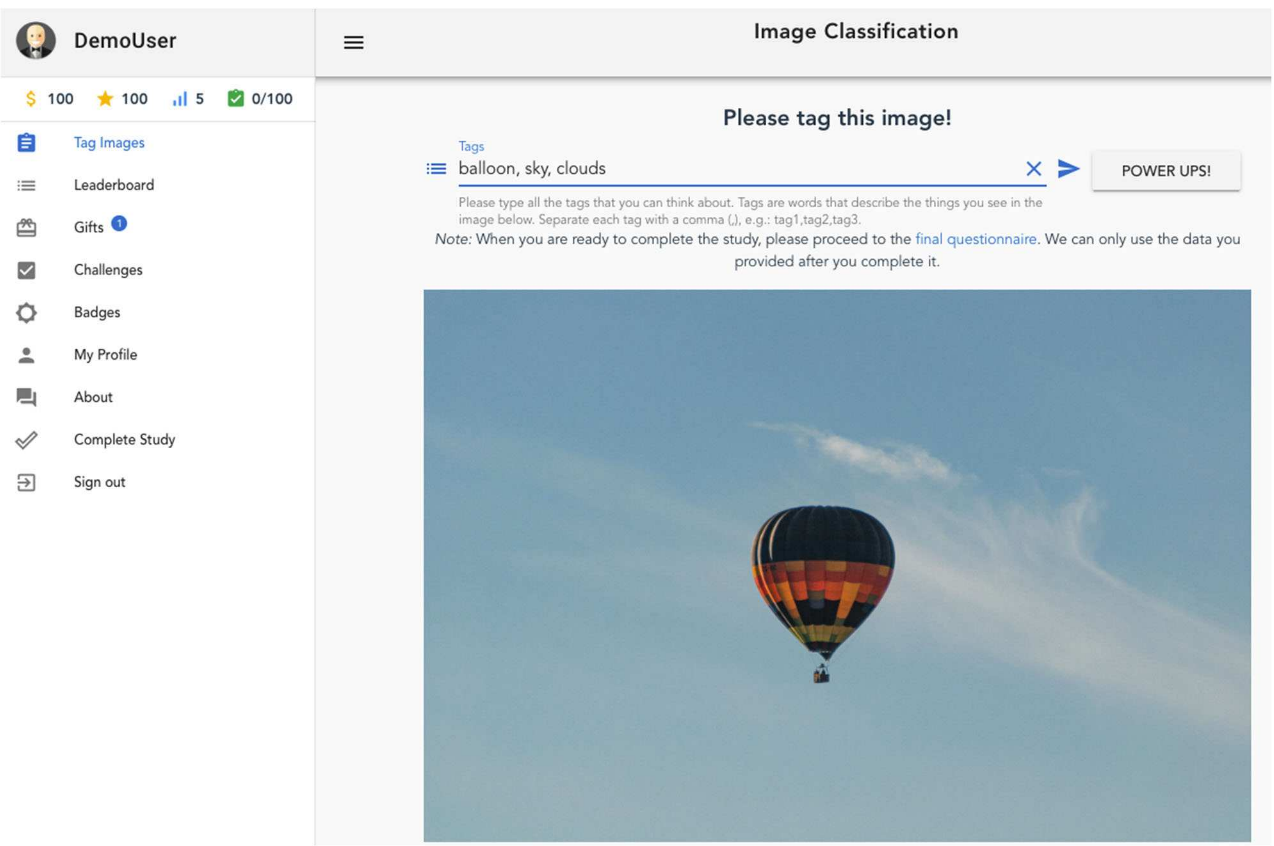


FIGURE 1 | Gameful image classification application used in this study.

participants in the generic (control) condition were not used to answer RQ1 as they were not given the chance to select game elements.

Based on the significant relationships between Hexad user type scores and game elements preferences observed by Tondello et al. (2016b,

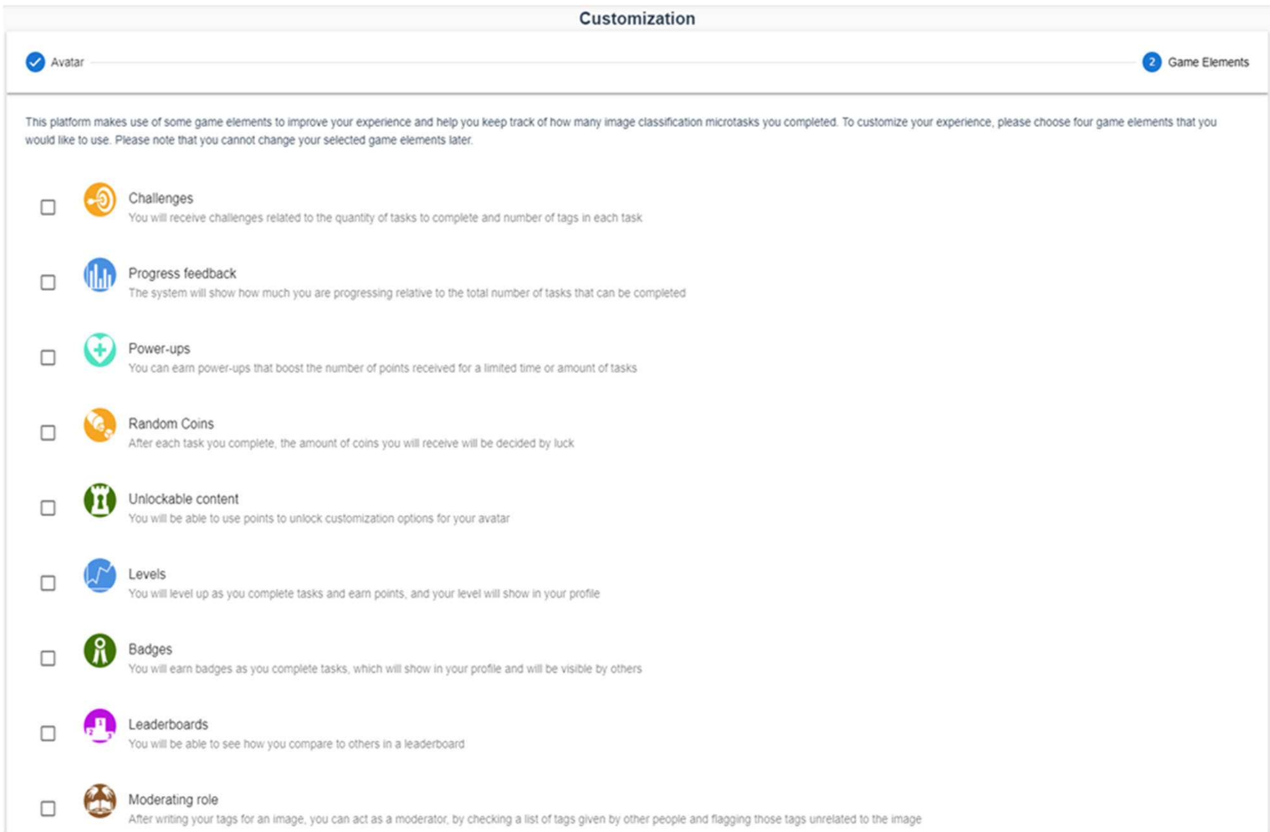


FIGURE 2 | User interface for game element selection. Participants could customize their experience by selecting any number of elements from this interface, based on the descriptions depicted in the image.

2017a) and Orji et al. (2018), we formulated the following hypotheses:

H1: The user type scores are different between participants who selected or not each game element in the application.

- **H1.1:** Participants who select Levels have higher Achiever and Player scores than those who do not select it.
- **H1.2:** Participants who select Moderating role have higher Philanthropist and Socializer scores than those who do not select it.
- **H1.3:** Participants who select Badges have higher Achiever and Player scores than those who do not select it.
- **H1.4:** Participants who select Unlockable content have higher Free Spirit and Player scores than those who do not select it.
- **H1.5:** Participants who select Leaderboards have higher Socializer and Player scores than those who do not select it.
- **H1.6:** Participants who select Challenges have higher Achiever, Player, and Disruptor scores than those who do not select it.
- **H1.7:** Participants who select Chance have higher Achiever and Player scores than those who do not select it.

Based on the significant relationships between personality trait scores and game element preferences observed by Jia et al. (2016) and Tondello et al. (2017a), we formulated the following hypotheses:

H2: The personality trait scores are different between participants who selected or not each game element in the application.

- **H2.1:** Participants who select Levels have higher Extraversion and Conscientiousness scores than those who do not select it.
- **H2.2:** Participants who select Moderating role have higher Extraversion scores than those who do not select it.
- **H2.3:** Participants who select Badges have lower Emotional Stability scores than those who do not select it.
- **H2.4:** Participants who select Leaderboards have higher Extraversion scores than those who do not select it.
- **H2.5:** Participants who select Challenges have higher Agreeableness scores than those who do not select it.

Based on the significant relationships between gender and game element preferences observed by Tondello et al. (2017a) and Codish and Ravid (2017), we formulated the following hypotheses:

H3: The frequency that each game element is selected is different by gender.

- **H3.1:** Men select Leaderboards and Moderating role more often than women.
- **H3.2:** Women select Badges, Unlockable content, and Power-ups more often than men.

Based on the significant relationships between age and game element preferences observed by Tondello et al. (2017a), we formulated the following hypothesis:

H4: The average participant age is lower for those who select Moderating role, Badges, Unlockable content, Challenges, and Chance than those who do not select it.

3.2.3. RQ2: Task Performance and User Engagement

Our second research question is "Are user performance and engagement better for a personalized gameful system than a generic system?" Because image tagging is the main user task, the quantity of images tagged, total number of tags for all images, and average number of tags per image are the direct measures of user performance in the task. Additionally, we wanted to evaluate if user performance would also improve for the measures generated by the game elements, which are total points earned and final level achieved. Although these are not direct indicators of performance in the image tagging task, they may represent how much the user was invested in the application. Finally, another measure that helps understand user involvement is the total amount of time spent in the application.

To measure user engagement, we employed the Intrinsic Motivation Inventory (IMI; McAuley et al., 1989) because it has been previously used in similar gamification studies. Additionally, we asked participants to directly rate their overall game selection experience on a Likert scale (see **Q2** in the next subsection), as this seemed a more direct form of participant feedback regarding their perceived engagement than the IMI questions. Therefore, direct participant rating is a more direct but non-standardized measure of engagement, whereas the IMI scale is a less direct but standardized measure.

As the literature reviewed in section 2.3 showed that user performance and engagement was generally better for personalized gameful applications than generic ones, we formulated the following hypotheses:

H5: User Performance measures are higher for participants in the experimental condition than in the control condition.

H6: User Engagement measures are higher for participants in the experimental condition than in the control condition.

3.3. Procedure

After following the link to the application, participants had to read and accept the informed consent letter. It described the image tagging tasks and framed the study as image classification research, without mentioning that we were actually studying gameful design elements. This initial deception was done to ensure that participants would interact naturally with the gameful elements without any bias.

Next, participants answered a short demographic information form that asked about their gender, age, Hexad user types, and Big-5 personality traits scale. Then, they were invited to customize their profile by selecting a nickname and an avatar. For the final step of the initial part, participants were assigned to one of the experimental conditions in counter-balanced order. Participants in the control condition were presented with a list of game elements for information only, whereas participants in the customized (experimental) condition were also able to select which game elements they wanted to use for the image classification task.

Upon completion of the initial part, participants were left to interact with the platform freely. Logically, the image tagging tasks were the focus point of the platform. In the first study, participants were recruited via Mechanical Turk and could complete as many tasks as they wanted (with no lower limit) up to the limit of 50 available images. The tasks were to be completed in one sitting. During this period, they could also interact with the features provided by the gameful design elements that they selected (experimental condition) or all elements (control condition). On the other hand, participants were recruited via social media for the second study and could interact with the application as many times as they wanted for 7 days. They could complete as many tasks as they wanted (with no lower limit) up to the limit of 100 available images. These participants also received a daily email reminder (sent by one of the researchers) that they needed to go back to the platform and complete the study by filling out the final survey.

When they felt they had tagged enough images, participants clicked the option "Complete Study" in the menu. At this point, they were asked to complete a questionnaire that included the Intrinsic Motivation Inventory (IMI) and the following free-text questions:

- **Q1:** Overall, how do you describe your experience with the image classification activities you just completed?
- **Q2:** How do you describe and rate the experience of selecting game elements to customize the platform for you? *(Likert scale with very negative, negative, neutral, positive, and very positive, in addition to the free-text answer)*
- **Q3:** Were you satisfied with the selection of game elements provided by the system? Why?
- **Q4:** Were you able to select game elements that matched your preferences? Why?
- **Q5:** How much do you feel that the selection of game elements you used to customize the platform for you influenced your enjoyment of the image classification tasks? Why?
- **Q6:** Now that you have used this system, which one was your preferred game element to use? Please explain why it was your preferred element. *(selection box with the eight game elements, in addition to the free-text answer)*
- **Q7:** Now that you have used this system, which game element do you feel most influenced how you tagged images? *(selection box with the eight game elements)*
- **Q8:** Which game element motivated you more to tag images? *(selection box with the eight game elements)*

After completing the post-study questionnaire, participants were presented with a post-study information letter and additional consent form. This additional letter debriefed participants about the deception used in the study. Thus, the letter explained that participants were initially told that we were interested in the tags to help us develop image classification systems; however, we were actually interested in studying their experience with the gameful design elements. It also explained that this was done to avoid bias in the participant's interaction with the game elements and their responses about their experiences. Participants were then given the chance to accept or to decline having their study data used after knowing the real purpose of the study and were instructed to contact the researchers by email if they had any question about the deception employed in the study. These procedures followed the guidelines for ethical participant recruitment established by the Office of Research Ethics at the University of Waterloo. Upon completion of this last step, the software then generated a completion code for participants recruited via Mechanical Turk, which they used to complete the task on the platform and receive their payment.

3.4. Participants

We planned to collect two data sets to answer our research questions. For the first study, we recruited participants through Amazon's Mechanical Turk, which is being increasingly used for HCI experiments (Buhrmester et al., 2018). This form of recruitment allowed us to determine the number of participants we wanted to recruit. Therefore, we planned to recruit a total of 200 participants (100 per condition). However, for **RQ2**, one concern was if participant motivation would have any effect on their performance. As Mechanical Turk participants were paid a fixed amount for completion of the task, it would be reasonable to assume that some of them might want to complete the task as quickly as possible to maximize their earnings. Therefore, we also collected a second data set only with volunteers that were not receiving a fixed payment for participation (although they were offered a chance to enter a draw as an incentive). This allowed us to also analyze data from participants that were presumably more willing to collaborate with the study without being too concerned with maximizing their time usage. For this second data set, we recruited participants through social media. Thus, it was hard to control how many participants would voluntarily complete the study. We aimed to recruit at least 100 participants and ended with 127 people creating an account, but in the end only 54 completed the study (27 per condition). Nonetheless, we considered that this sample size was sufficient to test hypotheses **H5** and **H6**. These two hypotheses were tested separately for each data set.

To answer **RQ1** and test the associated hypotheses, we used only the data from participants in the customized condition because participants in the generic condition were not allowed to select their game elements. Thus, only the customized condition contained data that we could use to test **H1–H4**. Considering that the number of participants in this condition was 99 per condition in the first study and 27 per condition in the second, we combined the data from the two studies because the groups of participants who selected or not each game element would otherwise be too small to carry out reliable statistical analyses, especially in the second study. Additionally, we have no theoretical reason to believe that the recruitment source (Mechanical Turk or social media) would make any difference in participants' preferred game elements according to their demographic characteristics. Even if their motivation to complete image tagging tasks was different depending on if they were being paid or not, we assumed that their gaming preferences would not be affected by it. Although **Table 2** shows that there were some differences in the user type scores, personality trait scores, and average age between the two datasets, these are the independent variables being analyzed in the statistical tests. Therefore, they are not confounding variables in the analyses. Therefore, we consider that combining the two datasets does not create a confounding factor in the analyses.

TABLE 2 | Description of participants' user type scores and personality trait scores.

	Study 1				Study 2			
User type	**Med**	**Mean**	**SD**	**α**	**Med**	**Mean**	**SD**	**α**
Philanthropist	5.75	5.44	1.10	0.879	6.00	6.00	0.66	0.633
Socialiser	4.75	4.68	1.31	0.893	5.62	5.26	1.26	0.887
Achiever	5.75	5.61	0.97	0.848	6.00	5.94	0.70	0.710
Free spirit	5.50	5.47	0.98	0.762	5.75	5.60	0.65	0.260
Player	5.75	5.64	0.96	0.786	5.75	5.43	1.02	0.713
Disruptor	3.25	3.42	1.24	0.783	3.50	3.67	1.10	0.630
	Study 1				**Study 2**			
Personality trait	**Med**	**Mean**	**SD**	**α**	**Med**	**Mean**	**SD**	**α**
Extraversion	3.25	3.42	1.69	0.694	4.00	3.79	1.39	0.723
Agreeableness	4.50	4.68	1.49	0.532	5.00	4.85	1.02	0.352
Conscientiousness	6.00	5.45	1.37	0.649	4.50	4.62	1.27	0.570
Emotional stability	4.50	4.59	1.69	0.762	3.50	3.57	1.59	0.855
Openness to experiences	5.50	5.32	1.35	0.463	4.50	4.57	1.46	0.572

Study 1: $N = 198$. Study 2: $N = 54$. Median and Mean values based on a 7-point Likert scale (range: 1.0–7.0). Cronbach's α calculated with 4 items per user type and 2 items per personality trait.

As mentioned above, we recruited a total of 200 participants through Amazon Mechanical Turk for the first study, with 100 per condition in counter-balanced order. Participants were required to have a HIT (high intelligence task) approval rate greater than 97%, a number of HITs approved higher than 5,000, and reside in the United States of America. This was done to ensure that only workers with a good history in the platform accepted our task. The HIT description on Mechanical Turk contained a brief description of the image classification task without mentioning the gameful elements and a link to the online system. Participants were informed that the estimated duration of the task was between 30 min and 1 h and were paid a fixed amount of $4.00 (four US dollars) after completion of the task. This remuneration was paid to all participants who submitted a completion code for the HIT, even if they did not complete all the steps of the study procedure, congruent to the ethical participant recruitment guidelines.

After verification, we had to remove two participants who did not complete the final survey with the final participation agreement. Therefore, the final dataset contained 198 participants (99 per condition). The sample contained answers from 90 women and 106 men (2 not disclosed), with ages varying from 19 to 72 years old ($M = 36.9$, $SD = 10.6$). They spent an average of 26.2 min on the platform ($SD = 23.4$), tagged 25.4 images on average ($SD = 19.2$) with a total of 118.9 tags on average ($SD = 135.6$), and earned a total of 873 points on average ($SD = 1,054$). Participants in the customized condition selected between zero and eight game elements ($M = 3.5$, $SD = 2.2$, $Med = 3.0$, $Mod = 1.0$, $N = 99$).

For the second study, we recruited participants through social media (Facebook, Twitter, and Reddit) and email lists of people interested in our research. They did not receive any direct compensation, but were offered the opportunity to enter a draw for one out of two $200 (two hundred US dollars) international gift cards. They could interact with the platform freely for a suggested limit of 7 days, but this limit was not enforced. However, the study actually ended when each participant decided to complete the final survey.

In total, 127 participants created an account and started interacting with the application. They were assigned to one of the two conditions in counter-balanced order. However, only 54 participants completed the study by filling out the end survey (27 per condition), which constitutes our final data set. The sample contained answers from 25 women and 28 men (1 not disclosed), with ages varying from 18 to 50 years old ($M = 25.8$, $SD = 5.8$). They were from Canada (17), China (7), India (6), France (5), United States of America (4), Iran (4), Nigeria (2), and nine other countries (only 1 participant each). They tagged 45.4 images on average ($SD = 37.3$) with a total of 345.8 tags on average ($SD = 391.1$), and earned a total of 2,243 points on average ($SD = 2,369$). Participants interacted with the platform between two and 13 different days ($M = 4.8$, $SD = 2.5$) and completed 1,089 action on average ($SD = 1,988$). Participants in the customized condition selected between zero and eight game elements ($M = 4.0$, $SD = 2.3$, $Med = 4.0$, $Mod = 3.0$, $N = 27$).

Table 2 summarizes the descriptive statistics for the Hexad user types and personality trait scores for all participants. Although there are some differences in the average values for these demographic variables between the two data sets, these were the independent variables being analyzed in the tests for hypotheses **H1–H4**. Therefore, we do not consider that this difference may have affected our results.

4. RESULTS

We present the results in this section for each one of the research questions. All statistical analyses were performed using SPSS v. 23 (IBM, 2015).

4.1. RQ1: Influence of User Characteristics on Element Selection

To answer **RQ1**, we carried out several splits of the data set according to whether the participant selected a specific element or not. For example, we compared participants who selected leaderboards with those who did not select it, participants who selected levels with those who did not select it, and so on. As explained in section 3.4, we combined the data from both samples and used only data from participants in the experimental (customized) condition because this was the only condition in which participants were given the chance to select the game elements they wanted to use.

Table 3 presents the results of the statistical tests comparing the Hexad and personality trait scores between participants who selected or did not select each element. Because the scores were not parametric, we employed the Mann–Whitney U-test. We also calculated the effect size $r = Z \div \sqrt{N}$, as suggested by Field (2009, p. 550).

There are several significant differences in the Hexad user type scores in relation to element selection:

- **H1.1: not supported.** Participants who selected **Levels** did not have higher **Achiever** ($p = 0.1595$, $r = 0.135$) and **Player** ($p = 0.160$, $r = 0.125$) scores than those who did not select it.
- **H1.2: not supported.** Participants who selected **Moderating role** did not have higher **Philanthropist** ($p = 0.449$, $r = 0.067$) and **Socializer** ($p = 0.333$, $r = 0.086$) scores than those who did not select it.
- **H1.3: partially supported.** Participants who selected **Badges** had higher **Achiever** scores than those who did not select it ($p = 0.015$, $r = 0.216$). However, they did not have higher **Player** scores ($p = 0.765$, $r = 0.027$).
- **H1.4: not supported.** Participants who selected **Unlockable content** did not have higher **Free Spirit** ($p = 0.787$, $r = 0.024$) and **Player** ($p = 0.641$, $r = 0.042$) scores than those who did not select it.
- **H1.5: partially supported.** Participants who selected **Leaderboards** had higher **Player** scores than those who did not select it ($p = 0.006$, $r = 0.244$). However, they did not have higher **Socializer** scores ($p = 0.116$, $r = 0.140$)
- **H1.6: partially supported.** Participants who selected **Challenges** had higher **Achiever** ($p = 0.005$, $r = 0.249$) and **Player** ($p = 0.045$, $r = 0.179$) scores than those who did not select it. However, they did not have higher **Disruptor** scores ($p = 0.682$, $r = 0.036$).
- **H1.7: partially supported.** Participants who selected **Chance** had higher **Achiever** scores than those who did not select it ($p = 0.005$, $r = 0.175$). However, they did not have higher **Player** scores ($p = 0.266$, $r = 0.099$).

On the other hand, the following significant differences were not predicted by the existing literature and were not part of our hypotheses, but appeared in the results:

- **Philanthropist** scores are higher for participants who selected **Badges** ($p = 0.027$, $r = 0.196$). This relationship was not suggested in any previous research.
- **Free Spirit** scores are higher for participants who selected **Chance** ($p = 0.050$, $r = 0.175$). This relationship was also not suggested in previous research.
- **Player** scores are higher for participants who selected **Power ups** ($p = 0.029$, $r = 0.194$). This makes sense because

TABLE 3 | Non-parametric tests (Mann–Whitney U) comparing the differences in Hexad user type scores and personality trait scores between users who selected or not each gameful design element.

Elements		Hexad user types						Personality traits				
		Phil	**Soc**	**Ach**	**Free**	**Play**	**Dis**	**Ext**	**Agr**	**Con**	**Emo**	**Ope**
Levels	$\widetilde{\textbf{No}}$ (42)	5.75	5.50	5.75	5.75	5.75	3.50	3.50	4.75	5.50	4.50	5.00
	$\widetilde{\textbf{Yes}}$ (84)	6.00	5.00	6.00	5.75	6.00	3.25	3.50	5.00	6.00	4.50	6.00
	U	1580.5	1577.5	1472.5	1489.0	1494.0	1693.0	1762.5	1529.5	1379.0	1706.5	1451.5
	Z	0.956	0.968	1.519	1.432	1.404	0.368	0.008	1.019	1.827	0.299	1.435
	p	0.339	0.333	0.129	0.152	0.160	0.713	0.994	0.308	0.068	0.765	0.151
	r	0.085	0.086	0.135	0.128	0.125	0.033	0.001	0.091	0.163	0.027	0.128
Moderation	$\widetilde{\textbf{No}}$ (99)	5.75	5.25	6.00	5.75	6.00	3.25	3.50	5.00	6.00	4.50	6.00
	$\widetilde{\textbf{Yes}}$ (27)	5.75	5.25	5.63	5.75	5.75	3.75	4.00	4.50	6.00	4.50	4.75
	U	1210.0	1321.0	1132.5	1125.0	1014.0	1119.0	1125.0	1280.0	1210.0	1258.5	1025.5
	Z	0.757	0.092	1.221	1.265	1.926	1.296	1.263	0.260	0.687	0.466	1.800
	p	0.449	0.926	0.222	0.206	0.054	0.195	0.207	0.795	0.492	0.641	0.072
	r	0.067	0.008	0.109	0.113	0.172	0.115	0.112	0.023	0.061	0.042	0.160
Badges	$\widetilde{\textbf{No}}$ (60)	5.75	5.13	5.75	5.75	6.00	3.63	3.00	4.50	5.75	4.50	5.00
	$\widetilde{\textbf{Yes}}$ (66)	6.00	5.25	6.00	5.75	5.75	3.25	3.50	5.25	6.00	4.50	6.00
	U	1531.5	1746.5	1487.5	1652.5	1919.0	1640.0	1738.5	1488.0	1598.0	1909.0	1529.5
	Z	2.205	1.144	2.422	1.609	0.299	1.665	1.184	2.286	1.748	0.348	2.083
	p	**0.027**	0.252	**0.015**	0.108	0.765	0.096	0.236	**0.022**	0.080	0.728	**0.037**
	r	**0.196**	0.102	**0.216**	0.143	0.027	0.148	0.106	**0.204**	0.156	0.031	**0.186**
Unlockables	$\widetilde{\textbf{No}}$ (78)	5.75	5.25	5.88	5.75	5.75	3.38	3.50	5.00	6.00	4.50	5.50
	$\widetilde{\textbf{Yes}}$ (48)	5.75	5.00	6.00	5.75	6.00	3.38	3.25	5.00	6.00	4.50	5.50
	U	1805.0	1853.5	1622.5	1818.5	1779.5	1757.0	1655.5	1778.0	1813.5	1765.0	1774.0
	Z	0.339	0.093	1.262	0.270	0.467	0.579	1.092	0.358	0.177	0.540	0.379
	p	0.735	0.926	0.207	0.787	0.641	0.562	0.275	0.720	0.859	0.589	0.705
	r	0.030	0.008	0.112	0.024	0.042	0.052	0.097	0.032	0.016	0.048	0.034
Leaderboards	$\widetilde{\textbf{No}}$ (65)	5.75	5.00	5.75	5.75	5.75	3.50	3.50	5.00	6.00	4.50	5.00
	$\widetilde{\textbf{Yes}}$ (61)	6.00	5.25	6.00	5.75	6.00	3.25	4.00	5.00	6.00	5.00	6.00
	U	1863.0	1661.5	1613.5	1874.0	1424.0	1976.5	1734.0	1940.0	1869.5	1693.5	1539.5
	Z	0.587	1.572	1.813	0.533	2.739	0.029	1.218	0.060	0.413	1.417	2.055
	p	0.557	0.116	0.070	0.594	**0.006**	0.977	0.223	0.952	0.680	0.156	**0.040**
	r	0.052	0.140	0.162	0.047	**0.244**	0.003	0.109	0.005	0.037	0.126	**0.183**
Challenges	$\widetilde{\textbf{No}}$ (71)	5.75	5.25	5.75	5.75	5.75	3.25	3.00	5.00	6.00	4.50	5.50
	$\widetilde{\textbf{Yes}}$ (55)	6.00	5.25	6.00	5.75	6.00	3.50	3.50	5.00	6.00	4.50	6.00
	U	1718.5	1914.0	1387.0	1816.0	1546.5	1869.5	1625.5	1875.0	1770.5	1846.5	1684.0
	Z	1.158	0.190	2.800	0.675	2.006	0.409	1.615	0.250	0.778	0.524	1.209
	p	0.247	0.849	**0.005**	0.499	**0.045**	0.682	0.106	0.802	0.436	0.600	0.227
	r	0.103	0.017	**0.249**	0.060	**0.179**	0.036	0.144	0.022	0.069	0.047	0.108
Chance	$\widetilde{\textbf{No}}$ (79)	5.75	5.00	5.75	5.75	5.75	3.38	3.50	4.50	5.75	4.50	5.00
	$\widetilde{\textbf{Yes}}$ (47)	5.75	5.25	6.13	6.00	6.00	3.38	3.75	5.50	6.00	4.50	6.00
	U	1673.0	1677.5	1309.5	1469.5	1637.0	1772.5	1715.5	1258.0	1505.5	1623.0	1426.0
	Z	0.931	0.906	2.778	1.964	1.112	0.425	0.714	2.882	1.615	1.183	2.019
	p	0.352	0.365	**0.005**	**0.050**	0.266	0.671	0.475	**0.004**	0.106	0.237	**0.043**
	r	0.083	0.081	**0.247**	**0.175**	0.099	0.038	0.064	**0.257**	0.144	0.105	**0.180**
Power-ups	$\widetilde{\textbf{No}}$ (57)	5.75	5.25	5.75	5.75	5.75	3.25	3.50	5.00	6.00	4.50	5.50
	$\widetilde{\textbf{Yes}}$ (69)	5.75	5.25	6.00	5.75	6.00	3.50	3.50	5.00	6.00	4.50	6.00
	U	1806.5	1964.0	1638.0	1843.0	1524.0	1916.5	1815.0	1854.0	1789.0	1846.0	1898.5
	Z	0.789	0.012	1.621	0.609	2.179	0.246	0.746	0.390	0.719	0.593	0.168
	p	0.430	0.990	0.105	0.543	**0.029**	0.806	0.456	0.697	0.472	0.553	0.867
	r	0.070	0.001	0.144	0.054	**0.194**	0.022	0.066	0.035	0.064	0.053	0.015

$N = 126$.
Bolded values are significant at the 0.05 level.
$\widetilde{\textbf{No}}$: median scores for users who did not select each element (range: 1.0–7.0).
$\widetilde{\textbf{Yes}}$: median scores for users who selected each elements (range: 1.0–7.0).
The numbers in brackets following $\widetilde{\textbf{No}}$/$\widetilde{\textbf{Yes}}$ are the number of participants for each row.
***U*/*Z*/*p*: results of the Mann–Whitney U-tests.**
r: *effect sizes, calculated as $r = Z \div \sqrt{N}$.*
*The absolute values of **Z** and **r** are displayed for improved readability.*

TABLE 4 | Distribution of users who selected or not each gameful design element by gender.

	Levels	Moderation	Badges	Unlockables	Leaderboard	Challenges	Chance	Power-ups
N-M/F	20/22	50/49	34/26	39/38	31/34	36/35	42/36	29/28
Y-M/F	44/39	14/12	30/35	25/23	33/27	28/26	22/25	35/33
χ^2	0.325	0.092	1.380	0.024	0.667	0.016	0.581	0.004
p	0.569	0.762	0.240	0.876	0.414	0.899	0.446	0.947

N = 126.
N-M/F: proportion of men and women who did not select each game element.
Y-M/F: proportion of men and women who selected each game element.
χ^2/p: results of Pearson's Chi-square tests comparing the proportions above (Crosstabs option on SPSS).

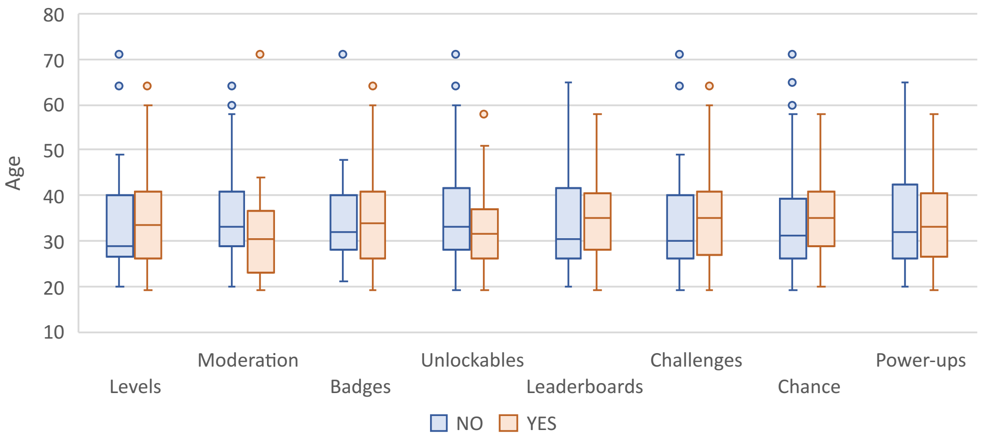


FIGURE 3 | Differences in age between users who selected or not each gameful design element.

power-ups allowed users to easily earn more points, which would be appealing to people with high scores on this user type.

Regarding participants' personality trait scores, none of the hypotheses were supported:

- **H2.1: not supported.** Participants who selected **Levels** did not have higher **Extraversion** ($p = 0.994$, $r = 0.001$) and **Conscientiousness** ($p = 0.068$, $r = 0.163$) scores than those who did not select it.
- **H2.2: not supported.** Participants who selected **Moderating role** did not have higher **Extraversion** scores than those who did not select it ($p = 0.207$, $r = 0.112$).
- **H2.3: not supported.** Participants who selected **Badges** did not have lower **Emotional Stability** scores than those who did not select it ($p = 0.728$, $r = 0.031$).
- **H2.4: not supported.** Participants who selected **Leaderboards** did not have higher **Extraversion** scores than those who did not select it ($p = 0.223$, $r = 0.109$).
- **H2.5: not supported.** Participants who selected **Challenges** did not have higher **Agreeableness** scores than those who did not select it ($p = 0.802$, $r = 0.022$).

On the other hand, there were some significant differences, which were not predicted by the existing literature and were not part of our hypotheses:

- **Agreeableness** scores are higher for participants who selected **Badges** ($p = 0.022$, $r = 0.204$) and **Chance** ($p = 0.004$, $r = 0.257$).
- **Openness** scores are higher for participants who selected **Badges** ($p = 0.037$, $r = 0.186$), **Leaderboards** ($p = 0.040$, $r = 0.183$), and **Chance** ($p = 0.043$, $r = 0.180$).

There were no significant relationships between the participants' selection of game elements and their genders (see **Table 4**). Therefore, **H3.1 and H3.2 are not supported**.

Regarding age, there was just one significant difference (see **Figure 3** and **Table 5**): participants who selected **Moderating role** were younger (*Med* = 30.5) than participants who did not select it (*Med* = 33.0, $p = 0.040$, $r = 0.183$; note that this is the absolute value of r because SPSS does not consider the direction of the relationship on the output of the Mann–Whitney *U*-test). However, age was not significantly different between participants who selected Badges, Unlockable content, Challenges, and Chance and the participants who did not select them. Therefore, **H4 is only partially supported**.

4.2. RQ2: Task Performance and User Engagement

To answer **RQ2**, we compared the participants' task performance between both conditions across the seven measures: total points earned, final level achieved, total images tagged, total tags entered

TABLE 5 | Non-parametric tests (Mann–Whitney U) comparing the differences in age between users who selected or not each gameful design element.

	Levels	Moderation	Badges	Unlockables	Leaderboard	Challenges	Chance	Power-ups
N/Y	42/84	99/27	60/66	78/48	65/61	71/55	79/47	57/69
$\widetilde{\textbf{No}}$	29.00	33.00	31.50	33.00	30.00	30.00	31.00	31.00
$\widetilde{\textbf{Yes}}$	33.50	30.50	34.00	31.50	35.00	35.00	35.00	33.00
U	1557.5	950.5	1882.5	1534.0	1810.0	1665.0	1457.0	1882.5
Z	0.866	2.049	0.319	1.595	0.702	1.294	1.918	0.246
p	.387	**0.040**	0.750	0.111	0.483	0.196	0.055	0.806
r	.077	**0.183**	0.028	0.142	0.063	0.115	0.171	0.022

N = 126.
Bolded values are significant at the 0.05 level.
***N/Y**: number of participants who did not select/did select each element.*
$\widetilde{\textbf{No}}$: median age for users who did not select each element.
$\widetilde{\textbf{Yes}}$: median age for users who selected each element.
***U/Z/p**: results of the Mann-Whitney U tests.*
***r**: effect sizes, calculated as $r = Z \div \sqrt{N}$.*
*The absolute values of **Z** and **r** are displayed for improved readability.*

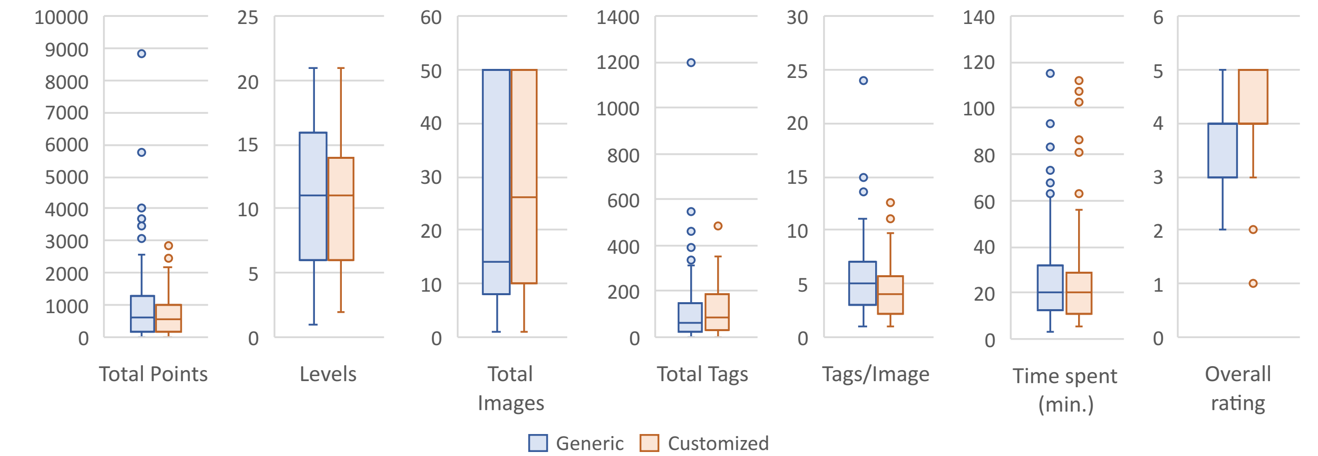


FIGURE 4 | Differences in performance and engagement variables between conditions (Study 1).

for all images, average tags per image, and time spent in the application (measured in minutes on study 1 and in days active in the application on study 2). User engagement was compared between both condition across the seven dimensions of the intrinsic motivation inventory (IMI). Participants' rating of their experience of selecting game elements (from their answer to **Q2**) was also compared as a measure of user engagement. Because the measures were not parametric, we employed the Mann–Whitney U-test and like in the previous subsection, we also calculated the effect size $r = Z \div \sqrt{N}$. We analyzed the data from each study separately to avoid the recruitment method (Mechanical Turk vs social media) as a confounding variable.

4.2.1. Study 1

Figure 4 displays the box plots comparing the performance and engagement variables between participants who selected or did not select each element. **Table 6** presents the results of the statistical tests.

Regarding task performance, users in the customized (experimental) condition classified more images ($Med = 26.0$) than in the generic (control) condition ($Med = 14.5$, $p = 0.013$, $r = 0.177$, a weak effect size). In this application, classifying more images means that participants contributed more to the systemic goal that was presented to them (collecting tags for images), and is therefore a relevant performance improvement. Nonetheless, the total number of tags did not change significantly between conditions. Because participants tagged more images in the customized condition, but wrote approximately the same total number of tags, the number of tags per image dropped significantly from $Med = 5.0$ in the generic condition to $Med = 4.0$ tags per image in the customized condition ($p = 0.008$, $r = 0.188$, a weak effect size). The other measures of task performance were not significantly different between conditions. Therefore, **H5 is partially supported in study 1**.

Regarding engagement, there were no statistically significant differences for any of the IMI measures. On the other hand, the experience rating was significantly higher in the customized condition than in the generic condition: $p = 0.025$, $r = 0.160$ (a weak effect size). Although the calculated median rating was 4.0 in both conditions, the boxplot in **Figure 4** shows that 50% of the

TABLE 6 | Comparison of performance and engagement variables and IMI scores between conditions (Study 1).

Variables	Median (generic)	Median (customized)	U	Z	*p*	r
Total points	611.5	551.0	4526.5	0.928	0.354	0.066
Level	11.0	11.0	4553.5	0.863	0.388	0.061
Total images	14.5	26.0	3915.5	2.495	**0.013**	**0.177**
Total tags	62.5	83.0	4548.5	0.873	0.383	0.062
Tags per image	5.0	4.0	3836.0	2.642	**0.008**	**0.188**
Time Spent (min.)	20.0	20.0	4807.5	0.231	0.817	0.016
Experience rating	4.0	4.0	3924.0	2.241	**0.025**	**0.160**
IMI scores	**Median (generic)**	**Median (customized)**	**U**	**Z**	***p***	**r**
Interest	5.00	5.17	4393.0	1.146	0.252	0.081
Competence	5.50	5.75	4184.5	1.782	0.075	0.127
Effort	5.50	6.00	4375.5	1.193	0.233	0.085
Pressure	2.00	2.25	4696.5	0.388	0.698	0.028
Choice	5.50	5.63	4388.5	1.159	0.246	0.082
Value	5.00	5.00	4362.0	1.225	0.220	0.087
Relatedness	3.83	4.00	4479.5	1.046	0.296	0.075

N = 198 (99 per condition).
Bolded values are significant at the 0.05 level.
Overall rating is a 5-point scale (range: 1–5).
***U/Z/p**: results of the Mann–Whitney U-tests.*
***r**: effect sizes, calculated as $r = Z \div \sqrt{N}$.*
*The absolute values of **Z** and **r** are displayed for improved readability.*

ratings in the generic condition were between 3 and 4, whereas 50% of the ratings in the customized condition were between 4 and 5. Therefore, **H6 is partially supported in study 1**.

4.2.2. Study 2

Figure 5 displays the box plots comparing the performance and engagement variables between participants who selected or did not select each element. **Table 7** presents the results of the statistical tests.

Regarding task performance, the number of images classified in the customized (experimental) ($Med = 51.0$) is higher than in the generic (control) condition ($Med = 25.0$); however, the difference is not significant: $p = 0.064$, $r = 0.132$. Although this effect is not significant in study 2, it is interesting to note on the box plot that only participants in the experimental condition classified all the available 100 images, but none in the control condition. Similarly to study 1, the total number of tags did not change significantly between conditions, but differently from the first study, this time the number of tags per image also did not change significantly between conditions. The other measures of task performance were once more not significantly different between conditions. Therefore, **H5 is not supported in study 2**.

Regarding engagement, participants scored higher in the IMI measure for competence in the customized condition ($Med = 5.25$) than the generic condition ($Med = 4.50$, $p = 0.022$, $r = 0.163$). The other IMI scores were not significantly different between conditions. In addition, the experience rating was significantly higher in the customized condition ($Med = 4.0$) than in the generic condition ($Med = 3.0$, $p = 0.012$, $r = 0.179$). This effect size is slightly larger than in the first study, and the difference in the medians is more pronounced, but the effect still has a similar order of magnitude (weak). Therefore, **H6 is partially supported in study 2**.

4.3. Participants' Perceived Usefulness of Each Element

Table 8 presents the number of times that each gameful design element was listed as the participant's preferred element, the element that most influenced them, or the element that most motivated them (in response to **Q6**, **Q7**, and **Q8**). The differences in the frequency distributions between conditions are significant for all three variables: $p = 0.003$ for preferred element, $p = 0.005$ for most influential element, $p = 0.001$ for most motivational element (Pearson's chi-square test; $N = 227$).

Although this was not one of the original research questions for this study, an analysis of this table provides interesting insights for personalized gameful design.

First, it is noticeable that the number of times that participants mentioned each element as preferred, influential, or motivating is similar, meaning that participants probably enjoy an element when they perceive it as influential or motivational. An interesting exception is that a few participants in the generic condition perceived Levels as the most motivating element even if it was not their preferred or most influential element.

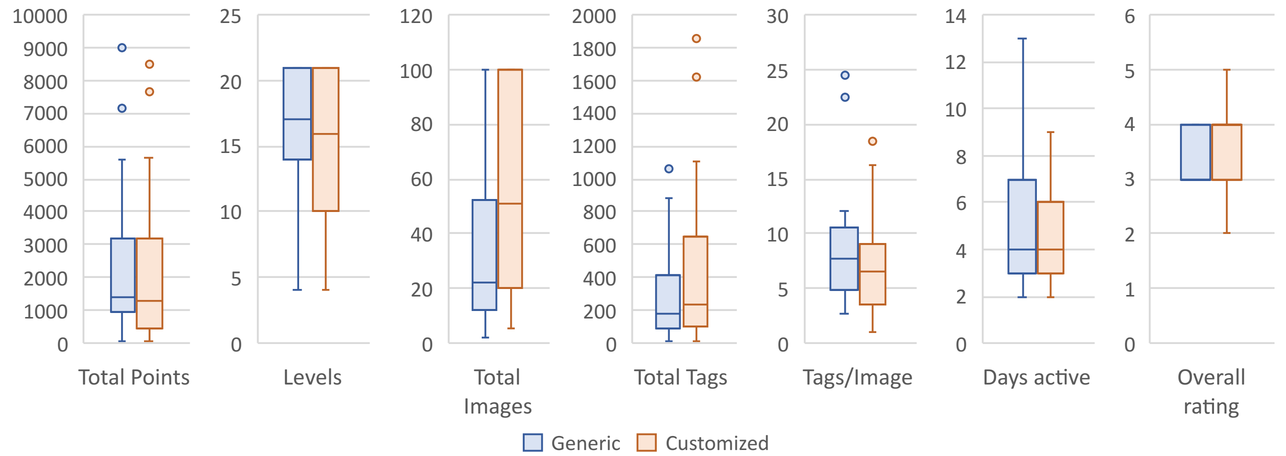


FIGURE 5 | Differences in performance and engagement variables between conditions (Study 2).

TABLE 7 | Comparison of performance and engagement variables and IMI scores between conditions (Study 2).

Variables	**Median (generic)**	**Median (customized)**	**U**	**Z**	***p***	**r**
Total Points	1622.0	1359.0	341.0	0.407	0.684	0.029
Level	18.0	16.0	331.0	0.588	0.556	0.042
Total images	25.0	51.0	258.5	1.854	0.064	0.132
Total tags	216.0	242.0	323.5	0.709	0.478	0.050
Tags per image	7.8	6.7	287.0	1.341	0.180	0.095
Days active	4.0	5.0	361.5	0.053	0.958	0.004
Experience rating	3.0	4.0	195.0	2.506	**0.012**	**0.179**
IMI Scores	**Median (generic)**	**Median (customized)**	**U**	**Z**	***p***	**r**
Interest	4.33	4.50	321.5	0.745	0.456	0.053
Competence	4.50	5.25	232.5	2.291	**0.022**	**0.163**
Effort	4.00	4.25	354.0	0.182	0.855	0.013
Pressure	2.25	2.50	344.5	0.347	0.728	0.025
Choice	5.75	6.00	274.0	1.574	0.116	0.112
Value	4.00	4.50	303.5	1.059	0.289	0.075
Relatedness	3.50	3.33	337.5	0.468	0.640	0.034

N = 54 (27 per condition).
Bolded values are significant at the 0.05 level.
Overall rating is a 5-point scale (range: 1–5).
***U*/*Z*/*p*: results of the Mann–Whitney U-tests.**
***r*: effect sizes, calculated as $r = Z \div \sqrt{N}$.**
*The absolute values of **Z** and **r** are displayed for improved readability.*

Another insightful observation is that Challenges and Power-ups were mentioned more often than any other element as the preferred and most influential elements, and as the second/third most motivating elements by participants in the generic condition. However, they were mentioned less often by participants in the customized condition, to the point that they are not mentioned more often than some of the other elements. In particular, Power-ups showed an accentuated decline.

On the other hand, Levels was only the third more cited element as preferred and most influential in the generic condition, but it appears as the sole element most often mentioned as preferred, most influential, and most motivating by participants in the customized condition. It was also the element selected more often by participants in the customization: 84 times. Similarly, Leaderboards, and Chance also received more interest by being mentioned more often as preferred,

TABLE 8 | Comparison of preferred, most influential, and most motivating elements per condition.

Elements	Generic			Customized			
	Pref	Infl	Mot	Sel	Pref	Infl	Mot
Levels	20	21	32	84	37	37	38
Moderating role	2	6	1	27	2	8	0
Badges	8	6	7	66	10	9	10
Unlockable content	2	2	1	48	2	2	2
Leaderboards	13	11	11	61	17	17	22
Challenges	29	32	28	55	23	16	14
Chance	2	5	2	47	11	13	14
Power-ups	35	28	30	69	14	14	14
N/A	15	15	14	–	10	10	12

N = 252 (126 per condition).

Sel, Number of times that each element was selected by participants in the customization step. Pref, Number of times that each element was listed as the participant's preferred element (Q6 in the end survey). Infl, Number of times that each element was listed as the participant's most influential element (Q7 in the end survey). Mot, Number of times that each element was listed as the participant's most motivating element (Q8 in the end survey).

most influential, and most motivating by participants in the customized condition than in the generic condition.

There were no significant relationships between participants' preferred element, most influential element, and most motivating element with their user type scores, personality traits, age (Kruskal–Wallis *H*-test), and gender (Pearson's chi-square test). However, with a sample of 252 participants distributed across eight gameful design elements, and some of the elements being mentioned very few times (e.g., moderating role and unlockable content), the sample was probably not large enough to detect any relationship.

4.4. Thematic Analysis

In this subsection, we examine participants' responses to the open-ended questions in our post-study survey. Three of the questions (**Q1**, **Q2**, and **Q3**) were meant to just obtain the general impressions about the use of the platform from participants in both conditions. The goal of this part of the analysis is to better understand the context in which participants' experience with the application occurred.

On the other hand, two questions specifically asked participants if the elements they selected matched their preferences and how they influenced the enjoyment of the task (**Q4** and **Q5**). While these questions make more sense in the customized condition, we also analyzed participants' responses in the generic condition to understand their experience. By having all the elements available to them, participants in the generic condition had to select elements for their experience by just deciding when to interact with them and when to ignore them, i.e., just by shifting their attention focus. Differently, participants in the customized condition were allowed to pre-select the elements they wanted to use, so their user interface was cleaner because only the selected elements were shown. The goal of this part of the analysis is to understand how participants experienced the customization and how their experiences differed by having all elements available to them (control condition) or being able to pre-select the desired elements (experimental condition).

These analyses were carried out by the first author using thematic analysis. The focus of our analysis was to identify themes that represented recurrent answers to the open questions answered by participants in the end survey. For example, **Q1** is "Overall, how do you describe your experience with the image classification activities you just completed?" Therefore, we focused our analysis in summarizing the themes frequently used by participants to describe their experience. Our analysis procedure was similar to reflexive thematic analysis (Braun and Clarke, 2006, 2019). Thus, the coding process was flexible, without a code book, and carried out by a single researcher. The process consisted on four steps: (1) *familiarization with the data*, i.e., an initial reading to become familiar with the content, (2) *coding*, i.e., labeling each participant's response with words extracted from the content of their answer, (3) *theme generation*, i.e., summarizing the themes from the codes that appeared more frequently, and (4) *writing up*, i.e., reporting the identified themes along with quotes from participants. These steps were carried out separately for each question in the survey (**Q1–Q5**). We combined the data from both studies for the analyses.

In the remainder of this subsection, we also present selected quotes from participants' responses to illustrate the identified themes.

4.4.1. Overall Experience

In response to **Q1**, some participants mentioned that they enjoyed their experience with the applications, but others did not. Participants who enjoyed the experience mentioned that it was fun, unique, easy, and interesting. Some specifically mentioned that the game elements contributed to making the experience fun or unique, whereas others mentioned that the photos were enjoyable, and some did not explicitly explain the reason for their enjoyment. For example:

> "I really enjoyed it more than I expected. The game elements captured my attention and made me want to do more of the tasks to earn more badges, complete challenges, etc." (P17, study 1, control condition)

> "It was interesting and enjoyable to come up with tags for the images. They were also nice photographs so it was fun to look at them." (P191, study 1, experimental condition)

Participants who reported less positive experiences mentioned that the task was boring or difficult, they had trouble understanding some of the instructions, or they felt that the game elements were not useful, for example, because the in-game rewards could not be carried out to the real world.

> "It was very dull, there was no real tangible reward outside your gamification systems. Without some kind of bonus this felt very 'Meh'." (P97, study 1, control condition)
> "It started out interesting and a bit exciting, but got boring after the first dozen or so images." (P3, study 2, control condition)

Regarding the experience of selecting elements, responses to **Q2** in the control condition were varied, which was expected because those participants did not actually customize their elements. Some participants just mentioned that interacting with the game elements was enjoyable, others said that they were not interested in the game elements, and some participants said that they did not actually select any game element:

> "I thought the game activities added a benefit to the classification task. It made it more fun and interesting." (P2, study 1, control condition)
> "I did explore the various game elements, but none of them were very interesting to me. I made use of the power-ups and claimed the challenges, but was a bit weirded out by the gifts feature and didn't really care about the levels, badges, or leaderboard. Also there were so many different elements that it was a bit confusing/hard to keep track of, so I mostly just stuck with the actual tagging." (P3, study 2, control condition)
> "I did not really do much in the way of customizing besides the avatar." (P103, study 1, control condition)

On the other hand, participants in the experimental condition did actually select game elements and so were able to explicitly comment about this experience. Participants said that the customization was easy, that they felt in control, and they tried to select the elements that matched their style or would help them in the task. Some participants enjoyed the possibility of customization because it is generally not offered or because they recognize that people may have different preferences. For example:

> "They were akin to filters on a shopping website in that I could choose the data that was most important/relevant to me and what I wanted to best assist me in my assessment of my progress." (P24, study 1, experimental condition)
> "I felt like I had control and like what I was doing mattered." (P44, study 1, experimental condition)
> "It was interesting because not many games allow you to do this." (P89, study 1, experimental condition)
> "It's a good idea, everyone can choose what they prefer, so every can play and be motivated with something they are interesting in." (P40, study 2, experimental condition)

However, there were also some participants who disliked the customization because it was not necessary or did not add much to their experience, they felt that the description of the elements was not enough for an informed choice, or that the application should allow them to modify their initial selection.

> "I thought it wasn't really necessary. I always try my best." (P10, study 1, experimental condition)
> "A bit arbitrary and there was little information given for each choice. I went in blind and I was stuck with what I chose." (P4, study 2, experimental condition)

With regards to the game elements offered by the system (**Q3**), participants who were satisfied mentioned that the elements made the task more fun or gameful, that they were varied enough, they were easy to choose, and provided a personalized experience. For example:

> "I was very satisfied. I felt like there was a good variety of options that I was familiar with. I liked some and disliked others, so I liked that I was able to pick." (P14, study 1, experimental condition)
> "I was more than satisfied by all the game elements provided. I knew that I could take any one of them and make the game more fun, but having more than one to choose from made it even more exciting." (P37, study 1, control condition)
> "Yes, lots of variety to cater to different personalities and improve user experience." (P53, study 2, experimental condition)

Some participants also reported not paying attention to the game elements, not interacting with them, or just feeling that they did not change anything. It seems that these participants had no specific issue with the offered elements, they just preferred to focus on the image classification task and were not interested in using the game elements. For example:

> "None of them make the task more interesting. The points mean nothing." (P35, study 1, experimental condition)
> "It really did not change anything for me." (P82, study 1, control condition)
> "Neutral, because I didn't use them." (P48, study 2, experimental condition)

4.4.2. Preference Matching and Task Enjoyment

When asked if they were able to select game elements that matched their preferences (**Q4**), some participants in the control condition responded that they could not select anything, which was to be expected as it was really the case. Some participants also mentioned that they were not aware of or did not understand what the game elements were. Echoing some of the responses in the previous subsection, there were also some participants who just did not care about the game elements or did not have any preference. But it is also interesting to note that some participants felt that they could select elements just because they could take a look at all of them and choose the ones they wanted to use and those they wanted to ignore. Other participants interpreted the ability to use some elements (for example, activating a power-up) as if it was an ability to select the game elements they wanted, which is understandable because they were not given a mechanism to better customize their experience like participants in the experimental condition.

> "No, I was just given game elements that would be in place with no options to choose." (P17, study 1, control condition)
> "Sort of. I played around with a whole bunch of them and they were all available to me as far as I could tell. My preferences are to unlock things which were available, so I would say the preferences were met." (P168, study 1, control condition)
> "Yes, it was mostly easy to ignore the ones I didn't care about (except the moderation feature, selecting yes/no for other people's tags, which got kind of annoying after a while since it popped up after each image)." (P3, study 2, control condition)
> "Didn't have a strong feeling with game elements. So no preferences really. I think it might be because that these techniques have been used too many times in a lot of applications, so people (or at least me) learn to ignore this and get to the core." (P8, study 2, control condition)

As expected, participants in the experimental (customized) condition responded more specifically about the task of selecting the game elements in the customization interface. Most participants said they were satisfied with the task of selecting game elements, mentioning that they were able to choose the elements that they preferred or that they thought would motivate them more. Only a few participants said that they did not appreciate the customization task because they would prefer to focus on the image classification task. Specifically, some participants on study 1 said they wanted to just classify the images and avoid interacting with the game elements so they would not decrease their hourly earnings. Logically, this reason did not appear on study 2 as they were participating voluntarily, not for payment like the Mechanical Turk workers from study 1.

> "Yes. I didn't want to examine other people's work, so it was nice that we had choices. If I was doing this long term, the game elements I chose would have added something to the activity." (P26, study 1, experimental condition)
> "Yes, I was able to find and select game elements that matched my preferences that would motivate me." (P78, study 1, experimental condition)
> "Not really—the only thing I really cared about was increasing my hourly earnings." (P91, study 1, experimental condition)
> "Yes because you can choose among a large set of game elements so you can easily find the one(s) that suit(s) you the best." (P42, study 2, experimental condition)

Finally, we asked participants if their selection of game elements influenced their enjoyment of the image classification task (**Q5**). A few participants in the control condition said that the game elements made the experience more enjoyable to them, but they did not relate this effect to the possibility of a customized experience, which was expected as they did not have a choice. However, many participants said that the game elements did not influence their enjoyment of the task. Explanations for this fact suggest that the task was already enjoyable enough without the game elements, or it was boring and the game elements could not change this fact.

> "The game elements made this a lot more enjoyable than a simple image classification task. I could see doing this for fun in my spare time." (P28, study 1, control condition)
> "I don't know if it influenced it too much. I was content doing the task without much customization, although I didn't explore it too deeply. I think if I had it would have become more enjoyable." (P55, study 1, control condition)
> "It didn't really. The task would've been the same without them." (P197, study 1, control condition)
> "Not that much. I mean, of course getting one badge made me feel accomplished and want to collect as many of them as possible but I did enjoy simply tagging the images without any gaming elements." (P5, study 2, control condition)

Responses from participants in the experimental condition generally followed the same themes, with some participants mentioning that the game elements made the experience more enjoyable, whereas others said that they did not make much difference. We were particularly interested in how participants felt that having customized their experience influenced their enjoyment; however, only a few participants specifically mentioned this aspect. Those who did said that customizing the game elements helped shape their experience and made them feel in control, or allowed them to choose their own goals or rewards.

> "I felt like I had control over the game." (P13, study 1, experimental condition)
> "I feel that my selection was important and really shaped my experience. I was motivated by the star rewards." (P51, study 1, experimental condition)
> "The selection of game elements allowed me to make the image classification suit my needs. It allowed me to make the classification more enjoyable and try to earn the highest score." (P77, study 1, experimental condition)
> "I don't think so because the task itself remained the same." (P89, study 1, experimental condition)
> "I think being able to choose rewards for myself made them more meaningful, choosing the elements that made me want to keep on going. Achieving those levels/badges/leaderboard spots/etc because I had decided that was the cool thing in this game made it more interesting than if all of those elements had been hardcoded and set for me by the game masters." (P2. study 2, experimental condition)
> "Not at all, I kind of forgot the game elements were there." (P14, study 2, experimental condition)
> "Like most people, I enjoyed being rewarded for my progress which allowed me to set specific goals and I felt accomplished when I was able to reach them. The game elements allowed me to be a little competitive with myself which is a good motivator for me." (P22, study 2, experimental condition)

5. DISCUSSION

5.1. Influence of User Characteristics on Element Selection

After analyzing the relationships between Hexad user type scores and gameful element selections, we found eight significant ones. From these, five were expected according to **H1** (Achiever-Badges, Achiever-Challenges, Achiever-Chance, Player-Leaderboards, and Player-Challenges); one was not expected, but is clearly understandable considering the description of the user type (Player-Power-ups); and two were not expected and cannot be easily explained (Philanthropist-Badges and Free Spirit-Chance). As reported in section 4.1, these results partially support **H1**, i.e., some of the expected differences in user type scores between participants who selected or not each game element were observed, but not all of them.

These results further support previous statements (such as Tondello et al., 2017a; Hallifax et al., 2019; Tondello, 2019) about the suitability of the Hexad user types as an adequate model of user preferences for the selection of gameful design elements in personalized gamification. Therefore, our work adds to the existing evidence that users with higher scores in specific user types are more likely to select specific game elements when given the choice, according to the eight pairs of user types and gameful elements listed above. By extension, we can assume that other relationships between user types and gameful elements proposed in the literature but not tested in this study may likely also hold true when tested in practice.

This contribution is important because the literature had relied so far on survey studies with only self-reported answers to establish relationships between Hexad user types and gameful design elements. Thus, the question remained if users would behave in an actual gameful system like they stated in their self-reported responses. The present work is the first one, to the best of our knowledge, to answer this question by demonstrating that participants' behavior (selection of gameful design elements) indeed correspond to their self-reported Hexad user type scores. While previous studies had compared two types of self-reported measures (user type scores and hypothetical game element preferences), we compared a self-reported measure (user type scores) with participant's actual behavior (their choice of game elements). This reinforces the confidence of gamification designers when using personalized gameful design methods that rely on selecting gameful design elements based on user types (such as Marczewski, 2018; Mora Carreño, 2018; Tondello, 2019).

On the other hand, some relationships between user types and gameful elements that were expected were not significant in this study (Philanthropist-Levels, Philanthropist-Moderating role, Socializer-Leaderboards, and Disruptor-Challenges). We believe that this happened because the context of the task was not favorable to create the type of experience that these users would enjoy. For example, the way that moderating role was implemented in our application did not seem very engaging as very few participants selected and enjoyed it; the leaderboard may have looked underwhelming because it was a very short experience and participants did not know and interact with each other. Better designs for these elements might have led to a higher appreciation by these participants. Additional studies will need to better evaluate these relationships.

Our results differ from those of Lessel et al. (2018) because they were not able to observe clear relationships between Hexad user types and gameful design elements like we did. But in their study, they asked participants to consider a few scenarios and try to design a gameful system for each one, which they thought they would enjoy. Although it provided many insights about how participants approached this task of designing a gameful experience for themselves, we believe that it speaks more about their capacity as designers than users because the designs were not implemented and tested. In contrast, our study allowed participants to actually use the gameful design elements, effectively testing how well each element worked for each participant.

Regarding the relationship between personality traits and gameful design elements, we found five significant ones (Agreeableness-Badges, Agreeableness-Chance, Openness-Badges, Opennes-Leaderboards, and Opennes-Chance). However, none of them were expected according to previous research or are not explained by the available literature. Therefore, **H2** was not supported. These results mirror previous literature, which also noticed inconsistent results when analyzing gameful design element preferences by personality traits (such as Tondello et al., 2017a; Lessel et al., 2018; Hallifax et al., 2019). Due to these variations in results across studies, it is hard to suggest how gamification designers could use this information in their practice. Therefore, we echo the existing literature in arguing that the Hexad user types are a better model for user preferences in personalized gamification than the Big-5 personality traits.

Finally, we found only one significant relationship between participants' age and their gameful element choices (moderating role was generally selected by younger participants) and none between gender and element choices. Thus, **H3** was only partially supported and **H4** was not supported. It is not clear why the differences identified in the existing literature were not observed in this study. More research will be needed to specifically try to observe in practice these different preferences by age and gender identified in the previous survey studies.

In summary, our response to **RQ1** "If allowed to choose the gameful design elements they prefer, do user choices correspond to the theoretical relationships with user types, personality, gender, and age reported in previous survey-based studies?" is that we found evidence that user choices do indeed correspond to their Hexad user type scores as reported in previous studies, at least partially. However, clear correspondences between element choices and participants' personalities, genders, and ages were not observed.

5.2. Task Performance and User Engagement

The results showed a significant improvement on the number of images tagged per participant in the experimental condition in study 1. Thus, **H5** was partially supported in study 1, but it was not supported in study 2. Additionally, results showed a higher rating for the experience of selecting game elements in both studies. Thus, **H6** was partially supported in both studies. However, participants spent approximately the same amount of time and wrote approximately the same number of tags for all images in both conditions. In addition, participants on study 1 did not want to lower their hourly rate of earnings in the Mechanical Turk platform, so they compensated the incentive to tag more images by writing less tags per image in the experimental condition. This effect was not observed on study 2, as the number of tags per image was not significantly different between conditions.

Therefore, it seems that personalization encouraged participants to achieve a higher task performance by classifying each image faster in order to complete more images in total. In a real application, this could be what designers wanted or not. In our application, this can be easily understood as a result of our design. Our application gave participants 10 points for each image classified and one additional point for each tag written for the image. It is reasonable to assume that participants quickly realized that they could earn more points by classifying more different images instead of spending time writing additional tags for the same image. If instead the design goal was to have participants adding more tags for each image, we could modify the design so that more points would be awarded for additional tags and less points for each classified image. We suppose that the performance change would have occurred in the opposite direction then, i.e., that participants would have classified less images, but provided more tags for each image.

This is evidence that personalization or customization can lead to higher task performance than generic gamification. Nonetheless, the design and incentives of the system must be well adjusted by the designers to achieve the intended goal. Our results showed that performance increased for the activity that was better rewarded by the system (classifying more images), even by perhaps decreasing the performance of other elements of the activity (e.g., adding more tags for each image). However, this should not be understood as an issue of personalization; it is just important to realize that personalization may not be able to automatically improve performance in all aspects of the task. It is part of the designer's job to fine tune the mechanics of gameplay to incentivize better performance where it is more important.

The intrinsic motivation measures did not differ significantly between conditions, except that perceived challenge was higher for participants in the customized condition on study 2. Looking at participants' free-text responses summarized in the thematic analysis, it is clear that some participants were already intrinsically motivated by the task and said that the game elements were not needed, whereas others said that they were bored by the task and the game elements could not change it. Considering this, it seems that the observed effects on task performance and engagement due to personalization did not occur because of changes in participants' intrinsic motivation. Therefore, future studies could consider different engagement or experience measures instead of the IMI to try and identify what are the mediators of these effects.

These findings are consistent with the evidence by Lessel et al. (2017, 2019), in which task performance was also higher for personalized than generic gameful systems. Even though our study is not the first to demonstrate the positive effects of personalized gamification for task performance, evidence of these effects is still scarce and additional studies are still needed to reinforce the preliminary findings. Our work contributes with additional empirical evidence of performance improvement with personalized gamification on an application context that is similar to that of Lessel et al. (image classification), but with a different application design and study design.

The analysis of participants' qualitative answers showed that the customization task was generally well received. However, designers should note that some participants asked for better descriptions of the game elements, for the possibility of changing the initial selection, or disabling all the game elements entirely. These are all features that should be included in the design of a customized gameful application. Moreover, Lessel et al. (2019) had already suggested that offering the possibility of disabling all the game elements may be desirable for some users, which is supported by some of the free-text answers from our participants.

In summary, our response to **RQ2** "Are user engagement and performance better for a personalized gameful system than a generic system?" is yes, user engagement and performance can be improved by adopting a personalized instead of a generic gamification design. However, designers must pay attention to clearly incentivize the behaviors that they want to improve in the gameful system, as providing more incentives for one type of behavior can lead to increased performance for that behavior in detriment of performance for different behaviors. Nonetheless, these findings are important because they demonstrate that it is worthy investing in personalized gameful design, which is undoubtedly more complex than generic gameful design, because it can lead to better achievement of the goals of the gameful system.

5.3. Participants' Perceived Usefulness of Each Element

The results from the analysis of participants' preferred, most influential, and most motivating elements suggest that users may perceive and experience some gameful design elements differently depending on whether they selected those elements themselves, or had no choice. Also considering the findings by Lessel et al. (2018), we can also suppose that participants would similarly experience elements differently if they were designing a system instead of just using a system previously built for them. This suggests an interesting line of investigation for future work because so far the relationships between user types and gameful design elements have been presented as universal. Future studies could investigate if the differences in user perception of each gameful design element depending whether they are designing, customizing, or just using the elements without modification can be replicated and mapped.

It is also noteworthy that we found no relationship between participants' preferred elements and their Hexad user type scores, even though there were relationships between those scores and the frequency of selection of specific game elements. In line with the comment above, it may be that the user type scores are currently better in capturing users' desire and intention regarding the use of specific game elements, rather than their perceived preferences after actually using the elements. It is possible that other factors may be in play during the actual user experience with the elements. For example, there are multiple ways of designing and implementing the same game element and Mora Carreño (2018, chapter 3) suggested that different designs can make each element more or less appealing for different user types. This is a question that requires more studies in future work.

5.4. Limitations and Future Work

Our study provided valuable findings about the correspondence between user types and gameful design elements in participant preferences, as well as the potential effect of personalized gamification on task performance. However, it was limited to one application context, which was image classification. We expect that similar results will be observed in different contexts and with different types of tasks, but this must be verified in future work. Therefore, we plan to conduct additional studies replacing image classification with different types of tasks.

Furthermore, we evaluated task performance considering only the number of tagged images and tags, but not the quality of tags. In future studies, it would be interesting to also consider tag quality by evaluating if the tags provided by participants corresponded to the presented images, to confirm that the quality of the tags remained the same or improved together with the improvement in the number of tagged images.

Additionally, participants in the first study were all Mechanical Turk workers residing in the United States of America. On the other hand, the second study had a more varied participation, with similar results to the first one, which suggest that the findings can probably be replicated with more diverse samples. Nonetheless, the difference in the number of images classified between conditions was significant in the first but not in the second study, despite a similar median difference. We believe that this was due to the smaller sample size in the second study. However, **Table 2** showed a few differences in the mean user type and personality trait scores between the two data sets. These differences may also have had any influence in the different results between the two studies. However, testing if the user type or personality trait scores would moderate the performance increase in **H5** was not one of the goals of this study. Therefore, we plan to carry out additional studies with participants from different countries to verify if our findings are similar for people with different cultural backgrounds. These additional studies may also test if demographic variables, such as user types, personality traits, age, and gender, may act as moderators of the performance difference between participants using a generic or a customized gameful application.

Finally, the personality traits inventory used in this study (Rammstedt and John, 2007) is very short, with just two items per trait. Although it has been validated and used frequently in HCI studies, its reliability is lower than longer scales, as the α values in **Table 2** show. This can have contributed to the inconsistent results in our analysis of the relationship between personality trait scores and element preferences. Thus, we plan to conduct additional studies using longer and more reliable personality trait scales to obtain more consistent results in the future.

6. CONCLUSION

In the present work, we showed that participants' choice of gameful design elements in a customizable gameful application partly corresponded to their Hexad user type scores, as predicted by models previously established from survey-based studies. This is the first study to demonstrate these relationships based on the actual observation of participants' experiences with a gameful application. This shows that personalized gameful design methods based on the selection of gameful design elements by user types can work in practice as suggested in the current literature.

On the other hand, these significant relationships were of weak effect sizes. Additionally, participants' user type scores were not related to their preferred, most influential, or most motivational game elements after they had interacted with the platform. This suggests that gameful designers can use the Hexad user types as one of the factors for personalization, but not the only one. There are yet other factors to be discovered in future work to determine with more precision what the preferences of a specific user will be in a gameful system.

Moreover, participants achieved a higher task performance and a better experience of selecting which game elements to use in a customizable version of our gameful application than a generic version with the same gameful design elements. These results show that personalization or customization of gameful design elements is a viable solution to increase task performance and improve the user experience. Nonetheless, the design of our application encouraged users to improve the number of images classified without at the same time improving the number of tags per image. This means that personalization may be more effective in increasing user behaviors that are more explicitly incentivized, and not necessarily all user behaviors in the application. This is something that designers should take in consideration when creating any gameful system, and especially personalized ones.

This contribution is valuable to the HCI and gamification communities because several personalized gameful design methods have been recently suggested in the literature. Our work shows that they are a promising approach to improve the design of gameful applications and make them more successful in achieving their goals.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Office of Research Ethics, University of Waterloo.

The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

GT designed the study, developed the digital application, collected and analyzed data, and wrote the manuscript draft. LN revised and approved the study design. All authors contributed to manuscript revision, read, edited, and approved the submitted version.

FUNDING

This work was supported by the CNPq Brazil, SSHRC (895-2011-1014, IMMERSe), NSERC Discovery (RGPIN-2018-06576), NSERC CREATE SWaGUR, and CFI (35819).

ACKNOWLEDGMENTS

The authors would like to thank GT's thesis committee (Daniel Vogel, Mark Hancock, Edith Law, James Wallace, and Juho Hamari) for the feedback on prior work leading to this contribution and Stuart Hallifax for helping with data collection.

REFERENCES

Adomavicius, G., and Tuzhilin, A. (2005). Personalization technologies: a process-oriented perspective. *Commun. ACM* 48, 83–90. doi: 10.1145/1089107.1089109

Altmeyer, M., Lessel, P., and Krüger, A. (2016). "Expense control: a gamified, semi-automated, crowd-based approach for receipt capturing," in *Proceedings of the 21st International Conference on Intelligent User Interfaces - IUI '16* (Sonoma, CA: ACM), 31–42. doi: 10.1145/2856767.2856790

Araújo Paiva, R. O., Bittencourt, I. I., Da Silva, A. P., Isotani, S., and Jaques, P. (2015). "Improving pedagogical recommendations by classifying students according to their interactional behavior in a gamified learning environment," in *Proceedings of the ACM Symposium on Applied Computing - SAC '15* (Salamanca: ACM), 233–238. doi: 10.1145/2695664.2695874

Bakkes, S., Tan, C. T., and Pisan, Y. (2012). "Personalised gaming,?? in *Proceedings of The 8th Australasian Conference on Interactive Entertainment Playing the System - IE '12* (Auckland: ACM), 1–10. doi: 10.1145/2336727.2336731

Barata, G., Gama, S., Jorge, J., and Gonçalves, D. (2017). Studying student differentiation in gamified education: a long-term study. *Comput. Hum. Behav*. 71, 550–585. doi: 10.1016/j.chb.2016.08.049

Bartle, R. (1996). Hearts, clubs, diamonds, spades: players who suit MUDs. *J. MUD Res*. 1.

Böckle, M., Micheel, I., Bick, M., and Novak, J. (2018). "A design framework for adaptive gamification applications," in *Proceedings of the 51st Hawaii International Conference on System Sciences (HICSS)* (Manoa: University of Hawaii), 1227–1236. doi: 10.24251/HICSS.2018.151

Böckle, M., Novak, J., and Bick, M. (2017). "Towards adaptive gamification: a synthesis of current developments," in *Proceedings of the 25th European Conference on Information Systems (ECIS)* (Guimarães).

Bouzidi, R., Nicola, A. D., Nader, F., Chalal, R., and Laboratoire, M. (2019). "A systematic literature review of gamification design," in *20th annual European GAME-ON Conference (GAME-ON'2019)* (Breda), 89–93.

Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. *Qual. Res. Psychol*. 3, 77–101. doi: 10.1191/1478088706qp063oa

Braun, V., and Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qual. Res. Sport Exerc. Health* 11, 589–597. doi: 10.1080/2159676X.2019.1628806

Buhrmester, M., Talaifar, S., and Gosling, S. D. (2018). An evaluation of Amazon's mechanical turk, its rapid rise, and its effective use. *Perspect. Psychol. Sci*. 13, 149–154. doi: 10.1177/1745691617706516

Butler, C. (2014). "A framework for evaluating the effectiveness of gamification techniques by personality type," in *HCI in Business. HCIB 2014. Lecture Notes in Computer Science, Vol. 8527*, ed F. Nah (Cham: Springer), 381–389. doi: 10.1007/978-3-319-07293-7_37

Chou, Y.-K. (2015). *Actionable Gamification - Beyond Points, Badges, and Leaderboards*. Octalysis Media, Kindle Edition.

Codish, D., and Ravid, G. (2017). "Gender moderation in gamification: does one size fit all?," in *Proceedings of the 50th Hawaii International Conference on System Sciences - HICSS 2017* (Waikoloa Village, HI), 2006–2015. doi: 10.24251/HICSS.2017.244

Costa, P. T. Jr, and McCrae, R. R. (1998). "Trait theories of personality," in *Advanced Personality*, eds D. F. Barone, M. Hersen, and B. Van Hasselt (Boston, MA: Springer), 103–121. doi: 10.1007/978-1-4419-8580-4_5

Deterding, S., Dixon, D., Khaled, R., and Nacke, L. E. (2011). "From game design elements to gamefulness: defining "gamification"," in *Proceedings of the 15th International Academic MindTrek Conference* (Tampere: ACM), 9–15. doi: 10.1145/2181037.2181040

Field, A. (2009). *Discovering Statistics Using SPSS, 3rd Edn*. London: Sage Publications.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *Am. Psychol*. 48, 26–34. doi: 10.1037/0003-066X.48.1.26

Hallifax, S., Serna, A., Marty, J.-C., Lavoué, G., and Lavoué, E. (2019). "Factors to consider for tailored gamification," in *Proceedings of the 2019 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '19* (Barcelona: ACM), 559–572. doi: 10.1145/3311350.3347167

Herbert, B., Charles, D., Moore, A., and Charles, T. (2014). "An investigation of gamification typologies for enhancing learner motivation," in *2014 International Conference on Interactive Technologies and Games (iTAG 2014)* (Nottingham, UK), 71–78. doi: 10.1109/iTAG.2014.17

Jia, Y., Xu, B., Karanam, Y., and Voida, S. (2016). "Personality-targeted gamification: a survey study on personality traits and motivational affordances," in *Proceedings of the 34th Annual ACM Conference on Human Factors in Computing Systems - CHI '16* (San Jose, CA), 2001–2013. doi: 10.1145/2858036.2858515

Kaptein, M., Markopoulos, P., De Ruyter, B., and Aarts, E. (2015). Personalizing persuasive technologies: explicit and implicit personalization using persuasion profiles. *Int. J. Hum. Comput. Stud*. 77, 38–51. doi: 10.1016/j.ijhcs.2015.01.004

Klock, A. C. T., Pimenta, M. S., and Gasparini, I. (2018). "A systematic mapping of the customization of game elements in gamified systems," in *XVII Brazilian Symposium on Computer Games and Digital Entertainment (SBGames 2018)* (Foz do Iguaçu).

Koivisto, J., and Hamari, J. (2019). The rise of motivational information systems: a review of gamification research. *Int. J. Inform. Manage*. 45, 191–210. doi: 10.1016/j.ijinfomgt.2018.10.013

Landers, R. N., Auer, E. M., Collmus, A. B., and Armstrong, M. B. (2018). Gamification science, its history and future: definitions and a research agenda. *Simul. Gaming* 49, 315–337. doi: 10.1177/1046878118774385

Landers, R. N., Bauer, K. N., and Callan, R. C. (2017). Gamification of task performance with leaderboards: a goal setting experiment. *Comput. Hum. Behav*. 71, 508–515. doi: 10.1016/j.chb.2015.08.008

Lessel, P., Altmeyer, M., and Krüger, A. (2018). "Users as game designers: analyzing gamification concepts in a "bottom-up" setting," in *Proceedings of the 22nd International Academic Mindtrek Conference - Mindtrek '18* (Tampere: ACM), 1–10. doi: 10.1145/3275116.3275118

Lessel, P., Altmeyer, M., Müller, M., Wolff, C., and Krüger, A. (2016). ""Don't whip me with your games": Investigating "Bottom-Up" gamification," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16* (San Jose, CA: ACM), 2026–2037. doi: 10.1145/2858036.2858463

Lessel, P., Altmeyer, M., Müller, M., Wolff, C., and Krüger, A. (2017). "Measuring the effect of "bottom-up" gamification in a microtask setting," in *Proceedings of the 21st International Academic Mindtrek Conference - AcademicMindtrek '17* (Tampere: ACM), 63–72. doi: 10.1145/3131085.3131086

Lessel, P., Altmeyer, M., Schmeer, L. V., and Krüger, A. (2019). ""Enable or Disable Gamification?": Analyzing the Impact of Choice in a Gamified Image Tagging Task," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI 2019)* (Glasgow: ACM), 150. doi: 10.1145/3290605.3300380

Marczewski, A. (2018). *Even Ninja Monkeys Like to Play: Unicorn Edition*. Gamified.

McAuley, E., Duncan, T., and Tammen, V. V. (1989). Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: a confirmatory factor analysis. *Res. Quart. Exerc. Sport* 60, 48–58. doi: 10.1080/02701367.1989.10607413

Monterrat, B., Desmarais, M., Lavoué, É., and George, S. (2015). "A player model for adaptive gamification in learning environments," in *Artificial Intelligence in Education. AIED 2015. Lecture Notes in Computer Science, Vol. 9112* (Cham: Springer), 297–306. doi: 10.1007/978-3-319-19773-9_30

Mora Carre no, A. (2018). *A framework for agile design of personalized gamification services* (Ph.D. Dissertation). Universitat Oberta de Catalunya, Barcelona, Spain.

Mora, A., Tondello, G. F., Calvet, L., González, C., Arnedo-Moreno, J., and Nacke, L. E. (2019). "The quest for a better tailoring of gameful design: an analysis of player type preferences," in *Proceedings of the XX International Conference on Human Computer Interaction - Interacción '19* (Donostia Gipuzkoa: ACM), 1–8. doi: 10.1145/3335595.3335625

Mora, A., Tondello, G. F., Nacke, L. E., and Arnedo-Moreno, J. (2018). "Effect of personalized gameful design on student engagement," in *Proceedings of the IEEE Global Engineering Education Conference - EDUCON 2018* (Tenerife). doi: 10.1109/EDUCON.2018.8363471

Nacke, L. E., Bateman, C., and Mandryk, R. L. (2014). BrainHex: a neurobiological gamer typology survey. *Entertain. Comput.* 5, 55–62. doi: 10.1016/j.entcom.2013.06.002

Nacke, L. E., and Deterding, S. (2017). The maturing of gamification research. *Comput. Hum. Behav.* 71, 450–454. doi: 10.1016/j.chb.2016.11.062

Nov, O., and Arazy, O. (2013). "Personality-targeted design: theory, experimental procedure, and preliminary results," in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work - CSCW '13* (San Antonio, TX: ACM), 977–984. doi: 10.1145/2441776.2441887

Orji, R., Mandryk, R. L., Vassileva, J., and Gerling, K. M. (2013). "Tailoring persuasive health games to gamer type," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13* (Paris), 2467–2476. doi: 10.1145/2470654.2481341

Orji, R., and Moffatt, K. (2018). Persuasive technology for health and wellness: state-of-the-art and emerging trends. *Health Inform. J.* 24, 66–91. doi: 10.1177/1460458216650979

Orji, R., Oyibo, K., and Tondello, G. F. (2017). "A comparison of system-controlled and user-controlled personalization approaches," in *UMAP 2017 - Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization* (Bratislava), 413–418. doi: 10.1145/3099023.3099116

Orji, R., Tondello, G. F., and Nacke, L. E. (2018). "Personalizing persuasive strategies in gameful systems to gamification user types," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '18* (Montreal, QC: ACM), 435. doi: 10.1145/3173574.3174009

Orji, R., Vassileva, J., and Mandryk, R. L. (2014). Modeling the efficacy of persuasive strategies for different gamer types in serious games for health. *User Model. User-Adapt. Interact.* 24, 453–498. doi: 10.1007/s11257-014-9149-8

Rammstedt, B., and John, O. P. (2007). Measuring personality in one minute or less: a 10-item short version of the Big Five Inventory in English and German. *J. Res. Pers.* 41, 203–212. doi: 10.1016/j.jrp.2006.02.001

Rapp, A., Hopfgartner, F., Hamari, J., Linehan, C., and Cena, F. (2019). Strengthening gamification studies: current trends and future opportunities of gamification research. *Int. J. Hum. Comput. Stud.* 127, 1–6. doi: 10.1016/j.ijhcs.2018.11.007

Roosta, F., Taghiyareh, F., and Mosharraf, M. (2016). "Personalization of gamification elements in an e-learning environment based on learners' motivation," in *2016 8th International Symposium on Telecommunications (IST)* (Tehran), 637–642. doi: 10.1109/ISTEL.2016.7881899

Seaborn, K., and Fels, D. I. (2015). Gamification in theory and action: a survey. *Int. J. Hum. Comput. Stud.* 74, 14–31. doi: 10.1016/j.ijhcs.2014.09.006

Sundar, S. S., and Marathe, S. S. (2010). Personalization versus customization: the importance of agency, privacy, and power usage. *Hum. Commun. Res.* 36, 298–322. doi: 10.1111/j.1468-2958.2010.01377.x

Tondello, G. F. (2019). *Dynamic personalization of gameful interactive systems* (Ph.D. thesis). University of Waterloo, Waterloo, ON, Canada.

Tondello, G. F., Kappen, D. L., Ganaba, M., and Nacke, L. E. (2019a). "Gameful design heuristics: a gamification inspection tool," in *Human-Computer Interaction. Perspectives on Design. Proceedings of HCI International 2019. LNCS 11566* (Springer), 224–244. doi: 10.1007/978-3-030-22646-6_16

Tondello, G. F., Kappen, D. L., Mekler, E. D., Ganaba, M., and Nacke, L. E. (2016a). "Heuristic evaluation for gameful design," in *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Extended Abstracts - CHI PLAY EA '16*, 315–323. doi: 10.1145/2968120.2987729

Tondello, G. F., Mora, A., Marczewski, A., and Nacke, L. E. (2019b). Empirical validation of the gamification user types hexad scale in English and Spanish. *Int. J. Hum. Comput. Stud.* 127, 95–111. doi: 10.1016/j.ijhcs.2018.10.002

Tondello, G. F., Mora, A., and Nacke, L. E. (2017a). "Elements of gameful design emerging from user preferences," in *Proceedings of the 2017 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17* (Amsterdam: ACM), 129–142. doi: 10.1145/3116595.3116627

Tondello, G. F., Orji, R., and Nacke, L. E. (2017b). "Recommender systems for personalized gamification," in *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization - UMAP'17* (Bratislava: ACM), 425–430. doi: 10.1145/3099023.3099114

Tondello, G. F., Wehbe, R. R., Diamond, L., Busch, M., Marczewski, A., and Nacke, L. E. (2016b). "The gamification user types hexad scale," in *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16* (Austin, TX: ACM), 229–243. doi: 10.1145/2967934.2968082

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Computer Science

ORIGINAL RESEARCH
published: 17 September 2020
doi: 10.3389/fcomp.2020.00040

A Taxonomy of Coping Strategies and Discriminatory Stressors in Digital Gaming

Cale J. Passmore and Regan L. Mandryk*

The Interaction Lab University of Saskatchewan, Department of Computer Science, Saskatoon, SK, Canada

Digital gaming's many benefits starkly contradict its well-cited toxicity. To accurately understand and compare how players cope with discriminatory stress in the context of play, 241 US players were surveyed on recurring sources of discrimination during gameplay and strategies for coping across ranging experiential prompts. Qualitative analysis created a taxonomy of discriminatory *targets*, discriminatory *acts*, and coping *strategies* specific to online digital play. We compare experiences, perceptions, and beliefs around coping across intersections of race, gender, and class (with notes on ability and age) and describe how player identities inform in-game behavior and exposure to types of discrimination and how coping strategies are navigated. We discuss the accumulative, anticipatory, and intergenerational nature of discriminatory stress in gaming, its stratified effects on well-being, and the role of discrimination in belief formation as well as ability to advocate for oneself and others.

Keywords: discrimination, stress, coping, ethnicity, gender, class, intersectional, digital game

OPEN ACCESS

Edited by:
Z. O. Toups,
New Mexico State University,
United States

Reviewed by:
Carlos Vaz De Carvalho,
Polytechnic Institute of Porto, Portugal
Elisa Mekler,
Aalto University, Finland

****Correspondence:***
Cale J. Passmore
cale.passmore@usask.ca

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Computer Science

Received: *01 February 2020*
Accepted: *05 August 2020*
Published: *17 September 2020*

Citation:
Passmore CJ and Mandryk RL (2020) A Taxonomy of Coping Strategies and Discriminatory Stressors in Digital Gaming. Front. Comput. Sci. 2:40. doi: 10.3389/fcomp.2020.00040

INTRODUCTION

> Before considering how to conceptualize, measure, and quantify health consequences of discrimination, one caveat immediately is in order: the purpose of studying health effects of discrimination is not to prove that oppression is "bad" because it harms health. Unjustly denying people fair treatment, abrogating human rights, and constraining possibilities for living fully expressed, dignified, and loving lives is, by definition, wrong—regardless of effects on health. (Krieger, 1999, p. 296)

In 2019, approximately two-thirds of the global online population play digital games on consoles, computers, and mobile devices (Wijman, 2018). As a form of play, gaming's unique appeal transcends many gendered, cultural, ethnic, national, ability, and socio-economic divides. This nearly universal endorsement is largely due to gaming's social (Koivisto and Hamari, 2014; Domahidi et al., 2018), cognitive (Baniqued et al., 2013; Oei and Patterson, 2013; Granic et al., 2014), and affective (Olson, 2010; Boyle et al., 2012; Dennis and O'Toole, 2014) benefits, combined with its ability to cater to diverse ways to play (Kafai et al., 2010; Gibbons, 2015). Digital gaming is often pursued as a leisure activity so that players can experience enjoyment, escapism, immersion, and challenge (e.g., De Grove et al., 2016; Abeele et al., 2020, but gaming's benefits exceed those of solely a pastime of pleasure. Playing games provides benefits to well-being by helping players recover from daily stressors (Reinecke, 2009), repair noxious moods (Bowman and Tamborini, 2015), build self-esteem (Bessière et al., 2007), promote mindfulness (Collins et al., 2019), combat loneliness (Depping et al., 2018), cope with life's challenges (Iacovides and Mekler, 2019), and practice emotional regulation (Villani et al., 2018).

Drawing an audience more diverse than other leisure activities (Media Entertainment, 2015), discrimination based on sexual orientation, gender, race, ethnicity, ability, and age is high in digital game contexts (Williams et al., 2009; De Schutter and Vanden Abeele, 2010; Kafai et al., 2010; Burgess et al., 2011; Shaw and Friesem, 2016; Behm-Morawitz, 2017; Edström, 2018; Shaw et al., 2019; Vella et al., 2020): 76% of women and non-binary digital game players experience sexism or genderism (McDaniel, 2016), rates of homophobia and transphobia vastly outweigh positive LGBTQ+ game content (Shaw et al., 2019), 92% of gamers feel that online platforms make others more critical and negative (Citrona, 2014), and systematic misrepresentation of race and ethnicity spans character design and game content, with players describing racism, tokenism, minstrelsy, and absence as norms in gaming (Shaw, 2012; Dietrich, 2013; Behm-Morawitz, 2017; Passmore et al., 2018). Game producers and players alike continue to struggle against norms that pander to gaming's stereotypical audience as young, able-bodied, Anglo-white, heterosexual men (Shaw, 2012). Over the last decade, player diversity has risen, yet studies show declining representation in industry and game content (IGDA, 2014; Passmore et al., 2017), with increasing reports of hate speech (Sengün et al., 2019). Thus, while most youth turn to online media and digital games as a means for coping with the challenges of life (Rideout et al., 2011), black, indigenous, people of color (BIPOC) spend 4.5 more hours per day consuming online media that leaves them more exposed to oppressive content than traditional "offline" sources (Tynes et al., 2014).

Exposure to recurring, systemic discriminatory stressors (e.g., ableism, sexism) requires strategies for long-term management. These strategies do not nullify the effects of discrimination, rather, as Krieger (1999), Luthar (2006), and Pascoe and Richman (2009) show, discrimination has extensive short- and long-term effects on *mental health* (e.g., lower self-esteem and life satisfaction; higher rates of depression, anxiety, and post-traumatic stress disorder), *physical health* (e.g., higher blood pressure, chronic pain), and *behavior* (e.g., higher aggression, self-harm; sleeplessness). These impacts starkly contrast gaming's recorded benefits.

Studies of discrimination commonly reference coping across a variety of digital contexts; however, few center coping in their study design. Of those that do, Gray's works (Gray, 2012a,b, 2018) and Ortiz (2019) provide unparalleled insight to how black players experience, manage, and resist heteropatriarchal White supremacist norms in gaming. Two studies focus on the coping strategies of (predominantly) White women (Fox and Tang, 2017; McLean and Griffiths, 2019), while Vella et al. (2020) show that women cope with targeted misogyny through manipulating their online appearance or "masking." The exceptional depth of these studies is owed to their restriction to specific intersections of player identities, affording for an ecologically valid account of plurality in experiences. However, such depth necessarily limits the breadth needed for comparisons *between* intersections of player identities. Best practices advise comparing such focused experiences across demographics, which requires a comprehensive taxonomy of context-specific strategies (Krieger, 1999; Skinner et al., 2003). Such a taxonomy has been absent from digital gaming literature.

The reasons for this gap are many. Stress management is often habitual (i.e., experientially suppressed) (Chen et al., 2016; Brosschot et al., 2017), making accurate elicitation of self-reported data notoriously difficult (Petitmengin, 2006). The wide range in how people describe both discrimination and coping as well as how we categorize that data underlies why studies identify 400+ coping strategies with no agreed-upon reduction (Skinner et al., 2003). Historically, studies of coping can lack ecological validity because they fail to account for systemic and historical relationships between stress, coping, and socially stratified identities or power dynamics (Krieger, 1999). This is to say nothing of the sample size and coding effort required for comparisons between demographic intersections of race, ethnicity, gender, class, ability, and so forth. Establishing a systemically accurate, context-specific, demographically comparative taxonomy of discriminatory sources and coping strategies is a formidable task.

Designing to account for these complexities, we conduct a qualitative study of coping with discrimination in digital games: its forms, frequencies, and effects. Thematic analysis constructs taxonomies for common *targets* of discriminatory stress, types of discriminatory *acts*, and coping *strategies* unique to digital gameplay across the compounding dynamics of race, gender, and class (with some notes on ability and age). We conclude by discussing the nature of discriminatory stress as an accumulative, persistent, anticipatory stressor biased toward feedback cycles of social inequity and describe their effects on behaviors, beliefs, and resilience in digital play.

Discrimination Across the Contexts Virtual and Analog

Discrimination is a threat against one's inherent legitimacy and agency along social categories of identity (e.g., axes of ethnicity, culture, gender, ability, age, nationality, *etc.*) (Crenshaw, 1991; Berjot and Gillet, 2011). A unique source of adverse stress, discrimination targets the most effective defense against its adverse effects: a positively affiliated sense of self (Luthar, 2006; Nadal et al., 2011; Bird, 2013; Romero et al., 2014; Umaña-Taylor et al., 2015). The social construction of axes of identity makes discrimination inseparable from historical norms, from social power dynamics. Anyone's identity or agency may come under threat; however, the mental, physical, emotional, and social effects of systemic threat are markedly different—significantly more harmful—for those culturally and/or historically marginalized (Krieger, 1999; Balsam et al., 2011; Chief Moon-Riley, 2017). Fundamentally accumulative, new discriminatory events combine with prior experiences common to marginalization such as intergenerational trauma, additional barriers to material conditions, and physical and psychosocial violence.

Direct sources of discrimination in online gaming run the gamut of discrimination's usual suspects. Cited forms include slurs, epithets, targeted threats, stereotypes, and targeted harassment or exclusion from other players, developers, and games themselves (Gray, 2012b; Fox and Tang, 2017; McLean and Griffiths, 2019; Ortiz, 2019). *Indirect sources* stem from systemic, historical sources, evidenced in the under- and misrepresentation of BIPOC (Kafai et al., 2010; Dietrich, 2013; Passmore et al., 2018; Srauy, 2019), women and non-binary players (Williams et al., 2009; Shaw, 2015; Behm-Morawitz, 2017), LGBTQIA2+ players (Gray, 2018; Shaw et al., 2019), disabled players (Gibbons, 2015; Holloway et al., 2019), and older players (De Schutter and Vanden Abeele, 2010). Indirect forms of discrimination reported by players extend from storylines to colonial, heterosexist, and/or racist game worlds and character choices; oversights in playtesting; inequal access to the time or technology to game; and a range of biases in developer hiring practices, determining whose perspectives are included in games, and restricted affordances for player interactions (Mukherjee, 2018; Passmore et al., 2018; Holloway et al., 2019; Spiel et al., 2019; Srauy, 2019).

Daily discrimination is unavoidable for 69% of Americans (American Psychological Association, 2016). The ubiquitously pervasive nature of identity violence thus requires strategies for management rather than avoidance (Anisman and Merali, 1999; Moghaddam et al., 2002; Brondolo et al., 2009b). While some do not game due to harassment or material inequity (McDaniel, 2016; Rankin and Han, 2019), most adopt strategies to reduce its impact. Players hide their racial and gendered axes through avatar and username selection, masking their digital self-representations to avoid harassment (Gray, 2012a,b; Fox and Tang, 2017; Ortiz, 2019; Vella et al., 2020). Players withdraw from online socialization altogether, forgoing chat, microphone use, and tools for gaming's social benefits (McDaniel, 2016; Fox and Tang, 2017; McLean and Griffiths, 2019; Vella et al., 2020). Players with non-Euro-American accents and/or neuro-physical atypicalities employ similar strategies to control their self-disclosure (Passmore et al., 2018; Ortiz, 2019; Rankin and Han, 2019). BIPOC players normalize near constant racial epithets, minstrelsy, and tokenization (Leonard, 2006; Gray, 2018; Passmore et al., 2018; Ortiz, 2019). Youth, older, and/or disabled players are discouraged from gameplay due to ability-restrictive interfaces, game mechanics, and exclusionary research practices (De Schutter and Vanden Abeele, 2010; Spiel et al., 2019). The need for relief is greater among these groups, yet discrimination complicates even these highly modifiable avenues for coping with life.

Intersectionality and Plurality

Regardless of how one *identifies*, people are *identified* through whatever representations are available: avatar skin tones, voices, slang, usernames, *etc.* (Kafai et al., 2007; Williams et al., 2009; Passmore and Mandryk, 2018). How one is identified often determines how they are treated. Where marginalized identity axes are concerned, experiences of discrimination are non-linear. Coined by Crenshaw (1991), *intersectionality* refers to the specific ways people are disempowered across compounding facets of identity. For example, LGBTQ2S+ women of color are exposed to significantly more discrimination than non-LGBT+ women of color (Balsam et al., 2011). Such experiences differ in the qualities and the quantities of violence experienced by women of color with disabilities. Intergenerational transmissions of trauma notwithstanding (Bird, 2013; Chief Moon-Riley, 2017), health outcomes differ significantly between intersections of identity not only due to increased rates of exposure to discrimination along one axis or another but also due to the unique ways marginalized axes compound (Krieger, 1999; Balsam et al., 2011).

Within gaming studies, Rankin (Rankin and Han, 2019), Gray (Gray, 2012b, 2018), Shaw (Shaw and Friesem, 2016), and Gibbons (Gibbons, 2015; Holloway et al., 2019) depict tensions between the benefits of online gaming, the costs of adapting to discriminatory violence, and the moments of successful strategies where systemic barriers are overcome. They demonstrate the plurality of player experiences at the ranging intersections of material inequalities and histories of stigma: how variable beliefs, perceptions, and experiences are even when analyses are limited to single demographic axes. Their work further supports Krieger's findings that the nuances of coping require qualitative methods of self-report (Krieger, 1999); quantitative generalizations often obscure these nuances in plurality, lacking ecological validity while encouraging demographic tokenization.

Decades of epidemiological research show health, well-being, social power, coping, and identity as inseparable (Krieger, 1999; Pascoe and Richman, 2009). Coping is culture- (Kuo, 2011), gender- (Szymanski and Henrichs-Beck, 2014), orientation- (Nadal et al., 2011), ethnicity- (Neal-Barnett and Crowther, 2000; Brondolo et al., 2009b), class- (Scott, 2004), education- (Lazarus and Folkman, 1984), and affiliation specific (Sellers et al., 1997, 2001). How one copes is determined by emotional responsivity (Pennebaker et al., 1988; Stanton et al., 1994), socio-historical contexts (Chief Moon-Riley, 2017; Mosley et al., 2017), awareness of privilege (Fujishiro, 2009; Black, 2016), novelty (Young et al., 2019), over-exposure (Miller et al., 2007; Brondolo et al., 2009a), beliefs surrounding both identity and what constitutes discrimination (Brondolo et al., 2009a; Dale et al., 2018), as well as individual preferences for coping strategy (Noh and Kaspar, 2003; Pascoe and Richman, 2009). Furthermore, coping is immediately contextual: how one copes with the stress of a sexist boss differs from coping with a sexist stranger or from coping with sexism in leisure (Walker et al., 1977; Bacchus, 2008; Szymanski and Henrichs-Beck, 2014).

Design Considerations for Stress and Coping

Quantitatively, marginalized identities compound in their exposure to violence and stress, with disabled non-binary Black and indigenous people of low income exposed to the highest rates of violence in the US. This does not mean that these groups *report* the highest frequencies of discriminatory experiences. For example, Greer (Greer et al., 2009) shows that African-American men more sensitively report experiences of discrimination than African-American women despite lower overall frequencies of exposure. Racism can be over-attributed to European-Americans (Burgess et al., 2011), and sexism is more attributed to men (Inman and Baron, 1996). Privilege awareness is often positively correlated with guilt, leading privileged *and* socially aware participants to over-report inequities (Black, 2016). Individual perspectives on what constitutes discrimination, personal life experiences, the relative novelty of exposure—what people experience and how they manage the stress of those experiences: each underlies who reports what types of experiences and to what degree.

Generally, the participants accurately report on their experiences when directly asked (Axt, 2011). As Krieger (1999), Lazarus (2000), and the above-mentioned authors show, self-reports where discriminatory stress are concerned can range greatly in their accuracy. This is largely due to how we adapt to high levels of chronic, systemic stress. Cortisol, the hormone responsible for initiating recovery from acute stress, accumulates when stressors (such as discrimination) occur with such frequency that the acute stress recovery response is incomplete when the next stress response is activated (Miller et al., 2007; Pascoe and Richman, 2009; American Psychological Association, 2016). Normalization (allostasis) is a coping response to this chronic saturation of cortisol due to interrupted recovery, lowering one's overall baseline for activation over time (Schulkin et al., 1998; Miller et al., 2007; Young et al., 2019). Generally, the experiential intensity of an acute stress response can be "dampened" at a cost of maintaining a higher baseline of stress (Alvarez and Juang, 2010; American Psychological Association, 2016). Being "used to it" or normalizing discriminatory stress, however, does not mitigate its long-term consequences to health (Pascoe and Richman, 2009; Karlamangla et al., 2013; Young et al., 2019), from higher risk of illness and neurological impairment (Miller et al., 2007; Treadway et al., 2019) to social reclusion (Willner, 1997; Riles et al., 2019). Resilience to stress declines over time due to deactivation of dopamine receptors (Treadway et al., 2019), tying chronic stress to lower motivation, impulse control, decision-making, focus, and effort discounting (Gassen et al., 2019; Treadway et al., 2019). Worse still, as an adverse, accumulative, chronic, and intergenerationally transmitted health factor, the immediate absence of discriminatory events does not necessarily indicate an absence of discriminatory stress (Miller et al., 2007; Mathur et al., 2016).

Considering the "toxic" norms of discrimination in gaming, stigmatized players appear substantially disadvantaged when gaming for relief. To better understand the extent, nature, and degree to which discrimination affects player experiences, the benefits they reap from gaming, and how these factors influence game behaviors across a spectrum of player identities, we build on Gray's, Ortiz's, and Fox and Tang's foundations. Acknowledging our breath necessarily lacks the depth of their studies, we attempt to shore up ecological validity by accounting for the factors and dynamics above.

METHODS

Background and Frameworks

To design with as much control over these factors, we reviewed literature on discrimination and coping across gender, ethnicity, age, disability, social class, and cross-cultural histories of modeling coping. These pre-study efforts helped inform (and limit) our questionnaire design. We integrated this knowledge into previous design standards for conducting research with marginalized groups in HCI (Passmore et al., 2018) informed by Critical Race Theory (Delgado et al., 2001; Finda Ogbonnaya-Ogburu et al., 2020), Identity-Based Motivation Theory (Oyserman, 2008), historical materialist epistemologies, and phenomenological elicitation. Patricia Hill Collins (Hill Collins, 2002), Helen Cixous (Sellers, 2003), Frantz Fanon (Fanon et al., 2004), Peggy McIntosh (McIntosh, 2003), Dean Spade (Spade, 2015), and Audre Lorde (Lorde, 2012) inform the theoretical background and language used in survey to ensure a shared, preliminary understanding of stratified experiences. Petitmengin (2006) and Giorgi (2010) inform design considerations for eliciting experiential self-reports, namely, how to use question order, word choice, and reflective prompts to prime participants, how to focus them on the experiential (rather than ideological) aspects of those experiences, and how to do so without biasing (Trnka and Smelik, 2020) their responses. Our analysis is deeply indebted to and influenced by intersectional frameworks; however, as non-Black settler researchers, we lack the situatedness required to employ it. Thus, our analysis is limited to a more general view of "compounding" (rather than intersecting) axes of identity.

Survey Design

Gathering accurate data for the purpose of comparing a wide range of experiences and degrees of privilege required, we design our survey iteratively, co-constructing questions with players of varied ethnicities, genders, socioeconomic and educational backgrounds, disabilities, and ages. Extensive pre-testing of question wording, descriptions, question types, and survey order was imperative to ensure that the data gathered under priming were sensitive, accurate, and non-leading. For example, after a battery of introspective demographic and gaming experience questions, we asked the participants to "select any (of the following) systemic source(s) of oppression you experience while gaming." Knowing that participants vary in their familiarity with, say, "classism," examples were given to *cue* the participants (e.g., "I experience relative poverty or constantly struggle with the cost of life") earlier in the survey. Pre-tests established that the participants who did not relate to class struggles overlooked these examples, opting to describe other phenomena in open fields later in the survey. Those who did relate often described highly detailed, direct experiences of class-based discrimination related to their gaming experiences. Techniques like these maintain a social-identity-centered focus, engender trust in *our* identification of systemic oppression, and, by providing a large range of questions and prompts, help mitigate numerous response and measurement biases (Trnka and Smelik, 2020).

Data were gathered across several axes in open and closed form. Appended for this survey's purposes, the sections include:

(i) Identity measures: With discrimination linked to identity and our focus on compounding axes of identity, we gathered substantial demographic information. After briefing the participants, the survey opened with the request asking them to self-describe. This allowed the categorization of participants on aspects of their identity that they felt were important. Specific identifiers of gender, age, household income, social class, education, sexuality, disability, gaming availability and habits, ethnicity, and generation were gathered (but not required). With a prior work identifying that the absence of representation is experienced as discrimination but often not labeled as such (Passmore et al., 2018), the participants were asked to describe instances (if any) where they related to or identified with game characters/worlds and why. This was partly for gathering data on indirect discrimination and partly to prime the participants to reflect on their gaming relationships.

(ii) Discrimination in digital games: Following Krieger (1999), we first presented check-all-that-apply questions about foci of systemic oppression in gaming contexts (e.g., racism, colorism, body-shaming, sexism). Prompts preceded open fields by asking the participants to describe instances of recurring discrimination in detail. Sources, situation reports, accompanying feelings, and emotional and behavioral reactions to these forms of discrimination (if any) were requested.

(iii) Debrief: Debriefing instructions, contacts for professional aid, and researcher contacts were provided, as was an open field for overlooked factors, comments, and survey feedback.

Sampling

Ethical approval was obtained from the University of Saskatchewan Research Ethics Board. Demonstrated as a reliable and validated platform for gathering representative US samples (Kittur et al., 2008; Mason and Suri, 2012), the participants were recruited through Amazon's Mechanical Turk (MTurk). To facilitate diversity in the participants, we released a pre-survey to gather demographic information and invited people from underrepresented groups to complete our full survey. Participation required an informed consent, mandating a minimum participant age of 18. The participants were paid $3.50 USD compensation for completion of the 20-min survey. They were informed that their identities would remain confidential and that no deception was involved but that they may leave contact information for follow-up or study release. Data were collected over 2 days and resulted in 241 total responses. The time spent per question was evaluated to screen for attentiveness to each question.

Data Analyses

We conducted both between- and within-group analyses per best practice for studies on perceptions and experiences of demographic groups (Cokley, 2007; Phinney and Ong, 2007).

Thematic analysis proceeded as per Braun (Braun and Clarke, 2006) and was conducted in SPSS 25.0. We closely integrated and followed best practices for analysis and construction of coping hierarchies as per Skinner et al. (2003). In addition to the SPSS dataset, a reflexivity journal was kept in all phases of analysis to track interpreter presumptions, codes, themes, and organizing families and to monitor analyst biases due to expectations.

Approach I (inductive, thematic): Each open-ended question was separated from other data and analyzed independently. Recurring experiential units, keywords, thematic trends, and proximal semantic units were recorded. An identical second round of analysis was conducted 1 week from the previous round, having bracketed prior results and randomized question response order. Lower-order codes, potential themes, and organizing categories (primary strategy, secondary strategy, *etc.*) were recorded and then compared to the first round's constructs for similarities and robustness. The results were grouped into "item pools" (Skinner et al., 2003) according to conceptual similarity and combined when differences in descriptions *and* codes were merely lexical.

Approach II (inductive, organizational): The participants' descriptions varied greatly in length and detail, with some participants describing multiple coping strategies for multiple forms of discrimination; proportion reporting, however, demanded that these experiences be segmented into units prior to thematic assignment. Following segmentation, themes were constructed and attached to each experiential "unit" from each description and then compared to the results from Approach I. From this comparison, a near-final draft of codes and themes was constructed and then organized into "families" (Skinner et al., 2003). The results were input to SPSS as new variables. A second round of this approach was conducted several days later, using an unmarked copy of the dataset in SPSS, and then compared to the first to test for consistent assignment of descriptions to codes and themes.

Approach III (deductive, verifying): With codes, themes, and organizing families finalized, code and category assignment took place cross-survey. Each participant's set of responses was treated as a case (considered in the context of *all* their other responses) and analyst interpretations were limited to assigning previously identified codes and themes. This "in-context" analysis constructed several new themes and another test of code unidimensionality (Bandalos, 2002). Some code assignments were modified as in-context interpretation clarified description meanings. We later checked for errors with a final pass, and the results were quantitatively analyzed in SPSS to report proportions.

Themes, subthemes, and coding structure are discussed in the results. To avoid contamination of our context-specific findings, a comparison between our taxonomies and others was conducted only after coding was completed.

Sample Composition

Of the 241 responses, two "participants" were deemed bots and nine participant responses were removed for low effort (e.g., one-word responses or, in one case, trolling, as determined by inconsistent self-reported identifiers with highly racist and sexist responses and low effort). After applying exclusionary criteria, the sample ($n = 230$) was binned into demographic categories (e.g., race, gender, class). If self-described identification conflicted with a participant's demographic data, self-description determined categorization. Other than those who preferred not to identify, the participants self-described as: White ($n = 78$, 33.9%), *A*sian ($n = 42$, 18.3%), *B*lack ($n = 49$, 21.3%), *H*ispanic ($n = 41$, 17.8%), or *N*ative American ($n = 1$, 0.4%). A total of 18 participants identified as "multi-racial" without an exclusive preference for a racial category (Mixed, 7.8%), two women identified as trans (0.9%), and one participant identified as non-binary (*Nb*, 0.04%). There were 99 participants who identified as Women (43%) and 130 as Men (56.5%). Furthermore, 16 participants identified with Disabilities (7%), and 24 participants identified with *LGBTQ*I2+ (10.4%). From yearly household income after taxes, education, and self-described socio-economic status, 20 participants identified as upper class (*Uc*, 8.7%), 109 as middle class (*Mc*, 47.4%), and 101 as lower class (*Lc*, 43.9%). Age ranged from 18 to 58 years ($M = 33.5$, SD = 10.36). Five participants did not currently play digital games, one indicated no time to game, and all others averaged at least 1–10 h of gaming per week.

RESULTS

Exclusionary and Inclusionary Criteria

There were 63 participants who describe no recurring experiences of discrimination. Of those who did, 14 observed others' experiences but described no first-hand experiences. As we asked for "recurring" experiences during digital gameplay, descriptions of a single event and events outside digital gaming ($n = 1$) were excluded. Some mistook "being annoyed" as systemic discrimination ($n = 5$, 2.2%): "I (made) just a little bit of a mistake (in game) but I was scolded by many people," *WWoUc32*. All five were among the 17 (7.4%) participants highly dismissive of discrimination as an experience altogether, identifying it as "unimportant," imagined, or a "tactic": "There is no discrimination in video games. Not to me, nor to anyone who I've played with online in the past 17 years", *HMeLc32*. To ensure that the results were exclusive to recurring, first-hand experiences of discrimination, these 33 cases were excluded from coding for discriminatory sources and coping strategies. We included 23 participants citing "no experiences of discrimination" but who described systemic discrimination. They believed that their experiences were unique to them (rather than systemic), were "deserved," or were universal (experienced by "everyone").

Topology of Discrimination

The players were asked a check-all-that-apply question for "systemic source(s) of oppression you experience most often while gaming?," including racism or colorism (overall 28.3%; 42.8% of BIPOC participants), sexism or genderism (30.9% total; 56% of non-cis-Men), classism (7.8%; 18.6% lower class), ableism or neurotypicalism (1.7%; 25% disabled), colonialism or imperialism (3.5%), cultural biases (16.5%), nationalism or politicalism (12.2%), body-shaming or attraction biases (11.7%), and none (40%). Other sources included: religious ($n = 3$), prejudice against new players ($n = 3$), ageism ($n = 2$), and motherhood ($n = 1$). Some participants did not affiliate with systemic oppression here but described recurring experiences of systemic oppression (e.g., "ableism" was not selected but ableist discrimination was described). Many participants indicated recurring discrimination across multiple axes, but the descriptions commonly focused on one axis (often race or gender).

We asked the participants two open-ended questions to collect data on recurring experiences of discrimination during digital game play: "Describe a recurring situation that left you feeling particularly discriminated against, over-looked, or mis-represented from your experiences in digital gaming. Please describe the game, situation, and what about this experience left you feeling this way," and "In your own words, please describe how you cope with discrimination (if any) in video games." The participants describe multiple coping strategies relative to the source and the type of discrimination. Almost all descriptions of discrimination were accompanied with strategies for management and emotional states. Thus, the descriptions were segmented into four categories: the *target* of discrimination or axes of identity, the discriminatory *act*, events, or stimuli considered as discriminatory; the participants' *feeling* during and after these events; and their coping *strategies* or reactive management of discriminatory stress.

Targets

Target codes were almost exclusively demographic descriptors (i.e., race, gender, class, appearance, sexuality neuroatypicality, nationalism, ability, age). Skill (in-game performance) was a minor subtheme. Outside of ableism, almost half of the skill themes were co-present with ageism: "undue" judgment of older players' performance: "I wasn't as fast as some of the people in multiplayer. I'm older and not as well-tuned with the controllers as the younger guys," *AMeMc52Lgbt*. The class was exclusively described in reference to inaccessible technology (e.g., high-fidelity inputs), purchasable game assets (e.g., "skins"), or material constraints on time: "I can't be accepted in (multiplayer games) because I don't have a boat load of time to play as some people do," *WWoLc44Lgbt*.

The *targets* are determined by how players are identified, not necessarily how they identify themselves. Real *and* digital attributes such as avatar skin tone, accents considered as "ethnic," character features conveying "normalized gender," atypicality—conveying these was described as creating vulnerability and risk and, in some cases, "inviting" discrimination. The *targets* of identity in digital play are hierarchical and demographic and correspond to real-world power dynamics and inequities. A player's identity is *inferred* to mirror a digital signifier's stereotypical, socio-historical meaning.

An illustrative example can be seen in the participants' descriptions of "mistaken identities":

> I am a male, but when I created my online character for GTA Online, I made a female character since you couldn't play with a female in the story mode. While playing, I never spoke on the mic, but I noticed how other players would assume that I was a female and that my gaming skills would be "lesser than." For example, when playing with a group of people on a heist, they would designate me with the "easier" jobs. WMeMc29

Mistaken identity descriptions commonly include: (i) detachment, as the player did not identify with the intended *target*, (ii) newfound empathy with those affiliated with the intended *targets* of that specific discrimination, and (iii) the

TABLE 1 | Four superordinate themes for discriminatory acts, with subthemes for each.

Rendering invisible (i.e., erasing/ minimizing identity)	Rendering grotesque (i.e., distorting identity through hyperbole/ stereotype)	Conflict (i.e., direct, violent action)	Gatekeeping (i.e., barring authority/ access)
Belittling	Sexualizing	Trolling	Barring access
Diminishing	Objectifying	Doxing	Lowering others' expectations
Silencing	Misrepresenting	Arguing	Presuming inadequacy
Shunning	Stereotyping	Mocking	Subordination
Dismissing	Tokenizing	Harassing	
Gaslighting	Targeting		
Absent representation	Outing		
	Slurs		
	Casual racism		

harasser maintaining and often escalating tactics when corrected (e.g., harassers' "doubling down" on discriminatory actions). Consistent with literature on the proteus effect (Yee and Bailenson, 2007; Gutierrez et al., 2014), these experiences were described as "enlightening" for those inexperienced with identity violence. Mistaken identification and second-hand observers of discrimination—especially those relating past discrimination against their *targets* to others—share descriptions of feeling sympathy, frustration, sadness, and guilt. These descriptions are distinct from first-hand experiences: more intellectualized, less intense, and shorter-lasting.

Acts

Discriminatory *acts* were divided into four superordinate themes with multiple themes using the criteria of 4+ independent descriptions to constitute a theme (see **Table 1**). All types are sharing qualities of "threat" —to one's physical, social, mental, or emotional well-being—and exclusion on the sole basis of identity axes. Discriminatory *acts* are described as harmful, negative stimuli in the form of presence (e.g., slurs, gatekeeping, profiling, targeting) and absence (of representation, similar players, respect, *etc.*). *Acts* were mostly described as "frustrating" or "annoying," with less marginalized players exclusively citing "surprise." Frustration—or agitation with impeded purpose—was often proximal to beliefs around the ease of *acts* relative to the burden of its effects and/or the "superfluous" presence of *acts* despite their irrelevancy to gameplay, performance, or enjoyment: "(They) called me a "dumb white bitch," also told me "Go have your daddy **** you again." This was literally over a healing issue in a video game where a sunflower is a healer. Crazy to me," *WWoMc37*.

Conflict *acts* were less common but described as most threatening to the players' real-life safety. They are direct, recurring, and focused. Doxing, harassment, and arguing exposes vulnerable player information (home addresses, real names, social media accounts, *etc.*). Rendering invisible subthemes describes feelings of identity or agency minimization (sometimes to the extent of erasure); rendering grotesque themes describe distortions of identity through hyperbole or inaccurate magnification (e.g., minstrelsy). Players describe both negatively impacting social and personal legitimacy and/or self-worth on personal and social levels. Gatekeeping, the most indirect of *act* themes, was sometimes explicit but more often the intended result of other *acts*. Women and LGBT+ players frequently referenced gatekeeping and sexual harassment. Subthemes of rendering invisible or rendering grotesque were frequently described by players of color—especially women of color.

Supporting the compounding nature of *acts*, Hispanic players recurrently cited the lack of representation, women subthemes of sexualizing and harassment, and Hispanic women recurrently lacking representation *and* sexual harassment. Black-coded players regularly cited slurs and tokenizing, while White-coded players often cited no discrimination and denial or dismissal of discriminatory experiences. Those who spoke multiple languages tended to describe the risk of being "outed" or *targeted* due to their accent. High performance in game (mention of "winning") was described as inviting rendering invisible/grotesque *acts* (most frequently, slurs) and, when combined with being outed, conflict.

Topology of Coping Strategies

Strategies cover a range of described strategies for managing with discriminatory actions (see **Table 2**). Every described *act* was accompanied by descriptions of learned strategies for its management. Some are generalized (e.g., normalization), some are context specific (e.g., altar avatar), and some are *act*-specific (e.g., blocking sexual harassers). Multiple codes for each participant require that we present proportions for ethnicity and gender per superordinate theme. The proportions for *strategies* are presented also by factors of race and gender (see **Figure 1**).

Endure/ignore descriptions are players' primary *strategy* despite many deeming it largely "unsuccessful" —especially when *acts* tend to increase in directness and/or frequency. No descriptions (beyond those mistaken about what constitutes discrimination) show passively enduring/ignoring as a "solution" to *acts* and rarely as "successful." Lack of functionality in addressing *acts* frames this *strategy* as a desire more than a behavior. Seek social support themes, which include support inside and outside digital contexts, never include reliance on support from other players; descriptions always reference friends, family, or community. Mute self and other themes for modifying the digital self are described as a resort, never a preference, which is unreliable due to its infrequent availability (e.g., premised with "if the option exists") and therefore sees low frequencies.

The second most frequent *strategy*, modifying experience, depicts players who change frames for understanding the *act* in response to discrimination. This can mean denying *acts* any reaction ("I cannot and will not allow their feelings to have any level of control over me"), forming generalizations or prejudices against gamers ("people online are toxic"), devaluing "gaming" as meaningful or significant, or, most frequently, combining expectations of discrimination in the game with those from other contexts:

TABLE 2 | Seven superordinate themes for strategies for coping organized by avoid to approach (top to bottom).

	Superordinate theme (count)	Description	Subthemes
	Endure/ignore (106) Asian−48% Black−53% Hispanic−51% Mixed−22% White−45%	Attempts to ignore, tolerate, or passively not engage with the act	Endure/ignore Tune-out Focus elsewhere Be silent Quietly hope
Avoid	Modify the digital self (19) Asian−12% Black−6% Hispanic−10% Mixed−17% White−5%	Altering the digital representation of targets to avoid acts	Hide Alter character/avatar/username Avoid/disable chat Mute self
	Modify the digital environment (28) Asian−7% Black−18% Hispanic−12% Mixed−11% White−12%	Removing or limiting the source(s) of discrimination from digital space	Mute/block players Play only with Friends Appeal to authorities/report player Switch server/game world
	Modify the experience (74) Asian−45% Black−22% Hispanic−51% Mixed−11% White−27%	Cognitive reframing to reduce the acute power of acts	Normalize discrimination Rationalize discrimination Empathize with discriminator Devalue players/game/gaming
	Modify/dismiss self (68) Asian−48% Black−16% Hispanic−34% Mixed−33% White−26%	Engagement editing or changing one's personal values, beliefs, or goals	Narrow interests Vow to assert future self and take pride Minimize/dismiss feelings Join-in with discriminator Switch tasks/game Return to familiar game Go offline/cease playing
Approach	Seek social support (33) Asian−14% Black−20% Hispanic−10% Mixed−22% White−12%	Seeking or involving others for support	Engage with family Engage with Friends Vent or relate to others online Seeking "like-minded" players Seek bystander intervention
	Direct confrontation (50) Asian−7% Black−20% Hispanic−29% Mixed−56% White−19%	Active, aggressive engagement with discriminatory sources within social spaces	Call out discriminator Draw attention to discrimination Outperform discriminator (revenge) Harass or dox discriminator

The (count) is the number of codes for that strategy; the percentages are the proportion of participants identifying in that group who reported that code. The columns provide a description of the strategy and the subthemes that were coded as belonging to that strategy.

> I've never experienced anything beyond the normal vitriol commonly experienced. WMeLc58Lgbt

Black men and Hispanic or White women tended to modify their digital environments through *strategies* that remove their perception of discriminatory players (mute/block players), appeal to authorities (moderators, guild leaders), or report players. Women exclusively modify their digital self, with those of mixed or Asian coding tending to report more than the other groups. Black and Asian-coded women tended to not report engaging in direct confrontation compared to other groups (Hispanic and Black men and White women), citing the "futility" of it "changing anything" and risking the escalation of abuse. Normalize, rationalize, and empathize codes were never positive, conveying a somber, regretful necessity of acknowledging oppression as a norm "to be accepted." Black men were unlikely to report modify/dismiss self, but Asian participants and Hispanic men tended to report this approach.

> I basically try not to get to "wrapped up" and emotional about the situation. I realize games represent a microcosm of how people act in the real world. HMeLc28

Specific approach strategies are relative to a player's situational agency, beliefs, and inherited culture, as discussed by Fragoso

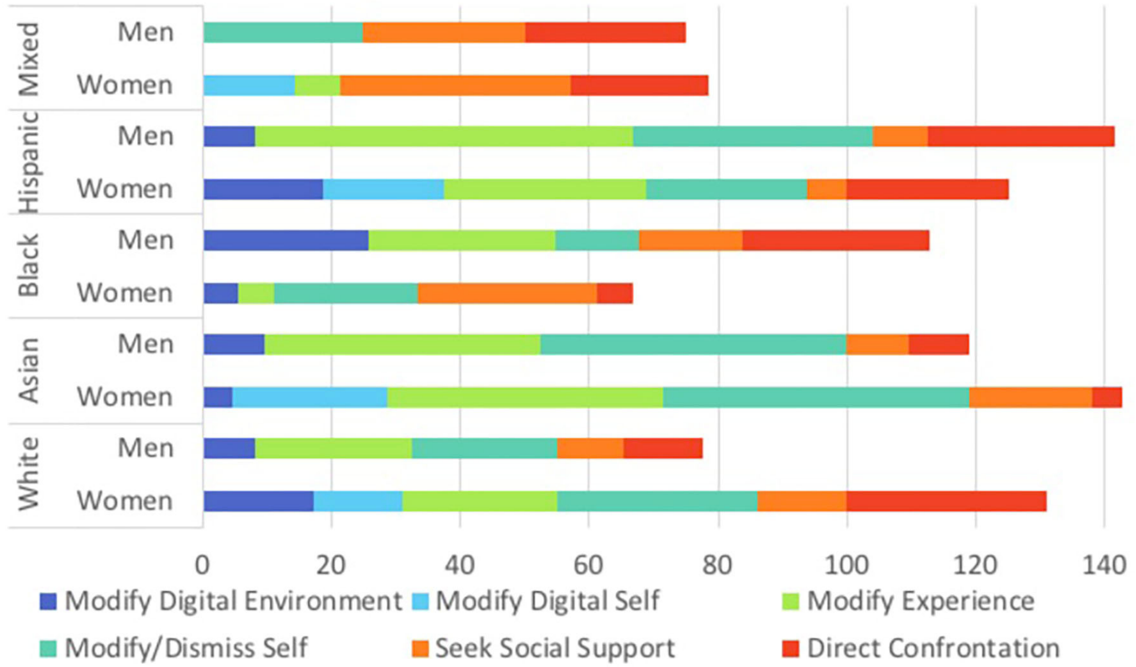


FIGURE 1 | Superordinate strategies for coping with discriminatory acts, by race and gender. Endure/ignore strategy is not included as it dominated the approaches (see **Table 2**) and was often the initial strategy reported. The bar length shows the percentage of participants, identifying by race and gender category, reporting each strategy. The counts exceed 100% as the participants reported multiple coping styles. The counts were normalized by the number of participants in each identity group, allowing inter- and intra-identity factor comparisons.

and Kashubeck (2000), Noh and Kaspar (2003) Yoo and Lee (2005) and Krieg and Xu (2018). For example, Asian men and women show low frequencies for direct confrontation compared to other groupings. Many Hispanic men reported modifying the experience of discrimination and Black-coded men tended to describe emphasis on pride and vows to assert future self and taking pride:

> I often see negative stereotypes about minorities in all aspects of life. It creates a problem, especially when trying to convey to my children about Black pride. BMeMc45

Strategies are complicated to parse given their intertwined, internalized, and anticipatory nature. LGBT+ players and women tended to cope with more acutely violent *acts* (e.g., doxing, harassment) by engaging in strategies of self-effacement (modifying/dismissing the self or digital self). Rarely is this through self-depreciation or joining-in with discriminators *via* self-directed humor; more often, players hide their identities [as in Gray (2012b), Fox and Tang (2017), Vella et al. (2020)], switch tasks/games, or alter their beliefs (narrow interests in games/genres, generalize, devalue the medium or players). As a result of this *strategy*, many players cite concerns around their *targets* being "outed" and/or "exposed," thus inviting subsequent *acts*. Outing commonly escalates in frequency and severity of rendering grotesque *acts*, culminating in more violent themes of conflict.

Seek revenge (outperforming the discriminator) is unique in its being both an *Act* and a *strategy*. These descriptions conveyed *gravitas*, a high-risk "gamble" of stereotype confirmation *combined* with risks of being outed for proving one's legitimacy through in-game performance. Some (mostly men) relished this gamble: "I'd target specifically them and kill just them in the most irritating ways possible," BMeMc44.

We coded for discrimination as *explicitly* normalized; 25% ($n = 42$) of participants who described first-hand recurring discrimination did so as a "given": a daily experience indistinguishable from discrimination experienced in other contexts of everyday life. The proportions for normalization codes generally match the sample demographics; however, those impacted across multiple *targets* more frequently described normalization.

Sequential Strategies for Coping

Players describe coping as relative to tools-at-hand. With in-game tools often unavailable, the chosen strategies depend on an initial assessment of their available resilience:

> You really have to pick your battles. AMeLc32

The coping strategies are tiered when *acts* persist. Lower-cost *strategies* are attempted and fail, or players with a higher vulnerability to identity violence anticipate discrimination from game spaces:

> At first, I will try to confront the problem head on. I will try to talk to the people committing the discriminatory behavior. If that doesn't work, then I will go to the game moderators (if there are any) and report the player. If that doesn't work, then I will just try to avoid communication with the player. BMeMc36D

Where discrimination persists after initial desires to endure/ignore, the players describe secondary and tertiary strategies. Avoid strategies are often subsequent to the failure of approach strategies mitigating discriminatory stress. Tertiary strategies (e.g., reverting to a prior game or going offline) often follow a significant accumulation of discriminatory stress. The most common sequence was endure/ignore, mute/block the source ("if possible"), and, if exposure persists, players hide, switch tasks/games. Players seek out previously "tried and true" games at this point, "even if I'm already bored of it," or "games where I have friend groups to laugh about these things with." Generally, modifying the digital self or environment is a second-to-last resort, proceeding failure of lower effort strategies. Seeking social support, going offline, or modifying the self were last resorts.

Compounding Privileges, Oppressions, and Pluralities

> It's alienating and it reminds me of how much discrimination still exists against Asians even when it doesn't manifest often in my daily life. AWoUc25

Levels of systemic privilege roughly correspond to both type and severity of *strategy*: Hispanic and Black men and Hispanic and White women without disabilities tended to report "approach" or "problem-focused" strategies (like direct confrontation); upper-class, able heterosexual White men reported few-to-no experiences of discrimination; upper-class, heterosexual White women and non-LGBT middle-class Black men tended to describe "fighting back:" high effort approach or interpersonally directed strategies. Black women, trans, non-binary, and disabled participants of lower class tended to describe "emotion-focused" or intrapersonally directed strategies when digital gaming. No players who describe experiencing discrimination also describe "waiting out" ableism, ageism, sexism, classism, or racism as a successful tactic, as without high cost, or describe a belief that wide-scale systemic stressors are addressable through game interactions. Be it explicit or implicit, marginalized players acknowledge the barriers to resolving sources of discrimination, opting to instead problem-solve *how to cope with the stressor's effects.* This contrasts more privileged participant descriptions of in-game stressors as problems with potentially direct resolution (ignoring infrequent slurs, blocking, *etc.*). Recognizing the difficulty of affecting systemic change, those who experience frequent discrimination across several *targets* tended to report "avoid" or inward-directed strategies to regain security, control, and agency.

Identity factors do not guarantee a player's experiences, beliefs, or values. Some highly privileged players report high frequencies and intensities of discrimination; some socially marginalized players describe few to no discriminatory experiences and hold oppressive beliefs:

> I've never felt discriminated against, but I'm not a millennial poofter or professional victim. WMeMc48Lgbtq

Generally, however, demographics—social identity factors replete with their socially stratified values and power—inform discriminatory stress exposure, amount, severity, tolerance, and coping strategies. Underlying the intertwined relationship between severity and type of *strategy*, tolerance for *acts*, self-reported marginalization, and number of compounding *targets* is a players' history of stress which they bring to the game context. Grouping players by total number of compounding *targets*, we observe similarities in their data. Mapped onto a u-shaped curve (**Figure 2**), with the x-axis containing the number of socially oppressed *targets* (from upper-class, able White men to lower-class LGBTQ+ Black women with disabilities), at each extreme we see low frequencies of self-reported discrimination and higher preferences for "avoid" coping strategies (e.g., endure/ignore). The midpoint represents players with the highest frequencies of self-reported discrimination, identification across one to two marginalized axes, and more frequent use of "approach" strategies (like direct confrontation). Labeling each fluid point on this spectrum "the privileged few," "the emboldened many," and "the conflict weary" is one way of descriptively representing the intertwined nature of *targets*, *acts*, and *strategies*. Using a descriptive spectrum rather than demographic labels better allows for plurality in experiences while resisting tokenization.

The Privileged Few

> The only "discrimination" I've seen in games is by the skilled against the unskilled. How you cope is you get better at the game. MMeMc50

Low in proportion and with few-to-no first-hand experiences of discrimination, players on the far left of this spectrum identify within privileged norms. Largely, but not exclusively, they are White middle-class heterosexual men who mistake "discrimination" in terms of players "being annoying" or "guilting" them over "hypersensitive" reactions. Discriminatory stress, defined by this group, is frustrating more than it is recalling of systemic oppression or trauma. There is confusion over the impact of oppression on agency. Thus, coping involves enduring/ignoring "annoyances," which is often deemed a successful strategy: "I do not deal with racial discrimination, but when I am met with negative situations during gaming, I just learn to live with it. That's the way I do not stress over it," WWoUc24.

Identifying with *status quo*, players here are skeptical of discrimination, assuming their experiences map universally onto others:

> None. Never happened. I could always make a character of color and use female voice during character creation. No one cares that I have a vagina, they just want to play the game. People forget that the Internet is the best anti-discrimination tool to date because

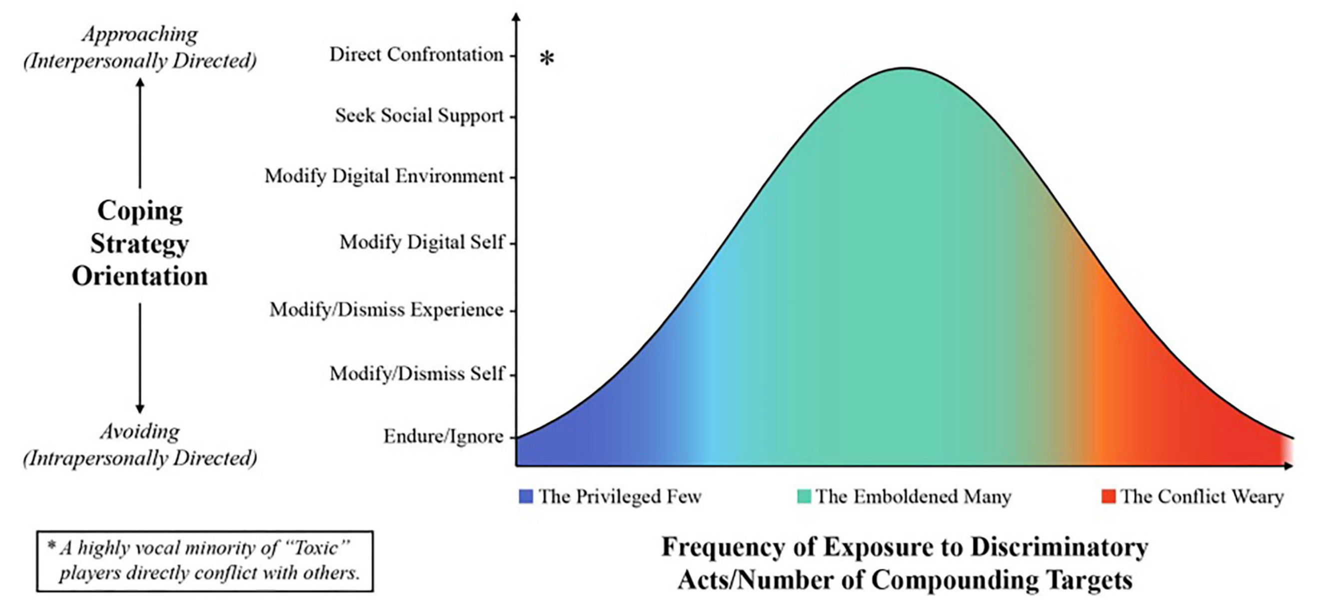


FIGURE 2 | *U*-shaped curve of the general relationships shared between coping strategy orientation and required effort (on the y-axis) and the number of targets upon which discrimination was experienced (on the x-axis). The outlying highly vocal minority of bigoted players with high approach and high conflict strategies are represented by an asterisk.

> no one knows who you are, what life you live, what your job is, etc., unless you tell them. If everyone shut the fuck up about themselves, people could back to judge people on their actions and not their bio. (participant trolling as BWoLc26)

Challenges to *status quo* or privilege (e.g., when other players draw attention to discrimination) is experienced as a challenge to oneself, frequently coped with through dismissal of others' experiences. Linguistic descriptions contrast greatly with those from the privileged few, where universalizing and certainty are common: "discrimination doesn't exist and everyone needs to stop whining," MMeMc50. While the privileged few describe their experienced *acts* decontextualized from systemic sources, and do so intensely, they project their high tolerance for *acts* to all others, deeming low-effort coping strategies such as ignore/endure sufficiently successful.

The Emboldened Many

> I tend to mostly just ignore it and anticipate it. One can completely mute the chat system in (League of Legends), so I tend to do that because I know that that's a simple way to overcome the unnecessary fog of uneducated, dumbed down and juvenile children. Even when I witness the explicit racism I try to ignore it because there's nothing to be done. HWoLc33

Experience for this group varies greatly as it captures the largest spectrum of *targets*, *acts*, and *strategies*. *Acts* for this group include both present stimuli (e.g., slurs, gatekeeping, profiling, stereotypes) and absent stimuli (e.g., lack of representation, similar players, respect, consideration). Contrary to demands from the privileged few to "get used to how Internet banter works" or "just ignore it if it's offensive," the emboldened many most often describe *acts* motivating ("inciting") a reaction and that such actions require "approach" *strategies* but range in belief of its effectiveness:

> Now that I'm older, I don't experience discrimination much. If I do, I simply confront the person and/or engage in trash talk against them. I feel that the gaming community is more toxic than ever, and you have to learn to confront people or they will keep doing it. WMeLc29

Across this spectrum, players *desire* non-approach strategies like ignore/endure but "know it won't go away." Discrimination is "a daily experience," in and out of leisure: "When it happens, I cannot help but be affected. It really depends on the severity," WWoMc37D. While some experience surprise at its severity in a game, most are familiar enough to incorporate discrimination into their identity in forms of resistance: "I am a Black man. I am used to it. It makes me work harder," BMeUc37. Further along the x-axis, coping strategies become withdrawn more than resistant, modifying the self through greater forms of distance from sources and contexts of oppression.

Following *strategy* patterns from the u-shaped distribution, the analysis suggests that this group engages in approach strategies (to the extent permitted by their relative level of stress) to avoid modifying the self in response to discrimination. Depending on available stress tolerance and proximity, the emboldened many is most frequent to leverage energy for conflict when met with discrimination, "proactively" coping. Discrimination is, in a sense, activating (with otherwise privileged White women showing higher frequencies of directly confronting than White men, for example). As energy or hope in changing oppressive norms reduces, proactive or systemically resistant coping is less described. Less desired *strategies* are required. Players modify the digital self then modify the experience before needing to cope through changes to self (personal beliefs, preferences):

> People are assholes when given the opportunity to interact with others anonymously. I feel bad about it sometimes, but it really pushes me to predominantly single-player experiences. AMeMc25D

The Conflict Weary

> Honestly, most of the time I don't talk at all. I know I should and need to move past it but over the years, people have just shown so much hate online (...) because of the sound of my voice. Now and then I'll confront someone but it takes too much energy to deal with these people. HNbLc24LgbtqD

Commonly, the participants in this group are marginalized across multiple (2+) identity axes. Higher frequencies of discriminatory experiences require higher normalization, underlying the belief for this group that conflict is a waste of already taxed energy:

> Character create I just roll my eyes and try to convince myself it's not the largest part of the game. For online interactions, I usually give a person a few strikes, (...) if it's blatantly racist, that's what mute and report options should be for, though if one is determined enough, they can easily make these options a lesson in futility. BWoMc28Ft

Self-modifying—particularly self-dismissal and narrowing interest—is highest in this group as desires to play certain games or genres is deemed a higher risk than reward. As compounding *targets* increase, we see greater anticipatory coping (like hiding from the outset) and lower tolerance for *acts*—especially from *acts* of rendering grotesque and gatekeeping. Secondary or tertiary strategies (blocking, muting, switching games, modifying self) are employed earlier, even primarily, prompted from few to no *acts*.

Descriptions from the conflict weary are linguistically distinct, containing caveats and non-absolutist words ("suppose," "sometimes,"). Moments of revenge-seeking and willingness to conflict are occasionally described but, here, coincide with descriptions of self-blame. These descriptions accompany high normalization (*e. g.*, "It wasn't that bothersome then, either, I guess," BWoLc22TD). Self-modification and self-denial are prevalent:

> For the most part, I just avoid it happening in the first place. Pick a popular character for my icon that doesn't scream "girl", avoiding voice chat with teammates like the plague. I only voice chat if I'm playing with close friends, and even then it's on a group chat so not with the team. On the off chance I "slip" or something I try to ignore them in chat. I also honestly just try to be better than them, like focusing them if I get the chance, and then asking why they feel so high and mighty when I'm beating them (In... not so kind words sometimes). HNbLc24LgbtD

Considering the multiple axes of oppression facing, say, Black women, or queer and disabled non-binary players (notwithstanding the indigenous, homeless, and larger non-binary, trans-, and disabled populations we failed to sample), the need to survive high levels of discrimination across multiple axes shows itself in the most extreme coping strategies: self-dismissal, self-effacement, and self-modification. These players try new games to find they are inhospitable, hostile to the point that risking hope is an exhausting, Sisyphean endeavor. Lowering expectation of non-discriminatory game spaces, these players accept *gaming* as toxic. With most desiring to seek social support and collective hiding, the risk of self-disclosure forces coping to occur outside game contexts, ceasing play.

DISCUSSION

> I wish it didn't, but it gets to me. AWoMc25

Despite digital utopian (Charles, 2009) slogans of "unprecedented freedom" and "power to the player" because "you are the controller" and "you deserve to game your way," players' lived realities of inequity import to digital worlds. The players bring their accumulative, daily experiences—of harassment, gatekeeping, tokenism, histories of enforced poverty, enslavement, homicide, deportation, imprisonment, residential schools, forced infertility, exile, scapegoating, ghettoization, medical experimentation, profiling, trafficking (Iwasaki et al., 2009; Bird, 2013; Chief Moon-Riley, 2017; Yuen et al., 2019)—into digital game spaces; 60% of players describe being recalled to these histories of systemic oppression during play through recurring forms of discrimination. The players would *like* to ignore such experiences, trading acute stress for chronic autonomic stress (Alvarez and Juang, 2010), especially if digital utopia's promises could be delivered. However, attempts to cope with life's stress through gaming is seen in a return to the inequitable burdens of daily life to which they turn to games for relief. Threats of physical, psychosocial, racial, and sexual violence—trends in silencing, harassing, and gatekeeping marginalized players—reveal identity violence in digital games common enough to be anticipated, normalized, mundane.

Recalling that generalized harassment leads to less rumination than sexual harassment in games (Fox and Tang, 2017), we find that the targeted nature of identity violence in games is a form of adverse stress burdening players already burdened across gender, orientation, ethnicity, race, class, ability, age, culture, attraction, body type, and/or nationality.

> I usually stay quiet or I quit the game/match. (...) Arguing with someone will just get me more anxious and depressed. I'll end up ruminating for the entire day. HMeMc28Lgbtq

The effects of this violence are shown in player descriptions of persistently higher stress, negative affect, self-dismissal. Players are forced to cope by avoiding pro-social tools beneficial to other players (e.g., hiding signifiers of identity, masking, code-switching, *etc.*) as discussed by Vella et al. (2020) or by using tools for unintended purposes (e.g., using helmets where skin tones are limited to pretend a character is not a White male). Clusters of the emboldened many may feel empowered enough to employ approach *strategies* but do so while coping with oppressive *acts* and the demands of escalation and further risk of self-disclosure. Discriminated players increasingly devalue fellow gamers, gaming as a medium, and, in extreme cases, people altogether. Preferences for games and genres are altered, oppression is universalized as inevitable, rationalized, and generalized, and coping occurs anticipatorily. These changes to in-game behavior, surrounding beliefs to perceptions of self and others, and compounding levels of stress and negativity demonstrate the pervasive effects of coping with discrimination in games.

Cycle 1: Self-Reifying Factors to the Atmosphere of Stigma

Results support the explanations for negative generalizations surrounding gamers and games as "toxic" (Shaw, 2012; Kuznekoff and Rose, 2013). Those impacted by *compounding* oppression more commonly recognize other forms of systemic disadvantage:

> While Black characters exist in games, I often find them more of a caricature and don't properly represent what we know as the struggle. Even putting race aside, something as simple as a diabetic in game would be more interesting, it wouldn't have to be the focus of the game but a part much like mana and health that ignoring it would be a detriment. BMeLc28D

Awareness of systemic oppression itself is shown to induce stress and guilt and impact relief and health factors (Fujishiro, 2009), contributing to stress and unease. Given the normalcy of discriminatory stressors in digital games, players familiar with systemic oppression on one axis describe apprehension, vigilance, an "atmosphere" of oppression—even when not directly targeted.

Though an exaggerated example, the "mistaken identity" cases show the intentions and the histories behind discriminatory acts as felt even by those who do not identify with them.

> I was getting some racial discrimination for this character due to his dark skin. Ultimately, I ended up changing the character for another one. The constant joking around was just too much after that. HMeMc41

Identities are often self-verified through moral behavior *toward* others, and witnessing discrimination can provoke a moral imperative to act in or toward said group (Stets and Carter, 2011). In these cases, one's agency is reinforced through sharing a struggle with others, which players describe through their newfound "appreciation" for experiences of oppression. Dynamics of active commitment (Downing and Roush, 1985), acts and feelings of solidarity, are seen among players who less often experience discrimination (or do so indirectly) but hold beliefs around its injustice.

For a small proportion from our sample, these challenges increase performance, self-assertion, "grit," compelling direct confrontation (seen in revenge strategies Consalvo, 2008; Cicchirillo, 2015; Leonard, 2020). In highly specific conditions, exposure to discrimination *can* benefit sympathy (as seen in literature on proteus effects Yee and Bailenson, 2007; Gutierrez et al., 2014; Ash, 2015) and provoke conflict against discriminatory sources. Players describe calling out discriminators, confronting them, and beliefs around solidarity; however, these instances are among the least common. Almost no participant described engaging in bystander intervention, explaining why no one described seeking aid from in-game strangers. Rather than acting on this moral imperative, we see descriptions of stereotype threat effects, escalation of abuse, and guarding against self-disclosure.

In lieu of gaming culture facilitating social support, players are left describing mostly negative aspects around identification: anxieties around failing one's social identity group or affirming stereotypes (Cadinu et al., 2005; Vella et al., 2020). Such threats impair working memory (Beilock et al., 2007) and executive functions (Cicchirillo, 2015) and provoke coping strategies even if discrimination is merely anticipated (Johns et al., 2008). Witnessing regular identity violence contributes to a general sense of insecurity, vulnerability, and social threat. Coping strategies of generalization, personal distancing, and low investment ("it's just a game") combine, leading players to attribute their experiences of discrimination and unease in *separate* game contexts to games and gamers in general. Recalling that 60% of players describe *recurring*, *first-hand experiences* of discrimination, these perceived elements of unease suggests that more players are impacted by discrimination in and from digital gaming environments than not.

Cycle 2: Desiring Visibility, Coping Through Collective Invisibility

> When I first started playing Fortnite, I started in random group games. In previous games, I would always turn off my voice chat, but I decided to leave it on for Fortnite. I went through a string of games where younger players were saying absolutely vile and racist rhetoric. (…) The only time I turn on voice chat now is if I'm playing with people I know. BWoMc36D

Players prefer to self-identify in spaces offline and online but cannot (Kafai et al., 2010; Barsamian Kahn et al., 2013; Shaw and Friesem, 2016; Passmore et al., 2018). The results further bolster criticisms like Shaw's, showing that representation is important to player experience, but without addressing surrounding systemic oppression, players can be left over-exposed and under-supported. *Fortnite* and *Overwatch* are celebrated for their diversity in character design and appeal to a wider audience (Conditt, 2019), yet player experiences depict a host of direct and indirect barriers even when such tools are available (Callahan, 2018). A curious case study in itself, nearly all women from our sample who mention either game describe: "not a lot of females that game in *Fortnite*, it makes me uncomfortable to speak on the microphone in groups." Marginalized players learn not to use the tools meant to benefit them:

> When I first started playing online, I didn't disguise my voice and used a female name. (...) They concentrated on hitting on me. Then, when they got rebuffed, they got pissy and concentrated on either actively sabotaging me during the run or just incessantly calling me names. This was so bad I had to stop using any kind of voice and had to change my player names from female to either male or neutral. HWoLc34Lgbtq

Women's "dislike" for first-person shooters has been referenced by players and developers (Au, 2018); however, our results support that players regularly risk discrimination out of desire to play these games (and play them as themselves) but face three times the harassment when "outed" by speaking on microphones or self-representing (Kuznekoff and Rose, 2013). Escalating harassment, conflict, and other forms of violence become less preferable to genre and representational preferences over time and repetition. Chronic discrimination is preference-forming (Dale et al., 2018). Learning to anticipate escalation and exhaustion from repetitive, cyclical conflict leads to more inward-turned coping, more modification of self. Risking overall resilience for potential benefits from gaming is an ill-advised cost–benefit analysis—especially when it involves betraying a learned history of failed attempts to enact lasting systemic change.

> I generally just accept it. There's not anything that can change their perception of my people, not in my lifetime anyway. Only time will change that. If my friends are with me, we do protect each other. WMeLc21LgbtqD

Enduring misogyny, chauvinism, ableism, racism, and classism, the human desire to be seen and find like-minded others is transformed when one is rendered grotesque. Invisibility is preferred to stereotype threats, tokenism, or exposure to abuse. It is often a well-learned distrust in others' construction of their identities that protects against further harm; it is less taxing to cope with being invisible than repeatedly being a target of violence. Unfortunately, invisibility exaggerates the absence of similar identities—inhibiting players from seeking social support *while* magnifying discriminatory voices. Those in the position to discriminate are given more space to do so; those who defiantly self-represent are left over-exposed, under-supported, and further stressed. For new players, coping through collective invisibility looks like an absence of similar others, associating that space with others rife with identity-targeted violence, provoking anticipatory coping strategies of hiding, masking, or rendering oneself invisible. Coping with digital spaces like this forms cycles that feed back into power imbalances, amplifying discriminatory stress on bodies already unduly stressed.

Cycle 3: The Ease of Discrimination, the Accumulative Burden of Coping

Few coping strategies in the context of gaming appear to be "positive" by Lazarus and Folkman's standards (Lazarus and Folkman, 1984). Positive, adaptive strategies are described, also evidenced in Gibbons (2015) and Gray (2018), but are exceptions to the norm. In general, this study presents a largely negative account of a medium notably beneficial and well-received. This negativity may be exaggerated by our study priming negative experiences and requesting players relate them as such or the high sensitivity and negative bias of our analyses. Most likely, this negativity is the result of both, allowing for a more *de*-normalized account of "mundane" identity violence. The barriers to coping through positive channels (social support seeking, in-game moments for reconciliatory dialogue with others, unfettered access to gaming's benefits) are many, but expectations for removal of those barriers are long gone. Feedback cycles create an atmosphere for growing norms of toxicity, for coping through collective invisibility. The results suggest high effort, and risk is needed to access gaming's more "positive" channels for coping even where they do exist.

Contrasting these difficulties are descriptions of the ease with which players and game content offend. Acknowledgment of "privilege" or mention of systemic injustice is perceived as a direct threat by some (often a highly vocal section of the privileged few). Outperforming others "invites" discrimination. Discussed with far greater nuance by Ortiz (2019), *acts* of rendering invisible and grotesque are far more harmful than "friendly banter." This is to show that players engage in low-effort discrimination to regain a sense of agency at the cost of another's. While these dynamics are common to transactional relationships of agency and power (Moghaddam et al., 2002; Berjot and Gillet, 2011), gaming pairs these relationships with a unique dynamic: control over self-disclosure.

> This misuse of (autism) greatly annoys me. As a result, I almost never disclose my ASD status nor discuss the problems that come with having it. AMeLc28D

Heightened rates of harassment in online games can be attributed to "social disinhibition," which suggests that anonymity facilitates violent behaviors (e.g., hate speech, gatekeeping). The relative anonymity of online play supports a lack of direct repercussions for abusers (Fox and Tang, 2017). Lack of accountability benefits discriminators. When combined with an absence of bystander intervention and in-game social supports due to both requiring risky self-disclosure and great effort, a vacuum of negativity is formed. Restorative or corrective player interactions are left without space.

Discrimination and coping share a preference for paths of least effort, even more so in leisure and play than in contexts of work or family (Walker et al., 1977; Yuen et al., 2019). The ease and the casualness with which players can create discriminatory stressors in a game (e.g., using a slur) contrast the length of recovery time required from acute stress events (Berjot and Gillet, 2011). This stress compounds with stress from systemic marginalization, often compounding with intergenerationally transmitted histories of stress and trauma.

Discrimination is dealt with at a cost, as is resistance to it. Learned anticipation of discrimination is informed by its mundane frequency and insidious variety of sources. Anticipating discrimination becomes an unfortunate necessity for self-protection across multiple exclusionary norms. Expectation reduces the intensity of a stress response, reducing acute stress but at a cost of greater chronic stress (Brondolo et al., 2009b; Liston et al., 2009). Here we see this cost as coping with acute, direct forms of discrimination through employing coping strategies that are, objectively, self-discriminatory: gatekeeping oneself from preferences, hiding, masking, code switching, self-denial. The realities of internalizing discrimination often resemble belief and value adjustments, cognitive reframing, and self-suppression. Ignoring or minimizing acute stress responses to discrimination results in autonomic stress, which can lead to *less* resilience to stress through deactivation of dopamine receptors (Chen et al., 2016), lower motivation, impulse control, decision-making, focus, and effort discounting (Gassen and Hill, 2019; Treadway et al., 2019). These costs to resilience when gaming is meant to be a coping activity are perceived and felt:

> Dealing with those type of people when I am trying to relax is exhausting. HMeLc32

Cycle 4: Equality, Equity, and the Broken Promises of Digital Utopia

> Mostly I'm numb to it as I grew up with the Internet and trolling is something I've dealt with for 25 years and I just don't care anymore. (preferred not to disclose)

The results here, as in Gibbons (2015), Fox and Tang (2017), Vella et al. (2020) and Gray (2012a,b, 2018), reflect a potential for coping strategies unique to digital games: like-minded players can connect over global networks; avatar customization, use of social features like chat, private server creation, and the ability to go offline promote greater control over one's digital environments; identity play for personal exploration and norm-bending (Martey et al., 2014) permits creativity and control over self-disclosure; players are given opportunities for cathartic revenge [also seen in Consalvo (2008) and Leonard (2020)]. When low-demand coping strategies fail, players engage in space-making and refuge-taking. They mute, report, or block players and game elements, exercising control over their digital worlds.

The ingenuity of players in overcoming toxic norms and shoring up agency warrants celebration. However, affordances and realities are distinct. The experiential divide between players utilizing tools for pleasure and those utilizing them to cope is massive. Where the privileged few use affordances to additional benefit, the emboldened many and conflict weary are further burdened by their use to mitigate discrimination. They are *forced from* adverse experiences with uncertainty of success rather than *motivated toward* positive ones. Even where gaming's potential tools are described as consistently available, reliable, and less burdensome, such tools service coping rather than agency. They re-center toxic norms and systemic oppression. This distinction is crucial. The benefits of gaming's tools and the power with which they are wielded are inequal.

When play is designed with the privileged as a frame of reference for new features, marginalized players are subject to increased stress. Microphones and avatars "out" players, inaccurate attempts at diversity further stigmatize, and social tools facilitate harassment. Providing *all* players with tools for greater agency when in unequal spaces results in a magnification of social power imbalances, disserving some groups while promoting social connection between more privileged groups. The inseparability of players from their lived identities means better player experiences must be approached through equity rather than equality, anti-oppression rather than utopian myths about potential and diverse self-fashioning. Digital worlds are not blank slates.

Understanding players means not just understanding how gaming fits into their lifeworld or what they *can* do but how intergenerationally transmitted social power relations determine what they *must* do. These relativities inform desires, perceived and suppressed stressors, transactional agency, and tiered obstacles to play. Interrupting and inhibiting the domino effects and feedback cycles of oppression discussed here are central to create *affiliative* digital spaces and lower identity-based violence. Well-cited discriminatory norms underlie both game spaces and their developers (IGDA, 2014; Srauy, 2019). A collective responsibility, a moral imperative to interrupt the normalization of discrimination is here. For developers, this means implementing features with equity in mind—not equality, benefits—not band-aids for coping, creating accountability in digital spaces, and risking reactions from the highly vocal minority of abusive players. For players, this means leveraging one's available privilege for changing norms in gaming through bystander intervention, providing social support, and demanding accountability around *acts*. For researchers, this means great effort, precision, and social responsibility in methods—in whose stories are told by data to what end.

DISCLAIMERS AND LIMITATIONS

The complications of self-report, normalization, recruitment methods (i.e., MTurk), and the plurality of subjective experiences of social power urge caution when drawing conclusions from this study. To reduce these barriers, we conducted an interdisciplinary literature review (>300 studies), used priming, trust engenderment, pre-study interviews, and co-constructive iterative survey design. Our data analyses were reviewed by non-participant players who confirmed our results, framings, and the implications, supporting that this study accurately reflects at least their experiences. Still objectively definitive, generalizable claims cannot be made from this study. Truly intersectional analysis requires a lens and sample size beyond our means; the quantitative analysis of this depth is infeasible even had we obtained an n of >20 for each combination of (an already problematic reduction to) three genders, five racial categories, two disability categories, and three classes. We took methodological efforts to correct for normalization, suppression, and the habitual reactivity of coping, but indirect sources of discrimination remain lower as reported than what analyses suggest. Where they appear, our identity-factor groupings are heuristic "more similar than not" categories that insufficiently represent identity (there is no more "a black" experience than there is "a way to cope" with "class" or "disability"). Experiences vary and are pluralistic. Just as we resist tying results to individual demographic axes at a disadvantage to traditional scientific "contributions," we did not evaluate the success or the efficacy of coping strategies. Without *knowing* the participants, judging the "health" of their coping can be discriminatory itself. Striking an optimal balance between experiential accuracy and generalizable comparisons formed an ongoing debate for us across all stages of the research process. With barely the space to represent an already reductive series of intersecting demographics, additional player variables (such as gamer profiles, game preferences, *etc.*) would increase the paper length and analyses exponentially and require an unfeasible sample size. All these complications led us to self-critique and multiple rounds of community consultation to ensure that an overview study of this nature—a taxonomy "writ large" —was accurate, justified, valuable in filling a crucial gap in literature on discrimination and digital games. Finally, we come from an interdisciplinary lens of historical materialist social power relations, which is a bias reflected throughout our study.

CONCLUSION

> I choose not to continue playing the game(s). I talk with others about my experiences. I try to choose healthier games for my children to play. MWoMc51Lgbtq

We provide additional evidence that most players cope with oppression during gaming. Relatively privileged players access lower burdens of stress to react across a greater range of interpersonally directed coping behaviors, while those more impacted by discriminatory stress are forced to cope inwardly, with more severe forms of anticipatory coping deployed earlier. Those more commonly marginalized in American society (disabled, queer, lower-class Black women) are most frequently targeted in digital play. Those more exposed to systemic oppression bring that chronic stress to their games. Of course, these exposure rates are deeply informed by socio-historically situated identity factors imported to the gaming context. With respect to plurality, however, player experiences are more accurately patterned by discriminatory stress exposure rates than demographic variables alone. This stress compounds, accumulates, and burdens players, leading them to seek games and play for relief only to experience further discriminatory stress. Within the context of psychological and epidemiological studies on chronic stress, the long-term detriments these experiences may have on player health are potentially substantial. As a supplement to future studies on precisely this, we provide a comprehensive taxonomy of discriminatory stressors and coping strategies in digital gameplay.

Against discourses of erasure, the results show little willing suspension of disbelief where reminders of physical, emotional, and cultural violence exist. Organizing players with respect to these lived realities of social power grants insight to the ways that discrimination shapes player experiences, beliefs, and behaviors—during gameplay and after. Accounting for spectrums of privilege, *most* players experience recurring discriminatory stress along at least one axis of identity or another. These experiences remain under-reported and over-normalized. The substantial disparities between player experiences along axes of social marginalization provide further evidence that coping with these stressors is itself a compounding burden. Having identified several feedback cycles serving inequitable norms, we see the responsibility for interrupting these cycles falling on those with an expendable privilege to act against gaming's toxic norms. This responsibility is as much a finding as it is a commitment by the authors of this study in our research and our play.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

All participants were 18+ and legally consenting per MTurk standards. Written, informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article. This research was approved by the Behavioral Research Ethics Board at the University of Saskatchewan.

AUTHOR CONTRIBUTIONS

RM and CP contributed to the formation of the research idea, the methods employed, the design of the interview questions, the data gathering, and contextualized the findings and wrote the paper. CP conducted all analyses and proposed the results frameworks. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by NSERC Discovery Grants and the NSERC SWaGUR CREATE Training Program.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomp.2020.00040/full#supplementary-material

REFERENCES

Abeele, V. V., Spiel, K., Nacke, L., Johnson, D., and Gerling, K. (2020). Development and validation of the player experience inventory: a scale to measure player experiences at the level of functional and psychosocial consequences. *Int. J. Hum. Comput. Stud.* 135:102370. doi: 10.1016/j.ijhcs.2019.102370

Alvarez, A. N., and Juang, L. P. (2010). Filipino americans and racism: a multiple mediation model of coping. *J. Couns. Psychol.* 57, 167–178. doi: 10.1037/a0019091

American Psychological Association (2016). *Stress in America: The Impact of Discrimination*. Available online at: www.stressinamerica.org (accessed January 29, 2020).

Anisman, H., and Merali, Z. (1999). Understanding stress: characteristics and caveats. *Alcohol Res. Heal.* 23, 241–249.

Ash, E. (2015). Priming or proteus effect? Examining the effects of avatar race on in-game behavior and post-play aggressive cognition and affect in video games. *Games Cult*. 11, 422–440. doi: 10.1177/1555412014568870

Au, H. (2018). *New World Notes: 20 Million+ Fortnite Players Are Female*. *Nwn.Blogs.Com*. Available online at: https://nwn.blogs.com/nwn/2018/07/fortnite-demographics-similarweb.html (accessed September 20, 2019).

Axt, J. (2011). Understanding how to best measure self-reported racial attitudes. *Proj. Implicit*. Available online at: https://implicit.harvard.edu/implicit/user/jaxt/blogposts/piblogpost002.html (accessed June 17, 2020).

Bacchus, D. N. A. (2008). Coping with work-related stress: a study of the use of coping resources among professional black women. *J. Ethn. Cult. Divers. Soc. Work* 17, 60–81. doi: 10.1080/15313200801906443

Balsam, K. F., Molina, Y., Beadnell, B., Simoni, J., and Walters, K. (2011). Measuring multiple minority stress: the LGBT people of color microaggressions scale. *Cult. Divers. Ethn. Minor. Psychol.* 17, 163–174. doi: 10.1037/a0023244

Bandalos, D. L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Struct. Equ. Model.* 9, 78–102. doi: 10.1207/S15328007SEM0901_5

Baniqued, P. L., Lee, H., Voss, M. W., Basak, C., Cosman, J. D., DeSouza, S., et al. (2013). Selling points: what cognitive abilities are tapped by casual video games? *Acta Psychol.* 142, 74–86. doi: 10.1016/j.actpsy.2012.11.009

Barsamian Kahn, K., Spencer, K., and Glaser, J. (2013). "Online prejudice and discrimination: from dating to hating," in *The Social Net* (Oxford University Press), 201–219. doi: 10.1093/acprof:oso/9780199639540.003.0011

Behm-Morawitz, E. (2017). Examining the intersection of race and gender in video game advertising. *J. Mark. Commun.* 23, 220–239. doi: 10.1080/13527266.2014.914562

Beilock, S. L., Rydell, R. J., and McConnell, A. R. (2007). Stereotype threat and working memory: mechanisms, alleviation, and spillover. *J. Exp. Psychol. Gen.* 136, 256–276. doi: 10.1037/0096-3445.136.2.256

Berjot, S., and Gillet, N. (2011). Stress and coping with discrimination and stigmatization. *Front. Psychol.* 2:33. doi: 10.3389/fpsyg.2011.00033

Bessière, K., Seay, A. F., and Kiesler, S. (2007). The ideal elf: identity exploration in world of warcraft. *Cyberpsychology Behav.* 10, 530–535. doi: 10.1089/cpb.2007.9994

Bird, K. (2013). "The intergenerational transmission of poverty: an overview," in *Chronic Poverty* (London: Palgrave Macmillan), 60–84. doi: 10.1057/9781137316707_4

Black, W. W. (2016). an E xamin ation of rel ation s among fear, guilt, self-comp assion, and mul ticultural a ttitudes in white ad ults. (PhD Thesis) *Dr. Diss. Univ. Kentucky UKnowledge RetriedProQuest* 10954403 2016, 74–98.

Bowman, N. D., and Tamborini, R. (2015). "In the mood to game": selective exposure and mood management processes in computer game play. *New Media Soc.* 17, 375–393. doi: 10.1177/1461444813504274

Boyle, E. A., Connolly, T. M., Hainey, T., and Boyle, J. M. (2012). Engagement in digital entertainment games: a systematic review. *Comput. Human Behav.* 28, 771–780. doi: 10.1016/j.chb.2011.11.020

Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. *Qual. Res. Psychol.* 3, 77–101. doi: 10.1191/1478088706qp063oa

Brondolo, E., Beatty, D. L., Cubbin, C., Pencille, M., Saegert, S., Wellington, R., et al. (2009a). Sociodemographic variations in self-reported racism in a community sample of blacks and latino(a)s. *J. Appl. Soc. Psychol.* 39, 407–429. doi: 10.1111/j.1559-1816.2008.00444.x

Brondolo, E., Ver Halen, N. B., Pencille, M., Beatty, D., and Contrada, R. J. (2009b). Coping with racism: a selective review of the literature and a theoretical and methodological critique. *J. Behav. Med.* 32, 64–88. doi: 10.1007/s10865-008-9193-0

Brosschot, J. F., Verkuil, B., and Thayer, J. F. (2017). Exposed to events that never happen: generalized unsafety, the default stress response, and prolonged autonomic activity. *Neurosci. Biobehav. Rev.* 74, 287–296. doi: 10.1016/j.neubiorev.2016.07.019

Burgess, M. C. R., Dill, K. E., Stermer, S. P., Burgess, S. R., and Brown, B. P. (2011). Playing with prejudice: the prevalence and consequences of racial stereotypes in video games. *Media Psychol.* 14, 289–311. doi: 10.1080/15213269.2011.596467

Cadinu, M., Maass, A., Rosabianca, A., and Kiesner, J. (2005). Why do women underperform under stereotype threat? Evidence for the role of negative thinking. *Psychol. Sci.* 16, 572–578. doi: 10.1111/j.0956-7976.2005.01577.x

Callahan, M. (2018). *Despite Inclusive Design, Fortnite Gamers Victim to Gendered Harassment*. *News @ Northeast.* Available online at: https://news.northeastern.edu/2018/04/10/despite-inclusive-design-fortnite-gamers-victim-to-gendered-harassment/ (accessed September 20, 2019).

Charles, A. (2009). Eludamos. *Eludamos. J. Comput. Game Cult.* 3, 281–294. Available online at: https://eludamos.org/index.php/eludamos/article/view/72 (accessed September 20, 2019).

Chen, J., Wang, Z. Z., Zuo, W., Zhang, S., Chu, S. F., and Chen, N. H. (2016). Effects of chronic mild stress on behavioral and neurobiological parameters - role of glucocorticoid. *Horm. Behav.* 78, 150–159. doi: 10.1016/j.yhbeh.2015.11.006

Chief Moon-Riley, K. (2017). *The Biological Impacts of Residential Schooling on the Development of Intergenerational Trauma Among Indigenous People*. *ProQuest Diss. Theses*, 128. Available online at: https://search.proquest.com/docview/1988779829?accountid=14468%0Ahttp://wx7cf7zp2h.search.serialssolutions.com?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rfr_id=info:sid/ProQuest+Dissertations+%26+Theses+Global&rft_val_fmt=info:ofi/fmt:kev:mtx:dissert (accessed September 20, 2019).

Cicchirillo, V. (2015). Priming stereotypical associations: violent video games and african american depictions. *Commun. Res. Rep.* 32, 122–131. doi: 10.1080/08824096.2015.1016148

Citrona, D. K. (2014). *Defining Online Harassment. Forbes.com*, 1–8. Available online at: http://www.forbes.com/sites/daniellecitron/2014/10/23/defining-online-harassment/#41df68414360 (accessed September 20, 2019).

Cokley, K. (2007). Critical issues in the measurement of ethnic and racial identity: a referendum on the state of the field. *J. Couns. Psychol.* 54, 224–234. doi: 10.1037/0022-0167.54.3.224

Collins, E., Cox, A., Wilcock, C., and Sethu-Jones, G. (2019). Digital games and mindfulness apps: comparison of effects on post work recovery. *JMIR Ment. Heal.* 6:e12853. doi: 10.2196/12853

Conditt, J. (2019). *Epic Games has 250 Million "Fortnite" Players and a Lot of Plans*. *Www.Engadget.Com*. Available online at: https://www.engadget.com/2019/03/20/fortnite-250-million-epic-games-sweeney-interview-gdc/ (accessed September 20, 2019).

Consalvo, M. (2008). *Cheating: Gaining Advantage in Videogames*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/1802.001.0001

Crenshaw, K. (1991). Mapping the margins: intersectionality, identity politics, and violence against women of color. *Stanford Law Rev.* 43, 1241–1299. doi: 10.2307/1229039

Dale, G., Sampers, D., Loo, S., and Shawn Green, C. (2018). Individual differences in exploration and persistence: grit and beliefs about ability and reward. *PLoS ONE* 13:e0203131. doi: 10.1371/journal.pone.0203131

De Grove, F., Cauberghe, V., and Van Looy, J. (2016). Development and validation of an instrument for measuring individual motives for playing digital games. *Media Psychol.* 19, 101–125. doi: 10.1080/15213269.2014.902318

De Schutter, B., and Vanden Abeele, V. (2010). "Designing meaningful play within the psycho-social context of older adults," *in ACM International Conference Proceeding Series*, 84–93. doi: 10.1145/1823818.1823827

Delgado, R., Stefancic, J., and Harris, A. (2001). *Critical Race Theory An Introduction.* New York, NY: NYU Press. doi: 10.1093/acref/9780195301731.013.51089

Dennis, T. A., and O'Toole, L. J. (2014). Mental health on the go: effects of a gamified attention-bias modification mobile application in trait-anxious adults. *Clin. Psychol. Sci.* 2, 576–590. doi: 10.1177/2167702614522228

Depping, A. E., Johanson, C., and Mandryk, R. L. (2018). "Designing for friendship: Modeling properties of play, in-game social capital, and psychological well-being," in *CHI PLAY 2018 - Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play* (New York, NY, USA: Association for Computing Machinery, Inc), 5–17. doi: 10.1145/3242671.3242702

Dietrich, D. R. (2013). Avatars of whiteness: racial expression in video game characters. *Sociol. Inq.* 83, 82–105. doi: 10.1111/soin.12001

Domahidi, E., Breuer, J., Kowert, R., Festl, R., and Quandt, T. (2018). A longitudinal analysis of gaming- and non-gaming-related friendships and social support among social online game players. *Media Psychol.* 21, 288–307. doi: 10.1080/15213269.2016.1257393

Downing, N. E., and Roush, K. L. (1985). From passive acceptance to active commitment: a model of feminist identity development for women. *Couns. Psychol.* 13, 695–709. doi: 10.1177/0011000085134013

Edström, M. (2018). Visibility patterns of gendered ageism in the media buzz: a study of the representation of gender and age over three decades. *Fem. Media Stud.* 18, 77–93. doi: 10.1080/14680777.2018.1409989

Fanon, F., Philcox, R., Sartre, J.-P., and Bhabha, H. K. (2004). *The Wretched of the Earth*. New York, NY: Grove Press.

Finda Ogbonnaya-Ogburu, I., Smith, A. D. R., To, A., and Toyama, K. (2020). "Critical race theory for HCI," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (New York, NY: Association for Computing Machinery) 1–16. doi: 10.1145/3313831.3376392

Fox, J., and Tang, W. Y. (2017). Women's experiences with general and sexual harassment in online video games: rumination, organizational responsiveness, withdrawal, and coping strategies. *New Media Soc.* 19, 1290–1307. doi: 10.1177/1461444816635778

Fragoso, J. M., and Kashubeck, S. (2000). Machismo, gender role conflict, and mental health in mexican american men. *Psychol. Men Masculinity* 1, 87–97. doi: 10.1037/1524-9220.1.2.87

Fujishiro, K. (2009). Is perceived racial privilege associated with health? Findings from the behavioral risk factor surveillance system. *Soc. Sci. Med.* 68, 840–844. doi: 10.1016/j.socscimed.2008.12.007

Gassen, J., and Hill, S. E. (2019). Why inflammation and the activities of the immune system matter for social and personality psychology (and not only for those who study health). *Soc. Personal. Psychol. Compass* 13:e12471. doi: 10.1111/spc3.12471

Gassen, J., Prokosch, M. L., Eimerbrink, M. J., Proffitt Leyva, R. P., White, J. D., Peterman, J. L., et al. (2019). Inflammation predicts decision-making characterized by impulsivity, present focus, and an inability to delay gratification. *Sci. Rep.* 9:4928. doi: 10.1038/s41598-019-41437-1

Gibbons, S. (2015). Disability, Neurological diversity, and inclusive play: an examination of the social and political aspects of the relationship between disability and games. *Loading J. Can. Game Stud. Assoc.* 9, 25–39.

Giorgi, A. (2010). *The Descriptive Phenomenological Method in Psychology: A Modified Husserlian Approach.* Pittsburgh, PA: Duquesne University Press

Granic, I., Lobel, A., and Engels, R. C. M. E. (2014). The benefits of playing video games. *Am. Psychol.* 69, 66–78. doi: 10.1037/a0034857

Gray, K. L. (2012a). Deviant bodies, stigmatized identities, and racist acts: examining the experiences of African-American gamers in xbox live. *New Rev. Hypermedia Multimed.* 18, 261–276. doi: 10.1080/13614568.2012.746740

Gray, K. L. (2012b). Intersecting oppressions and online communities: examining the experiences of women of color in xbox live. *Inf. Commun. Soc.* 15, 411–428. doi: 10.1080/1369118X.2011.642401

Gray, K. L. (2018). Gaming out online: black lesbian identity development and community building in xbox live. *J. Lesbian Stud.* 22, 282–296. doi: 10.1080/10894160.2018.1384293

Greer, M. T., Laseter, A., and Asiamah, D. (2009). Gender as a moderator of the relation between race-related stress and mental health symptoms for African Americans. *Psychol. Women Q.* 33, 295–307. doi: 10.1111/j.1471-6402.2009.01502.x

Gutierrez, B., Kaatz, A., Chu, S., Ramirez, D., Samson-Samuel, C., and Carnes, M. (2014). "fair play": a videogame designed to address implicit race bias through active perspective taking. *Games Health J.* 3, 371–378. doi: 10.1089/g4h.2013.0071

Hill Collins, P. (2002). *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. New York, NY: Routledge doi: 10.4324/9780203900055

Holloway, C., Gerling, K., Power, C., Spiel, K., Barbareschi, G., Cox, A., et al. (2019). "Disability interactions in digital games: from accessibility to inclusion," in *CHI PLAY 2019 - Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play* (Association for Computing Machinery, Inc), 835–839. doi: 10.1145/3341215.3349587

Iacovides, I., and Mekler, E. D. (2019). "The role of gaming during difficult life experiences," in *Conference on Human Factors in Computing Systems - Proceedings* (New York, NY: Association for Computing Machinery). doi: 10.1145/3290605.3300453

IGDA (2014). *IGDA Developer Satisfaction Survey. Summary Report*. Available online at: http://www.igda.org/news/179436/igda-developer-satisfaction-survey-summary-report-available.htm (accessed January 29, 2020).

Inman, M. L., and Baron, R. S. (1996). Influence of prototypes on perceptions of prejudice. *J. Pers. Soc. Psychol.* 70, 727–739. doi: 10.1037/0022-3514.70.4.727

Iwasaki, Y., Bartlett, J. G., Gottlieb, B., and Hall, D. (2009). Leisure-like pursuits as an expression of aboriginal cultural strengths and living actions. *Leis. Sci.* 31, 158–173. doi: 10.1080/01490400802686011

Johns, M., Inzlicht, M., and Schmader, T. (2008). Stereotype threat and executive resource depletion: examining the influence of emotion regulation. *J. Exp. Psychol. Gen.* 137, 691–705. doi: 10.1037/a0013834

Kafai, Y. B., Cook, M. S., and Fields, D. A. (2010). "Blacks deserve bodies too!": design and discussion about diversity and race in a tween virtual *world. Games Cult.* 5, 43–63. doi: 10.1177/1555412009351261

Kafai, Y. B., Fields, D. A., and Cook, M. S. (2007). "Your second selves: *Resources*, agency, and constraints in avatar designs and identity play in a tween virtual world," in *Third International Conference of the Digital Games Research Association (DiGRA)* Tokyo.

Karlamangla, A. S., Friedman, E. M., Seeman, T. E., Stawksi, R. S., and Almeida, D. M. (2013). Daytime trajectories of cortisol: demographic and socioeconomic differences-Findings from the national study of daily experiences. *Psychoneuroendocrinology* 38, 2585–2597. doi: 10.1016/j.psyneuen.2013.06.010

Kittur, A., Chi, E. H., and Suh, B. (2008). "Crowdsourcing user studies with mechanical turk," in *Conference on Human Factors in Computing Systems - Proceedings* (New York, NY: ACM Press), 453–456. doi: 10.1145/1357054.1357127

Koivisto, J., and Hamari, J. (2014). Demographic differences in perceived benefits from gamification. *Comput. Human Behav.* 35, 179–188. doi: 10.1016/j.chb.2014.03.007

Krieg, A., and Xu, Y. (2018). From self-construal to threat appraisal: understanding cultural differences in social anxiety between Asian Americans and European Americans. *Cult. Divers. Ethn. Minor. Psychol.* 24, 477–488. doi: 10.1037/cdp0000194

Krieger, N. (1999). Embodying inequality: a review of concepts, measures, and methods for studying health consequences of discrimination. *Int. J. Heal. Serv.* 29, 295–352. doi: 10.2190/M11W-VWXE-KQM9-G97Q

Kuo, B. C. H. (2011). Culture's consequences on coping: theories, evidences, and dimensionalities. *J. Cross. Cult. Psychol.* 42, 1084–1100. doi: 10.1177/0022022110381126

Kuznekoff, J. H., and Rose, L. M. (2013). Communication in multiplayer gaming: examining player responses to gender cues. *New Media Soc.* 15, 541–556. doi: 10.1177/1461444812458271

Lazarus, R., and Folkman, S. (1984). *Stress, Appraisal, and Coping.* Available online at: https://books.google.com/books?hl=en&lr=&id=i-ySQQuUpr8C&oi=fnd&pg=PR5&ots=DfIVhuicSa&sig=Zft5tkr5wZfpvy-VPpnJgKw669E (accessed September 20, 2019).

Lazarus, R. S. (2000). Toward better research on stress and coping. *Am. Psychol.* 55, 665–673. doi: 10.1037/0003-066X.55.6.665

Leonard, D. J. (2006). Not a hater, just keepin' it real: the importance of race- and gender-based game studies. *Games Cult.* 1, 83–88. doi: 10.1177/1555412005281910

Leonard, D. J. (2020). Virtual anti-racism: pleasure, catharsis, and hope in mafia III and watch dogs 2. *Humanity Soc.* 44, 111–130. doi: 10.1177/0160597619835863

Liston, C., McEwen, B. S., and Casey, B. J. (2009). Psychosocial stress reversibly disrupts prefrontal processing and attentional control. *Proc. Natl. Acad. Sci. U.S.A.* 106, 912–917. doi: 10.1073/pnas.0807041106

Lorde, A. (2012). *Sister Outsider: Essays and Speeches*. Available online at: https://books.google.ca/books?hl=en&lr=&id=kE3ek_-FGWgC&oi=fnd&pg=PA6&dq=lorde&ots=iF3_2BNLZ9&sig=7_GZZoYf5Jll3AmX-8at8a7lVPU (accessed June 16, 2020).

Luthar, S. S. (2006). "Resilience in development: a synthesis of research across five decades," in *Developmental Psychopathology: 2nd Edn.,* eds D. Cicchetti and D. J. Cohen (wiley), 739–795. doi: 10.1002/9780470939406.ch20

Martey, R. M., Stromer-Galley, J., Banks, J., Wu, J., and Consalvo, M. (2014). The strategic female: gender-switching and player behavior in online games. *Inf. Commun. Soc.* 17, 286–300. doi: 10.1080/1369118X.2013.874493

Mason, W., and Suri, S. (2012). Conducting behavioral research on amazon's mechanical turk. *Behav. Res. Methods* 44, 1–23. doi: 10.3758/s13428-011-0124-6

Mathur, M. B., Epel, E., Kind, S., Desai, M., Parks, C. G., Sandler, D. P., et al. (2016). Perceived stress and telomere length: a systematic review, meta-analysis, and methodologic considerations for advancing the field. *Brain. Behav. Immun.* 54, 158–169. doi: 10.1016/j.bbi.2016.02.002

McDaniel, M. (2016). *Women in Gaming: A Study of Female Players' Experiences in Online FPS Games*. Available online at: https://aquila.usm.edu/honors_theses/427 (accessed September 20, 2019).

McIntosh, P. (2003). "White privilege: unpacking the invisible knapsack," in *Understanding Prejudice and Discrimination* (McIntosh, Peggy; Wellesley, MA: Wellesley College Center for Research on Women; McGraw-Hill), 191–196.

McLean, L., and Griffiths, M. D. (2019). Female gamers' experience of online harassment and social support in online gaming: a qualitative study. *Int. J. Ment. Health Addict.* 17, 970–994. doi: 10.1007/s11469-018-9962-0

Media and Entertainment (2015). *How Diverse Are Video Gamers–and the Characters They Play? Nielsen*. Available online at: http://www.nielsen.com/us/en/insights/news/2015/how-diverse-are-video-gamers-and-the-characters-they-play.html (accessed September 20, 2019).

Miller, G. E., Chen, E., and Zhou, E. S. (2007). If it goes up, must it come down? Chronic stress and the hypothalamic-pituitary-adrenocortical axis in humans. *Psychol. Bull.* 133, 25–45. doi: 10.1037/0033-2909.133.1.25

Moghaddam, F. M., Taylor, D. M., Ditto, B., Jacobs, K., and Bianchi, E. (2002). Psychological distress and perceived discrimination: a study of women from India. *Int. J. Intercult. Relations* 26, 381–390. doi: 10.1016/S0147-1767(02)00012-3

Mosley, D. V., Owen, K. H., Rostosky, S. S., and Reese, R. J. (2017). Contextualizing behaviors associated with paranoia: perspectives of black men. *Psychol. Men Masculinity* 18, 165–175. doi: 10.1037/men0000052

Mukherjee, S. (2018). Playing subaltern: video games and postcolonialism. *Games Cult.* 13, 504–520. doi: 10.1177/1555412015627258

Nadal, K. L., Wong, Y., Issa, M. A., Meterko, V., Leon, J., and Wideman, M. (2011). Sexual orientation microaggressions: processes and coping mechanisms for lesbian, gay, and bisexual individuals. *J. LGBT Issues Couns.* 5, 21–46. doi: 10.1080/15538605.2011.554606

Neal-Barnett, A. M., and Crowther, J. H. (2000). To be female, middle class, anxious, and black. *Psychol. Women Q.* 24, 129–136. doi: 10.1111/j.1471-6402.2000.tb00193.x

Noh, S., and Kaspar, V. (2003). Perceived discrimination and depression: moderating effects of coping, acculturation, and ethnic support. *Am. J. Public Health* 93, 232–238. doi: 10.2105/AJPH.93.2.232

Oei, A. C., and Patterson, M. D. (2013). Enhancing cognition with video games: a multiple game training study. *PLoS ONE* 8:e0058546. doi: 10.1371/journal.pone.0058546

Olson, C. K. (2010). Children's motivations for video game play in the context of normal development. *Rev. Gen. Psychol.* 14, 180–187. doi: 10.1037/a0018984

Ortiz, S. M. (2019). "You can say i got desensitized to it": how men of color cope with everyday racism in online gaming. *Sociol. Perspect.* 62, 572–588. doi: 10.1177/0731121419837588

Oyserman, D. (2008). Racial-ethnic self-schemas: multidimensional identity-based motivation. *J. Res. Pers.* 42, 1186–1198. doi: 10.1016/j.jrp.2008.03.003

Pascoe, E. A., and Richman, L. S. (2009). perceived discrimination and health: a meta-analytic review. *Psychol. Bull.* 135, 531–554. doi: 10.1037/a0016059

Passmore, C. J., Birk, M. V., and Mandryk, R. L. (2018). "The privilege of immersion: racial and ethnic experiences, perceptions, and beliefs in digital gaming," in *Conference on Human Factors in Computing Systems - Proceedings* (New York, NY: ACM Press), 1–19. doi: 10.1145/3173574.3173957

Passmore, C. J., and Mandryk, R. (2018). "An about face," in *Association for Computing Machinery (ACM)* (New York, NY: Association for Computing Machinery), 365–380. doi: 10.1145/3242671.3242711

Passmore, C. J., Yates, R., Birk, M. V., and Mandryk, R. L. (2017). "Racial diversity in indie games: patterns, challenges, and opportunities," in *CHI PLAY 2017 Extended Abstracts - Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play* (Association for Computing Machinery, Inc), 137–151. doi: 10.1145/3130859.3131438

Pennebaker, J. W., Kiecolt-Glaser, J. K., and Glaser, R. (1988). Disclosure of traumas and immune function: health implications for psychotherapy. *J. Consult. Clin. Psychol.* 56, 239–245. doi: 10.1037/0022-006X.56.2.239

Petitmengin, C. (2006). Describing one's subjective experience in the second person: an interview method for the science of consciousness. *Phenomenol. Cogn. Sci.* 5, 229–269. doi: 10.1007/s11097-006-9022-2

Phinney, J. S., and Ong, A. D. (2007). Conceptualization and measurement of ethnic identity: current status and future directions. *J. Couns. Psychol.* 54, 271–281. doi: 10.1037/0022-0167.54.3.271

Rankin, Y. A., and Han, N. E. (2019). "Exploring the plurality of black women's gameplay experiences," *in Conference on Human Factors in Computing Systems - Proceedings* (New York, NY: Association for Computing Machinery). doi: 10.1145/3290605.3300369

Reinecke, L. (2009). Games and recovery: the use of video and computer games to recuperate from stress and strain. *J. Media Psychol.* 21, 126–142. doi: 10.1027/1864-1105.21.3.126

Rideout, V., Alexis, L., and Ellen, W. (2011). *Children Media and Race Media Use Among White, Black, Hispanic and Asian American Children*. Northwestern University. 1–24.

Riles, J. M., Behm-Morawitz, E., Shin, H., and Funk, M. (2019). The effect of news peril-type on social inclinations: a social group comparison. *J. Mass Commun. Q.* 107769901985563. doi: 10.1177/1077699019855633

Romero, A. J., Edwards, L. M., Fryberg, S. A., and Orduña, M. (2014). Resilience to discrimination stress across ethnic identity stages of development. *J. Appl. Soc. Psychol.* 44, 1–11. doi: 10.1111/jasp.12192

Schulkin, J., Gold, P. W., and McEwen, B. S. (1998). Induction of corticotropin-releasing hormone gene expression by glucocorticoids: implication for understanding the states of fear and anxiety and allostatic load. *Psychoneuroendocrinology* 23, 219–243. doi: 10.1016/S0306-4530(97)00099-1

Scott, L. D. (2004). Correlates of coping with perceived discriminatory experiences among African American adolescents. *J. Adolesc.* 27, 123–137. doi: 10.1016/j.adolescence.2003.11.005

Sellers, R. M., Morgan, L., and Brown, T. N. (2001). A multidimensional approach to racial identity: implications for African American children. *Forg. links Clin. Perspect. African Am. Child.* 23–56.

Sellers, R. M., Rowley, S. A. J., Chavous, T. M., Shelton, J. N., and Smith, M. A. (1997). Multidimensional inventory of black identity: a preliminary investigation of reliability and construct validity. *J. Pers. Soc. Psychol.* 73, 805–815. doi: 10.1037/0022-3514.73.4.805

Sellers, S. (2003). *The Hélène Cixous Reader*. London: Routledge. doi: 10.4324/9780203408483

Sengün, S., Salminen, J., Mawhorter, P., Soon-Gyo, J., and Jansen, B. J. (2019). "Exploring the relationship between game content and culture-based toxicity," in *HT 2019 - Proceedings of the 30th ACM Conference on Hypertext and Social Media* (Association for Computing Machinery, Inc), 87–95. doi: 10.1145/3342220.3343652

Shaw, A. (2012). Do you identify as a gamer? *Gender, race, sexuality, and gamer identity. New Media Soc.* 14, 28–44. doi: 10.1177/1461444811410394

Shaw, A. (2015). *Gaming at the Edge: Sexuality and Gender at the Margins of Gamer Culture*. University of Minnesota Press: Minnesota Scholarship Online. doi: 10.5749/minnesota/9780816693153.001.0001

Shaw, A., and Friesem, E. (2016). *Where is the Queerness in Games? Types of Lesbian, Gay, Bisexual, Transgender, and Queer Content in Digital Games*. Available online at: http://ijoc.org (accessed September 20, 2019].

Shaw, A., Lauteria, E. W., Yang, H., Persaud, C. J., and Cole, A. M. (2019). Counting queerness in games: trends in LGBTQ digital game representation, 1985-2005. *Int. J. Commun.* 13, 1544–1569.

Skinner, E. A., Edge, K., Altman, J., and Sherwood, H. (2003). searching for the structure of coping: a review and critique of category systems for classifying ways of coping. *Psychol. Bull.* 129, 216–269. doi: 10.1037/0033-2909.129.2.216

Spade, D. (2015). *Normal Life : Administrative Violence, Critical Trans Politics, and the Limits of Law*. Durham, NC: Duke University Press. doi: 10.1215/9780822374794

Spiel, K., Frauenberger, C., Keyes, O. S., and Fitzpatrick, G. (2019). Agency of autistic children in technology research - a critical literature review. *ACM Trans. Comput. Interact.* 26:38. doi: 10.1145/3344919

Srauy, S. (2019). Professional norms and race in the north American video game industry. *Games Cult.* 14, 478–497. doi: 10.1177/1555412017708936

Stanton, A. L., Danoff-Burg, S., Cameron, C. L., and Ellis, A. P. (1994). Coping through emotional approach: problems of conceptualization and confounding. *J. Pers. Soc. Psychol.* 66, 350–362. doi: 10.1037/0022-3514.66.2.350

Stets, J. E., and Carter, M. J. (2011). The moral self: applying identity theory. *Soc. Psychol. Q.* 74, 192–215. doi: 10.1177/0190272511407621

Szymanski, D. M., and Henrichs-Beck, C. (2014). Exploring sexual minority women's experiences of external and internalized heterosexism and sexism and their links to coping and distress. *Sex Roles* 70, 28–42. doi: 10.1007/s11199-013-0329-5

Treadway, M. T., Cooper, J. A., and Miller, A. H. (2019). Can't or won't? immunometabolic constraints on dopaminergic drive. *Trends Cogn. Sci.* 23, 435–448. doi: 10.1016/j.tics.2019.03.003

Trnka, R., and Smelik, V. (2020). Elimination of bias in introspection: methodological advances, refinements, and recommendations. *New Ideas Psychol.* 56:100753. doi: 10.1016/j.newideapsych.2019.100753

Tynes, B. M., Rose, C. A., Hiss, S., Umaña-Taylor, A. J., Mitchell, K., and Williams, D. (2014). Virtual environments, online racial discrimination, and adjustment among a diverse, school-based sample of adolescents. *Int. J. Gaming Comput. Simulations* 6, 1–16. doi: 10.4018/ijgcms.2014070101

Umaña-Taylor, A. J., Tynes, B. M., Toomey, R. B., Williams, D. R., and Mitchell, K. J. (2015). Latino adolescents' perceived discrimination in online and offline settings: an examination of cultural risk and protective factors. *Dev. Psychol.* 51, 87–100. doi: 10.1037/a0038432

Vella, K., Klarkowski, M., Turkay, S., and Johnson, D. (2020). Making friends in online games: gender differences and designing for greater social connectedness. *Behav. Inf. Technol.* 39, 917–934. doi: 10.1080/0144929X.2019.1625442

Villani, D., Carissoli, C., Triberti, S., Marchetti, A., Gilli, G., and Riva, G. (2018). Videogames for emotion regulation: a systematic review. *Games Health J.* 7, 85–99. doi: 10.1089/g4h.2017.0108

Walker, G. J., Deng, J., and Dieser, R. B. (1977). *Leisure Sciences. Taylor & Francis.* Available online at: https://www.academia.edu/8381342/Ethnicity_Acculturation_Self-Construal_and_Motivations_for_Outdoor_Recreation (accessed September 20, 2019).

Wijman, T. (2018). *Global Games Market Revenues 2018. Newzoo, Newzoo.* Available online at: https://newzoo.com/insights/articles/global-games-market-reaches-137-9-billion-in-2018-mobile-games-take-half/ (accessed September 20, 2019).

Williams, D., Martins, N., Consalvo, M., and Ivory, J. D. (2009). The virtual census: representations of gender, race and age in video games. *New Media Soc.* 11, 815–834. doi: 10.1177/1461444809105354

Willner, P. (1997). Validity, reliability and utility of the chronic mild stress model of depression: a 10-year review and evaluation. *Psychopharmacology.* 134, 319–329. doi: 10.1007/s002130050456

Yee, N., and Bailenson, J. (2007). The proteus effect: the effect of transformed self-representation on behavior. *Hum. Commun. Res.* 33, 271–290. doi: 10.1111/j.1468-2958.2007.00299.x

Yoo, H. C., and Lee, R. M. (2005). Ethnic identity and approach-type coping as moderators of the racial discrimination/well-being relation in Asian Americans. *J. Couns. Psychol.* 52, 497–506. doi: 10.1037/0022-0167.52.4.497

Young, E. S., Farrell, A. K., Carlson, E. A., Englund, M. M., Miller, G. E., Gunnar, M. R., et al. (2019). The dual impact of early and concurrent life stress on adults' diurnal cortisol patterns: a prospective study. *Psychol. Sci.* 30, 739–747. doi: 10.1177/0956797619833664

Yuen, F., Ranahan, P., Linds, W., and Goulet, L. (2019). Leisure, cultural continuity, and life promotion. *Ann. Leis. Res.*, 1–22. doi: 10.1080/11745398.2019.1653778

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Computer Science

ORIGINAL RESEARCH
published: 22 October 2020
doi: 10.3389/fcomp.2020.530309

Check for updates

Technology Facilitates Physical Activity Through Gamification: A Thematic Analysis of an 8-Week Study

Dennis L. Kappen[1]*, Pejman Mirza-Babaei[2] and Lennart E. Nacke[3,4]

[1] Faculty of Applied Sciences and Technology, Industrial Design, Humber Institute of Technology and Advanced Learning, Etobicoke, ON, Canada, [2] Applied User Experience Research Lab, Faculty of Business and Information Technology, Ontario Tech University, Oshawa, ON, Canada, [3] HCI Games Group, Games Institute, University of Waterloo, Waterloo, ON, Canada, [4] Stratford School of Interaction Design and Business and Department of Communication Arts, University of Waterloo, Waterloo, ON, Canada

Gamification has enabled technology to facilitate behavior change through increasing the engagement and motivation of people in health and wellness domains. While research on physical activity (PA) and why older adults engage in PA exists, there are not many long-term studies on how gamification influences technology use and adherence to PA by older adults. We conducted a synchronous, 8-week, experimental study with older adults in the 50+ age group. Participants were randomized into three groups: Gamified technology, non-gamified technology and a control group. We conducted a weekly semi-structured interview with them focused on their PA motivations, setting up goals, accomplishments, fears or barriers, (immediate and long-term) rewards, and tracking in PA. Thematic analysis (TA) of the interview data showed these distinct variations in themes for the three groups over the 8-week period. This indicates that motivational affordances or gamification elements can be customized for older adults to suit their current health conditions and PA participation barriers. We define gamification design guidelines for PA motivation of older adults based on self-determination theory, setting up progressive goals, accomplishments to track PA quality, intangible rewards, and activity tracking.

Keywords: motivational affordances, gamification (GAM), physical activity, older adults, thematic analysis, adaptive engagement, persuasive technology, personalization and customization

OPEN ACCESS

Edited by:
Marc Herrlich,
University of Kaiserslautern, Germany

Reviewed by:
Jan David Smeddinck,
Newcastle University, United Kingdom
José Baptista Coelho,
University of Lisbon, Portugal

****Correspondence:***
Dennis L. Kappen
dennis.kappen@humber.ca

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Computer Science

Received: *28 January 2020*
Accepted: *07 September 2020*
Published: *22 October 2020*

Citation:
Kappen DL, Mirza-Babaei P and Nacke LE (2020) Technology Facilitates Physical Activity Through Gamification: A Thematic Analysis of an 8-Week Study. Front. Comput. Sci. 2:530309. doi: 10.3389/fcomp.2020.530309

INTRODUCTION

Research emphasizes positive aspects of gaming and technology for seniors (Zelinski and Reyes, 2009; Gerling et al., 2010, 2011, 2012; Marston, 2013; Bleakley et al., 2015; Kappen et al., 2016). The increase in the popularity of health and fitness apps provides users with the ability to track their activities, sleep patterns, and caloric intake (Lister et al., 2014). While technology artifacts like Fitbit ("FitBit", 2015) and Google Fit ("Google Fit", 2015) have enabled people to track their PA, it is important to understand the factors that motivate older adults' to engage in PA.

The factors that motivate older adults to participate in PA are influenced by their age-related impairments and health related challenges (Schutzer and Graves, 2004; Dacey et al., 2008; Fife, 2008). Motivation to engage in PA is also influenced by their own personalities, attitudes toward technology and social interaction (Kuroda et al., 2012). While PA can be a chore or a mundane activity for many, usage of game elements, can help with behavior change (Deterding et al., 2011; Kappen and Nacke, 2013; Hamari and Koivisto, 2015).

Much research is needed to provide insights into designing and tailoring fitness programs for older adults from a motivation and goals-based initiative as opposed to a point-based system. In this manuscript, the terms gamification elements and motivational affordances are used interchangeably. While research has explored older adults PA and motivation, to the best of our knowledge, there is limited research with the application of gamification elements or motivational affordances over longer durations. Prior analysis of interview data from the 8-week study using grounded theory (GT) resulted in *A Theory of Motivational Affordances for Older Adults*' (Kappen et al., 2018). This theory posited the relevance of intrinsic, extrinsic and feedback affordances to facilitate PA in older adults to sustain engagement in PA over longer periods of time. Customization and personalization in the context of the design of PA technology relate to the system to be designed (tailored) to suit the age-related infirmities of older adults mirrored with their abilities and the flexibility of the system to match their short-term and long-term PA goals, respectively (Kappen, 2017; Kappen et al., 2018, 2019). However, in order to identify intrinsic, extrinsic and feedback *gamification elements* or *motivational affordances*) specific to older adults PA, we used the findings from GT as a framework to develop a codebook and re-analyse the interview data using thematic analysis (TA). Our paper addresses this gap and investigates the implementation of motivational affordances through gamification technologies in the domain of older adults' PA, provides a thematic mapping of affordances (**Figure 2**) and proposes design guidelines for technology facilitation of PA (section Technology Facilitation of PA).

RELATED WORK

Motivation to Participate in PA

While adults with age 65 years and older are categorized as seniors or elderly, many studies in the canon of research on PA interventions have qualified older adults to be 50 years and older (King et al., 1998; King, 2001; King and King, 2010; Weber and Sharma, 2011). Research on this demographic is important because many individuals ≥50 years do not meet the national guidelines for PA (Brawley et al., 2003). Although many researchers have studied motivation to participate in PA (Schutzer and Graves, 2004; Dacey et al., 2008; Mullen et al., 2011; Chase, 2013; Bethancourt et al., 2014; Stathi et al., 2014), there is limited research on the intersection of PA, motivation and technology facilitation for older adults PA.

A long-term study of computer tailored PA intervention for older adults carried out on adults over 50 years of age was effective in inducing long-term behavioral changes in PA of older adults (van Stralen et al., 2011). The efficacy of print-based intervention was stronger than web-based intervention in adults over 50 years, measured over a 12-month period indicating the need for improved web-based interventions for better sustainability of PA over the long-term (Peels et al., 2013). Research also indicates that baby boomers those aged 50–64 are increasingly more adept at using web applications and technology artifacts (Keenan, 2009; Irvine et al., 2013; Mouton and Cloes, 2013). This indicates the need to explore the usage of novel strategies like gamification applied to the PA domains.

Gamification and Older Adults PA

While research has shown that gamification facilitates the interjection of motivational affordances into mundane activities as exercise routines (Lister et al., 2014; Kappen et al., 2017), it is critical to investigate the specific types of motivational affordances that foster older adults PA. Different approaches have been taken with regard to the promotion PA for older adults. Non-commercial games like *UbiFit Garden* provided feedback in the form of flowers based on daily PA (Consolvo et al., 2006). *Flowie* provided feedback on increased number of daily steps taken by the participants (Albaina et al., 2009) and aimed to encourage PA through persuasive technology intervention (Fan et al., 2012). Mobile apps also promoted activity, focusing on individually tailored feedback (Geurts et al., 2011). While playful persuasive solutions (Romero et al., 2010), embodied gaming (Aarhus et al., 2011), and augmented gaming (Mahmud et al., 2010) facilitated fun and socially engaging activities, key intrinsic and extrinsic reasons for sustaining older adults motivation in these activities have not been identified. Therefore, our research investigation on identifying intrinsic and extrinsic motives of older adults' usage of PA technology addresses this need.

METHOD

While many PA technology were available commercially, Spirit50 was selected as a gamified intervention because it was specifically designed for older adults taking into consideration their age-related impairments and health conditions. Spirit50 incorporated the following gamification elements: goal definition (quest), daily challenges (sub-goals), goal progression meter, points and badges (stars) as motivational affordances. Pedometers was used as a non-gamified second type of PA intervention (Kappen, 2017; Kappen et al., 2019). We acknowledge that our prior publication (Kappen et al., 2018) was presented as a summary paper of the same mixed-method study which incorporated quantitative measures and collected interview data over an 8-week period. GT method was used to analyse the interview data in our prior publication (Kappen et al., 2018) which focused on summarizing the quantitative and qualitative data. However, in this manuscript, we focus only on the qualitative data and present the analysis of the interview data using TA. Therefore, an abridged version of the study design is presented in this section to provide context to the TA.

Participants

While many challenges exist with older adult's health, motivation and PA, our investigation specifically focused on determining whether gamification as a tool, could help older adults with an active lifestyle sustain, maintain, and even initiate new PA over longer durations of time. Therefore, our focus involved researching the needs and motivations specific to older adults

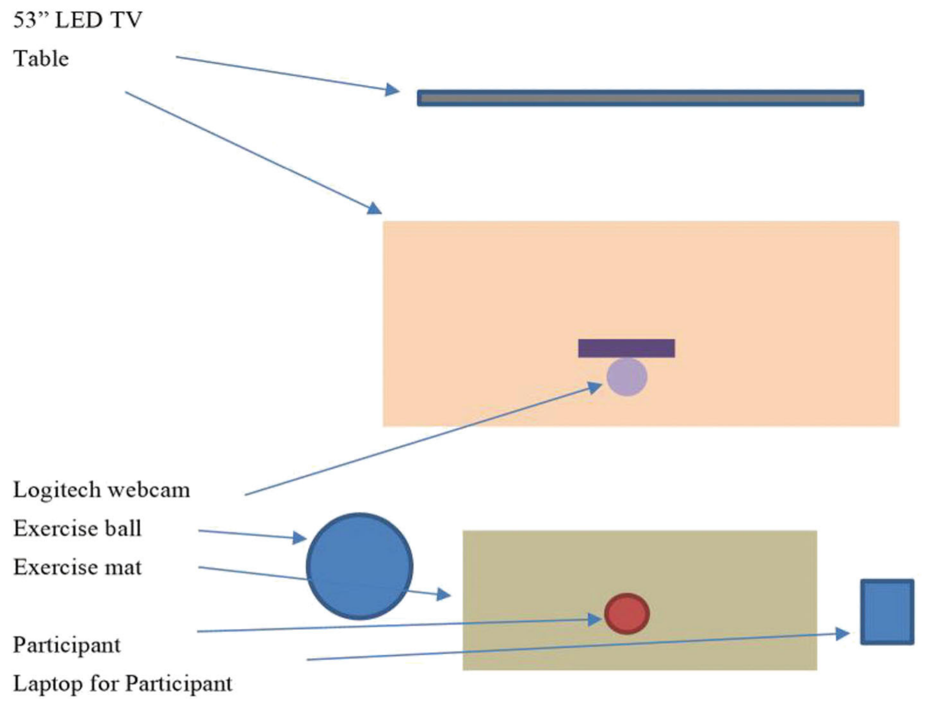


FIGURE 1 | Experimental setup for Group 1.

with an active lifestyle, identifying motivational affordances with the interest of developing guidelines for the design of gameful PA technology. Participants (unpaid) with an active lifestyle were informed about the three-arm study and randomly allocated to three groups. Eligibility was also ascertained with the PA Readiness Questionnaire (PAR-Q)[1]. Current PA intensity levels of all participants was ascertained using the International PA Questionnaire (IPAQ), a validated instrument (IPAQ Group, 2005; Hagstromer et al., 2006). Selection criteria were: (1) age 50+, (2) active lifestyle, and (3) minimal computer literacy.

Procedure

Based on prior literature, a minimum effective exercise program for habit formation was 6 weeks (van der Bij et al., 2002; Martinson et al., 2010; Kaushal and Rhodes, 2015). Therefore, PA over an 8-week duration was studied in this experimental study. This study had a total of thirty participants (**Supplementary Table 8**) and randomized into one of three conditions:

1. Physically active and use a gamified PA app (Spirit50)
2. Physically active and use a pedometer
3. Physically active (control group)

Group 1: Spirit50

Participants were provided with a login and password for Spirit50. Weekly exercise sessions and interviews were carried out at the LiveLabs—Humber College. Setup comprised of a laptop, 53" screen and a Samsung webcam as shown in **Figure 1**. Participants selected their *long-term goals and specific goals* (Kappen, 2017), barriers to doing PA and answered questions regarding their health situations. This enabled the gamified application to identify a low, medium or high intensity exercise routine for an 8-week period. All participants selected a common specific goal (i.e., "Get up and down off the floor with ease").

Group 2: Pedometer (Non-gamified)

A standard clip-on pedometer was provided to all participants from this group and asked to continue their PA as normal. The pedometer screen provided information on the number of steps, distance walked, calories burnt, and time taken. Interviews and questionnaires were done on a weekly basis or at times in an online format.

Group 3: Control

Participants continued with their normal activities. They were interviewed and completed the questionnaire in-person or through a link to the long-form questionnaire on a weekly basis.

Interview Protocol

Semi-structured interviews focussed on eliciting answers related to PA which were as follows:

1. What was your *motivation* to do PA this week?
   1.1. Were there any triggers that helped you be motivated to do the PA this week?

[1]("Physical Activity Readiness Questionnaire - PAR-Q" 2002).

2. How do you *decide on setting up goals* to help you do PA or exercises?
3. Were there any accomplishments or *feeling of accomplishments* this week (completion of a task is also an accomplishment)?
4. With regards to PA, were there any *fears or barriers* that you faced this week?
5. Were there any *rewards* (tangible or intangible) that you received or felt/received this week?
6. What kinds of *tracking* information or feedback would you have liked to receive?

These semi-structured interview questions were used as a starting point of discussion with participants from all three groups.

Data Collection

Data were gathered in the form of audio recordings, skype interviews, or answers to long form questionnaires.

THEMATIC ANALYSIS

TA is a common qualitative analytic method which involves the identification of themes and patterns within the data (Boyatzis, 1998; Alhojailan and Ibrahim, 2012). While qualitative data analysis is interpretive in nature, TA provides a structured method for analysis through six stages which are: (1) familiarizing yourself with the data, (1) generating the initial codes, (3) searching for themes, (4) reviewing the themes, (5) defining and naming the themes, and (6) producing the report (Braun and Clarke, 2006). To the best of our knowledge, while the TA method in qualitative research has evolved over time, there is no single literature that outlines the various aspects and critical stages of this method. This method has also evolved in analysis and interpretation by many researchers. Therefore, we had to refer to multiple sources to support the rigorous and methodical manner of the way in which TA for this study was conducted. Overall, a combined technique of deductive and inductive thematic analysis used the data-driven inductive approach to define emergent themes and a deductive a priori template of codes/research questions to formulate categorizations (Boyatzis, 1998; Fereday and Muir-Cochrane, 2006). This combined method helped to relate the data to the six semi-structured questions while providing the flexibility of themes to emerge from the data inductively within each question category. The definition of themes (**Supplementary Table 1**) has been adapted from Fereday and Muir-Cochrane (2006) to frame the development of the Codebook (**Supplementary Tables 2–7**) and was used to code 20% of the data to establish inter-rater reliability and code the remaining data.

DATA ANALYSIS

Participant Demographics

While the PAR-Q instrument was used to qualify all participants, baseline PA levels of participants were determined using the IPAQ instrument and based on metabolic equivalent tasks (MET) (IPAQ Group, 2005). **Supplementary Table 8** indicates details of participant information in the three groups.

MET scores are categorized as low, moderate (at least 600 MET-min/week) and high (at least 3,000 MET-min/week) (IPAQ Group, 2005; Hagstromer et al., 2006). Based on this, participants in all three groups had high levels of PA (**Supplementary Table 8**), also qualifying them as *active lifestylers* (Kappen et al., 2016, 2017).

Interviews

Audio recordings ($n = 100$, $t_{av} = 15\,\text{min}$) of participant interviews from the three groups were transcribed using Transcribe (https://transcribe.wreally.com/). Answers to interview questions provided in written, online data or email format were collated into six spreadsheets.

Codebook Design

Themes for PA

To allow for better clarity and simplicity of usage of the codebook, the themes relevant to each question was represented in a separate table. Using the inductive and deductive method proposed by Fereday and Muir-Cochrane (2006), the following six questions were used as deductive generalizable categories:

1. Motivation for PA
2. Setting up goals
3. Feeling of Accomplishments
4. Fears and barriers
5. Rewards and PA
6. Tracking of PA

Operationalization

First author's analysis of the interview data using the GT method resulted in a list of themes (axial codes) that was published prior (Kappen et al., 2018). However, the same dataset when analyzed using TA resulted in the same set of themes inductively in each of the question categories and a few additional themes in specific question categories. Therefore, a second coder was used to independently evolve the themes for an initial 1% of the dataset from each question category. Differences in interpretations were discussed, resolved and explanations were noted regarding the method used to resolve such differences. Essentially, transcripts were coded line-by-line to break up the data into its component parts or properties (Charmaz, 2006; Corbin and Strauss, 2015). Open coding was done on each sentence of the transcripts to identify the interpreted meaning of the interview data into phrases that represented each sentence by the participant (Corbin and Strauss, 2015). The above process was done for all participant responses for each of the six questions. These open codes were aggregated into a higher category or themes (Boyatzis, 1998; Alhojailan and Ibrahim, 2012). The interview responses were sorted based on the group number and themes to gather interview responses and to evolve characteristics of the categories.

The following rules were used to identify themes for the six question categories.

1. Each emergent theme was specific to either of the six question categories.
2. For each list of participant responses specific to the above question category, themes were allocated to the responses.

3. Once themes were assigned to all participant responses from one question category, the set of transcripts for the next category was coded similarly.

These rules were used to design a nuanced codebook for further analysis of the entire dataset by the first author and 20% of the dataset from each question category by the second coder as explained in section Evaluating the Codebook. Details of the codebook items for each question category and themes (themes and open codes), definitions and properties specific to each group are indicated in (**Supplementary Tables 2–7**).

Evaluating the Codebook

To review the operationalization of the codebook against participants' responses, as a pilot, another researcher (second coder) coded 1% of the data from each of the six question categories. Cohen's Kappa coefficient (Cantor and Lee, 1996) and Krippendorff's Alpha (De Swert, 2012) was above 80% indicating good inter-rater reliability (Fereday and Muir-Cochrane, 2006; Guest et al., 2012). Based on this review, a few explanations were added to the "descriptions" column of the codebook. Subsequent to this, 20% of the data were coded by the second coder. Details of the total number of participant responses, data used in pilot coding and for reliability analysis is shown in **Supplementary Table 9**.

Coding Reliability

Reliability in TA is also a measure of predictability of the findings (Guest et al., 2012; Miles et al., 2014) and is determined using inter-coder reliability (Mouter et al., 2012). A random selection of 20% of the data for each question category was coded for themes by a second coder (Cantor and Lee, 1996; De Swert, 2012; Vaismoradi et al., 2016). Additionally, care was taken to ensure that this dataset included responses for each theme and moderator. The k-alpha values above 0.8 and kappa values above 0.66 were considered to be good metrics for ratings by two coders, respectively (Ryan and Bernard, 2000; Guest et al., 2012) (**Supplementary Table 10**).

RESULTS AND DISCUSSION: EXPERIMENTAL STUDY

TA method of the same interview dataset gave us the same themes in the six question categories with a few extra themes in specific question categories. Therefore, while the list of themes is same as our prior publication (Kappen et al., 2018), the TA helped us to identify a detailed list of motivational affordances (gamification elements) to help facilitate older adults PA technology (**Figure 2**). Additionally, the TA also helped to identify a nuanced set of open codes and properties for each specific theme (**Supplementary Tables 2–7** for the Discussion Section). Furthermore, based on the comparison of the themes emergent for the three groups (**Supplementary Material–section Themes From Thematic Analysis**), we discuss the themes influencing PA for the six interview questions. Details of sample participant responses for all groups are indicated in **Supplementary Material (section 8)**.

Motivation for PA

Comparison of themes (**Supplementary Material—section Themes From Thematic Analysis**) indicate many intrinsic motivation categories (concepts) emerging from the open coding process. These categories are discussed in relation to the technology characteristics and applicability of these characteristics.

Accomplishing a Goal

Participants in Group 1 were motivated by several factors: completing a goal with ease; realizing health improvements and being active. They were inspired by in-app progress reports and did outdoor activities to increase their level of PA. Multitasking to do app activities and household chores, and outdoor activities pushed them to continue to do more (P31). They felt energized by the routines provided by the app (P08). The app introduced the delineation between immediate goals and long-term goals, which allowed participants to acknowledge the value of doing simple tasks, doing short bursts of exercise routines of various intensities to help them feel like accomplishing a lot. The app helped applaud participants achievements of small steps of exercise routines which were bigger successes from a feel-good perspective, which was similar to the result from the study about beliefs around PA among older adults in rural Canada (Schmidt et al., 2016).

Group 2 participants were motivated by factors such as accomplishing challenges with ease, hitting pedometer targets, doing outdoor or fitness activities. Additionally, inspiring situations such as reminiscing about former fitness levels, lack of PA and, increase in pedometer numbers enabled participants to persevere with PA. Increased awareness of the benefits of PA were also triggers for accomplishing a goal, similar to the results seen in a prior study (Jancey et al., 2009).

For Group 3 (control), the motivating factors for PA were performing tedious outdoor activities and completing challenges with ease. Additionally, this group allocated more time to outdoor activities because of scheduled morning workouts. This inclination could have been due to increased freedom of choice between many outdoor activities such as walking the dog, playing with grandchildren (P07), swimming (P25), dancing (P27), and participating in aerobic exercises or Tai Chi (P20). This indicated that for Group 1 and Group 2, the presence of the app and the pedometer influenced the participants to do more activities while accomplishing a goal.

Aging Well

Being conscious and accepting of growing older (P04), working out to age gracefully (P16), and the interest to overcoming age-related challenges (P05) served as motivations for PA in Group 1 and Group 2. This theme was not evident in Group 3.

Challenged by Activity

Increasing the intensity of PA routines or exercises, as well as trying to complete hectic and difficult activities, were relevant motivational elements for Group 1. Additionally, being prompted with higher intensity routines, or new exercise routines through the app, provided greater motivation to stay with the app. Combining regular outdoor PA routines like yard work,

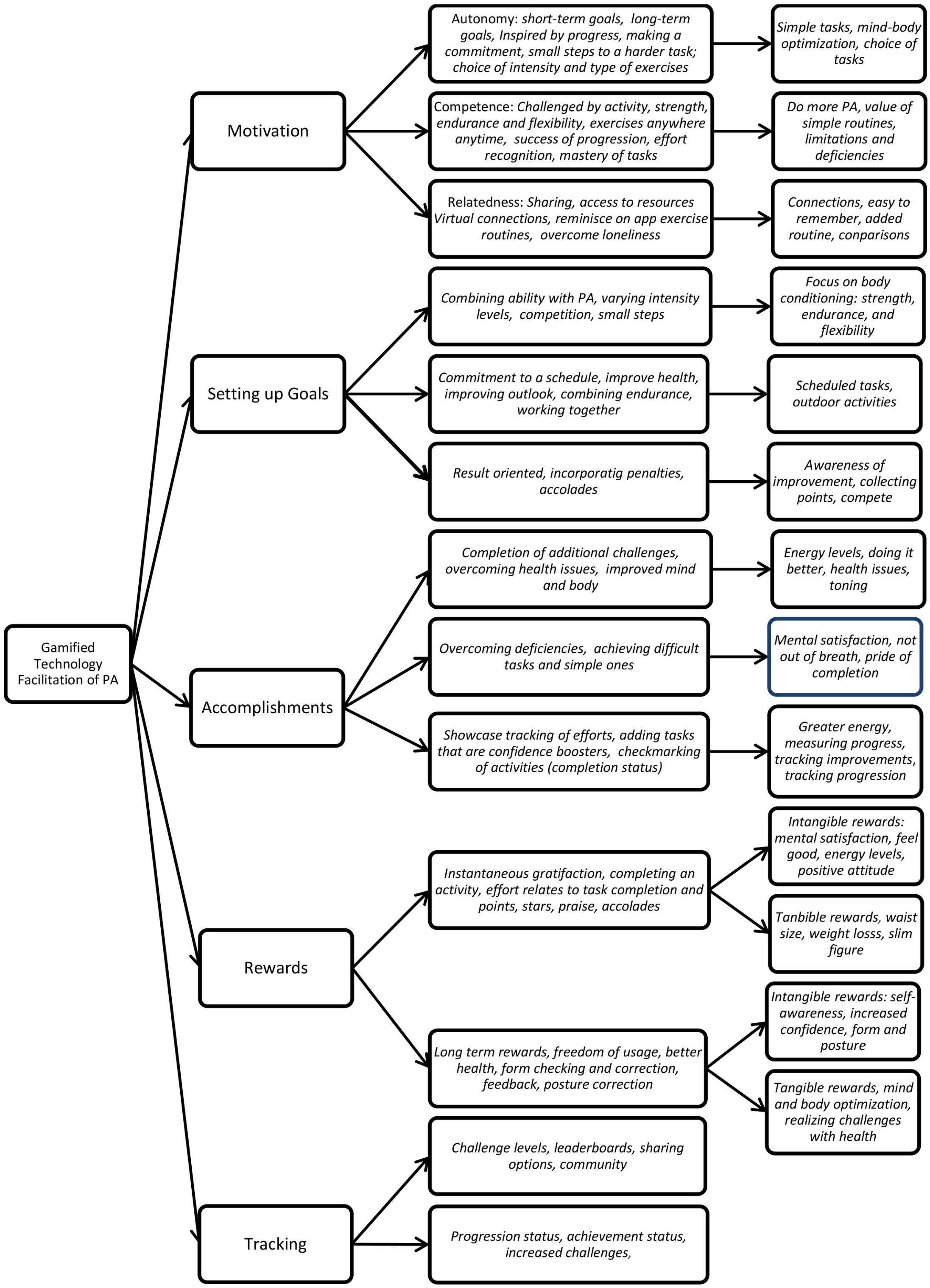


FIGURE 2 | Evidential chain showing motivational affordances from TA.

gardening, or raking with indoor exercise routines helped with sustaining a variety throughout the week. Personal life challenges hindered PA activity (P04) and the stress of preparing for a future hiking activity (P31) added the stress in the week for doing PA. Overcoming lethargy and laziness by using simple exercises (P10) also helped as a motivating element.

In Group 2, overcoming boredom (P13) and lack of patience in dealing with daily chores (P06) fostered walking. The pedometer influenced participants to increase walking distance (P23) and challenges to their existing routine such as going up the hills and adding additional hills.

In Group 3, pushing to the point of pain or aches to achieve predetermined challenges, like preparing for a 12-km run (P25) and the excitement of participating in a competition (P27) were triggers for being challenged by an activity.

Easy Access to Resources

Challenges with distasteful gym environments, flexibility to do the app applications anywhere (P10) in their home environment and any attire served as triggers in this category (P31). Additionally, costs played into the acceptability of online programs. While there is a commitment and a sense of obligation toward an in-home physical trainer, the plausibility of a virtual coach through the app was well-received. Participants from Group 2 and Group 3 did not indicate any motivational triggers in this category.

Enjoying Outdoors

In Group 1, the preferences for doing the app exercises outside the home environment was desirable due to the potential of too many distractions at home (P11, P31). PA outdoors (e.g., climbing in and out of the boat to empty its gas tank) did help to bring back certain routines learned from the app (P04). There was a general consensus to do outdoor activities such as hiking, walking, bike-riding, playing tennis, or running when the weather was good (P01, P11, P16). These activities also helped participants to reminisce about a few routines practiced from the app. This allowed for easy recall of app routines and participants realized its value while working outdoors (P08).

In Group 2, participants were open to enjoy the outdoors (P18, P21) resulting in more step count on their pedometers, similar to a subconscious side-effect of doing an activity. Group 3 participants did not bother much about any tracking their steps and spent more time enjoying the weather (P12, P15).

Experiences

Participants from Group 1 were interested in feeling good (P01), improve energy levels, increase excitement levels and have a positive outlook toward fitness (P08). Participants felt good about a workout commitment. They did not consider the app complex (P08, P16), but they believed that it definitely helped with the realization of taking small steps toward a bigger goal (P31).

Feeling good to see the numbers increase (pedometer) after a vigorous activity (P23) and doing something for the sake of doing an activity (P18) were common motivation concepts in Group 2.

This theme was not evident in Group 3.

Fear of Being Unhealthy

In Group 1, working on the simplistic app routines was one way of doing something to be active and allay the fears of being inactive and overcome the fear of getting stale (P01). The participants found the app routines to be simple and doable (P04) and helped them overcome their fears of being unable to do these routines (P11).

In Group 2, one participant had a "sick" feeling due to inactivity and decided to move around and add some steps on the pedometer (P18).

Group 3 participants were mostly concerned about inactivity and their future health which was a motivational trigger for them to be active (P15). Predominantly, the fear of being unhealthy due to inactivity was a prime concept that motivated participants to feel like doing something (P25).

Focussing on Appearance

In Group 1, there was a desire to improve their physical appearance by controlling their weight, slim down and overcome obesity (P08). The simple exercise routines from the app encouraged the participants to realize the simplicity of being able to do the app exercises anywhere and anytime (P11, P16, P24). This enabled the app to be used frequently to improve their flexibility and posture (P04).

Participants from Group 2 added more step-count on their pedometers to shed some weight, slim down for a future trip and have a more toned body (P13). Adding more steps with the intention of "shedding some fat" (P02) with the realization that in summer there are "less clothes to hide behind" (P02, P18). This theme was not evident in Group 3.

Focussing on Motivational Affordances

In Group 1, participants acknowledged and accepted the motivational affordances as trigger elements to do more PA. These ranged from monitoring or the desire to measure PA as a means of reassurance and validation for effort being done toward PA (P08). Simple tracking of effort, receiving badges, points, and progression information and comparing effort with peers and/or spouses (P11) were triggers for participants. Participants suggested the addition of motivational slogans and inspirational imagery to help establish aspirational goals (P24). While points and stars seemed inconsequential, it was good to earn (P16) and push further to earn more points. Initially, the existence of points and stars were not acknowledged, however, as they progressed through the weeks (P11), these elements did get noticed and inspired the desire to reach higher levels within the app. Participants took time to realize that the increase in points and stars corresponded to the completion of daily and weekly exercise routines.

In Group 2, it was noticed that participants walked more than their step-count in the first week to achieve daily step targets (P05). One participant who was unaware of projected daily step targets, and was used to doing <1,000 steps/day, pushed herself to achieve more than 10K steps by week 3 (P23). Participants also indicated that step-count, time and distance done, were key triggers to do more PA (P21). Additionally, calories and weight loss information also served as triggers to do more PA (P18).

In Group 3, participants mentioned that while they noticed their time investment in PA (P15), they were keen on weight loss as seen on their bathroom scale (P25).

For a Healthy Lifestyle

Participants in Group 1 were motivated to do PA to live longer with good health benefits, improve mind-body optimization

(P11), be mobile, be more active (P08), fit and healthy (P10). The app exercises were geared toward body flexibility were seen to be stress-relieving and more relaxing which lead to a mind-body optimization (P11) contrary other apps (Fitbit) in the marketplace.

Interestingly, Group 2 participants felt the need to maintain a healthy lifestyle (P02), aspire to live longer (P21), be independent and mobile prompted more walking (P13). This also encouraged them to do PA like gardening, household chores and trekking (P04). Participants indicated that the tracking of steps provided them the reassurance of reaching their PA targets and to their goal of becoming active, fit and healthy.

In Group 3, overall triggers for PA motivation were to lead a normal life (P12), be mobile and healthy (P25).

Freedom of Usage

Participants in Group 1 expressed this to be a value/benefit provided by the Spirit50 app. Key characteristics that allowed for motivated engagement in the app were: ability to understand the steps and timing of the exercise routines. This was because the videos showcased in the app helped to monitor the correctness of the routines and provided flexibility of usage of the app anywhere and anytime and in any attire (home or outdoor). Additionally, simplicity of instructions and simple routines enabled participants to recall routines from memory and do the routines anywhere even without the app. Participants also suggested the inclusion of affordances such as reminder bells, voice commands, timers, and adding a variety of routines to choose from. Working with the app also helped with bringing some routine into one participant's daily life that was full of many incidental things that had to be done (P11).

This theme was not evident in Group 2 and Group 3.

Fun and Recreation

Overall, this category represented more aspects of motivation to do PA from a generic prescriptive as opposed to specific advantages from using the app or the pedometer. A few responses indicated that accomplishment of the scheduled tasks and the surprise element of discovering a new way to do difficult exercise in a simple manner became the fun aspect of learning and interacting with the app (P01, P04, P08, P10).

In Group 2, the frustration of having to endure boring exercise routines from online sources and DVDs' for example, encouraged a participant to do more walking with her dog where the fun part was the changing sceneries (P23). Progression with increased step-count on a daily basis added to the competitive nature of the activity and incorporated the fun element (P02).

In Group 3 the aspect of retirement and availability of unlimited time to do anything anywhere and the freedom to do choose different locations to do PA were motivating and fun elements (P12, P25 P27).

Influenced by the App/Artifact

Flexibility of using the app anywhere and anytime (P31), appreciation of improved ability (P04), trying out different combinations of the app and regular PA activities (P11) were common motivational elements in Group1.

In Group 2, many participants were receptive to using the pedometer and used it as a tool to monitor their steps, calories and distance traveled (P18). Many participants reported increased number of steps on a daily basis leading to a higher average each week. There was a consistent effort to add more steps to improve their prior daily average (P13, P21).

Group 3 participants did not report any influences from technology artifacts such as Fitbit or Nike-Plus, however, relied on their watch as time keepers for their daily routines.

Inspirational Influencers

In Group 1, participants were influenced to do more PA by watching team performances (P11), usage of team PA apps by family members and reviews by online fitness experts (P24). Key influencers were: doctors, coaches, physiotherapists and family members (P11, P24, P31). Participants were willing to do the routines provided in the app which were considered to be low intensity (P29). Participants did not find the app routines to be limited by their personal challenges of arthritis, back pain or poor posture and helped to compliment the recommendations from their doctors or physiotherapists (P31).

In Group 2, participants were also influenced positively by younger persons with (buff) toned bodies in their gym environments' (P02). At the same time, participants felt inadequately fit in comparison to younger persons (P12, P23).

Group 3 participants did not report any such influences but relied on the internet for sharing health and wellness information.

Limitations of Resources

Participants from Group 1 were challenged by expenses for physical trainers, cost of gym memberships and lack of interest in standard routines (P04). They also faced discouragement due to lack of challenging exercise routines and not being able to make time for daily exercises (P24, P26). While change was frowned upon by a few participants, they preferred to have the option of a variety of routines and a choice of higher intensity routines on the app to supplement their regular routines (P08, P24). However, these limitations were, in fact, the reason to motivate them to do PA and improvise the routines themselves.

From Group 2, inclement weather forced the need to use transportation to get to their gym routines, therefore the use of treadmills was a limitation because they could not adhere to their original routines (P02).

No *limitations of resources* were reported by Group 3 participants.

Mental Well-Being

In Group 1, key characteristics such as desire to overcome lethargy (P04), lower stress, fear of boredom (P10), feeling of failure and guilt for not doing anything (P11) served as triggers for motivating PA. Overcoming sedentary activity (P29), overcoming commitment issues toward enrolling in a routine program (P24) were additional triggers. Participants indicated that the app helped to reminisce about past laurels and fitness successes and desired to overcome their inertia to improve their current effort to doing PA (P24). The app helped to serve as a medium to enable a simple and small-step approach (P29) toward appreciating a set of routine exercises.

In Group 2, participants desire to work outdoors and achieve satisfaction of completing a visually pleasing and satisfying result (P05) motivated them to do PA. Additionally, the guilt of not having done any PA for some time (P13) also goaded them mentally to do some PA.

In Group 3, building up one's self-confidence and overcoming the feeling of laziness (P25), and guilt for doing nothing (P09) coerced participants into doing PA.

Routine/Lifestyle

Existing daily routines (P01), self-regulation (P08), and self-monitoring of weekly PA routines (P10) fostered habit formation in Group 1 participants. The app exercises being simple to do and easy to remember provided easy recall during existing scheduled daily activities (P31). It was easy to integrate these app exercises and combine with daily routines (P29) with the added flexibility to increase and lower the intensities at will.

Group 2 and Group 3 participants were set into doing their routine activities of walking and jogging over the past many years leading to habit formation (P05, P18, P27). In Group 2, walking was done routinely and the pedometer helped to reinforce their efforts (P18).

Social Connections

All three groups engaged in motivated PA to connect with people and expand their social network while doing their PA. Social connections were not engendered through the app, as this module was not included into the design for this experimental study. However, this motivational category helped reinforce the value of regular fitness due to the realization that they were not alone (G1), and were not singled out in the attempt to maintain and improve their fitness and health (G1). Many participants in G1 wanted to be left alone to their routine activities' in the gym, but wanted the presence of people around them, even if they were not interested in interacting with them. One participant (P24) also suggested a virtual connection with others through the app so that they did not feel that they were doing the PA alone.

In Group 2, participants preferred to compare their step-count with friends and family members (P13, P18, P23). Group 3 participants used their routine PA activity of walking to meet with friends (P07, P09, P14).

Spontaneous and Subconscious Activity

Participants in Group 1 and 2 were motivated by activities done on the spur of the moment, which were either in their indoor or outdoor activities. Many app exercises afforded the possibility to be done indoors and outdoors. These activities helped generate PA subconsciously leading to more engagement and realization that their effort was worthwhile which was based on the results of the activity. This also helped to overcome the drudgery (P01, P08) imposed by routine regular fitness exercises. Additionally, the app used in the study helped influence a change in thought process by helping them differentiate and recognize that they were doing valuable PA (P04, P08) when doing regular chores.

Treatment for a Health Issue

Motivation to participate in PA was also triggered by this category as indicated by the properties gleaned from the participant responses. There was a general consensus of being forced to considering changing sedentary lifestyle by including simple and general fitness routines to overcome health issues. While the gravity of the health conditions was specific to individual participants, all three groups indicated using PA as an means to provide treatment for specific health issues.

Participants in Group 1 recognized the value of the app in helping them to understand some of their deficiencies such as lack of arm-strength (P16), bad posture (P04), low upper body strength (P08) to mention a few and resolved to do more specialized training to overcome these issues.

Setting Up Goals

The comparison of themes between the groups indicated many intrinsic and extrinsic motivation categories emerging from the open coding process. While many of these categories do not directly relate to the usage of technology artifacts, these categories help with a granular understanding of how and why older adults set up goals for PA.

Combining Exercise Types

Participants in Group 1 were more interested in combining the flexibility of exercise routines from the app with strengthening exercises. They also wanted the option to select different intensity levels from the app so that they could aspire for to more challenging routines. They preferred to combine sports, app exercises and strengthening routines. This indicated the desire to combine endurance, flexibility, and strengthening routines to add variety in the collection of PA activities.

Group 2 participants were interested in combining exercise and following their set routines.

Group 3, participants were keen on continuing their set daily routines due to habit and they did not have to complete a set category of exercises. They were also keen on doing PA so that they felt the result of their effort (felt the burn – P16).

Committing Time for Activity

Participants in Group 1 felt obliged to keep their commitment to stick to a specific routine of activities (8-week study, yoga, Pilates, gym). As a result, they were dedicated to meeting their time commitment, and they also felt they need to show their trainer what they had done. In comparison to a physical trainer, the presence of a virtual coach in the form of the app was welcomed as long as there was a real person on the screen with an expectation of certain work to be done per week and monitoring of weekly progression. They did not want to let down their coach (virtual coach) and felt obliged to do something even though they would not have felt like doing any PA for a specific week.

Enjoying Combination of Activities

This category was established separate from *combining exercise types* category indicated prior due to the experiential aspect of participant engagement. The combination of exercise types and activity types (sports, outdoorsy, and/or indoor) provided the excitement of competing with people. Encouraging results from competing in sports like activity (hockey, golf) helped to improve their interest in setting up goals for future. From the app perspective, it was suggested to have "competing with people" as a goal setting so that the enjoyment was better.

Focussing on Specific Goals

Participants in Group1 were keen on setting up goals based on the focus of what their goals aimed to achieve. This contributed toward acknowledging the specificity of the focus of the goals. Recognizing their deficiencies through the app, many participants were keen on focussing on immediate goals, rather than short-term goals or long-term goals. This is also because of their reasoning of the possibility of immediate achievable results that could be visually monitored or measured.

In Group 2, small achievable goals such as weight loss, improvement in posture were key characteristics of the concepts for focussing on goals. Walking more to increase step count from 500 to 10,000 steps/day in a week time was a measurable achievable focused goal for a participant (P23). Reducing excess weight (P18), improving posture and reducing lower back-pain (P13) were a few of the other specific goals from participants.

Focussing on Appearance

was also a focussed/specific goal for participants in Group1, where participants were keen on '*waist management*' (P11, P24) resulting in either maintaining or reducing one belt buckle position as a measurable specific goal.

Focussing on Motivational Affordances

In Group 1, measuring progress, results, accomplishing something, and getting rewarded for efforts were indicated by participants' to be concepts in helping them set up their goals for PA. Checking off a list of activities, receiving points, and stars served as validation for their efforts. While participants were not initially keen on such rewards, as they became aware of their progression, they began to notice the presence of the point and stars (in the app) and contributed to their feeling good about their efforts.

While many participants in Group 2 indicated that they did not need any badges or rewards, accomplishment of certain tasks (daily walking targets, completing set routines) were rewards in itself. This feeling of achievement was critical to help validate their efforts and helped set up goals for future activities.

Group 3 participants were interested in setting up goals based on doing their daily PA routines such as walking, jogging, tracking distance, and measuring weight loss.

Improving Health Outlook

Setting up goals was also based on participants improving their health outlook. Recognizing benefits of specific PA, reminiscing on past achievable laurels and aspiring to reach past glory helped participants from Group 1 to set up their goals. Overcoming laziness/lethargy to avoid sedentary lifestyle also helped with setting up goals for PA. This enabled participants to remember simple routines and aspiring to be fit and mobile inspired them to do the routines anywhere and anytime.

With participants from Group 2, the challenges of health issues (osteoporosis, arthritis, back pain) forced them to modulate their prior vigorous exercise routines to a more simplified form of walking activity.

Participants from Group 3 wanted to *outlive their pension,* monitor their dietary habits and set up goals based on doing something good to their body.

Increasing Challenges Progressively

Working on muscle groups, improving their stamina, working on muscle groups, trying to hit daily targets were some of the specific concepts that emerged in Group 1 in this category to help set up their goals. The app served as a medium for them to push for more challenging routines, and, made them realize their potential of progressively increasing their challenges for improvement.

Participants in Group 2 were also keen on adding more walk time to their daily routines to see how far they could push themselves.

Working toward increasing walking and measuring heart rate and weight loss were important concepts from Group 3 participants.

Self-Regulating Routines

Many participants in Group 1 (P04, P10, P11, P08) indicated the need to control their own routines from the perspective of having the choice to change up exercise intensity based on the flexibility of their schedules and monitoring results. The app helped with setting a schedule for 8-weeks with routine activities. Lack of the option to increase exercise intensity and choice of exercise types was suggested to be a requirement by the participants.

Participants in Group 2 wanted to exercise more control on the amount of time spent on exercise activities like walking, running or riding a bike. The self-regulated the desire to do more and used the pedometer to monitor their progress each day.

In Group 3 participants preferred to watch videos and exercise routines on the internet and try out different activities based on their ability and fitness level.

Social Interaction

Participants in Group 1 were keen on associating with others for companionship in their journey to better health goals (P11, P24). Participating in walkathons, group exercises comparing within a group and with others were key characteristics for setting up goals. The suggestion was to have the app connect them with comparing levels and accomplishments of others.

In Group 2, the key concepts in this category was to have a workout partner, to share in the pain and the journey of doing PA.

In Group 3, overcoming loneliness walking with a partner, and comparing distances and time taken for PA with the group or with a spouse was welcomed.

Spontaneous and Subconscious Activity

This category emerged mainly in Group 1, where participants considered doing exercises to be a forced activity. The best exercise of PA was when they did not realize that while they were engrossed in doing their daily routines and scheduled activities, they were actually exerting themselves and getting a good work out.

Feeling of Accomplishment

The comparison of themes emerging from within the three groups are as shown in **Supplementary Material** (section Themes From Thematic Analysis). We discuss these categories in relation to participant responses (**Supplementary Material section 8**).

Adding New Challenges

In Group 1, participants were more inclined to feel a sense of accomplishment when noticing an improvement in their flexibility and greater endurance in working out. Feeling confident with existing routines allowed them the opportunity to add new ones and combine different types of PA activities. The simplicity of exercises (app) *"...the exercises are simple and can be done anywhere...(P08)",* also resulted in participants feeling that they could do more from the point of pushing themselves to more exertion and thereby more points. A few added more routines or participated in outdoor activities (P01, P08, P16).

Participants in Group 2 added more walking to their daily routines for the sole interest of adding more steps to the pedometer (P13, P21).

In Group 3, they went about their daily tasks to keep themselves busy and, when suggested by friends, also took part in competitive outdoor activities.

Influencing Activity Through App

Feeling of accomplishment was also supported by this category where participants felt that the app helped identify certain deficiencies or short-comings with their self. Realization of one's ability to do better and feeling energetic were also key properties of this category. This category was specific only to Group 1 participants.

Completing Difficult Challenges

Participants felt that increased intensity of exercise routines, completion of task contributed to the feeling of accomplishment. The aspect of completing a few weeks of the app routines was also a feeling of accomplishment for few participants (P08, P10) because they had started out the program with a lot of skepticism. The app did provide increases in intensity which led to participants feeling more challenged and felt the sense of accomplishment on completion of the activity.

Completing the walking or daily outdoor activities were a few properties from this theme for Group 2.

Completing the marathon, daily tasks and repair work contributed to participants from Group 3 to feel a sense of accomplishment for this category.

Feeling of Mental Satisfaction

In Group 1, understanding ones' body to know that it feels better after doing a workout, feeling tiredness in a good way, contentment at the aspect of doing the PA well, were most common properties in this category of mental satisfaction. From an app perspective, the completion of the difficult app routines gave participants the feeling of mental satisfaction contributing to the feeling of accomplishment.

Participants in Group 2 felt energetic (satisfaction of having the energy) and content after completing the outdoor and the PA. The pedometer contributed to pushing participants to do more and feel the satisfaction of completing the activity.

In Group 3, enthusiasm and feeling of achievement contributed to the feeling of happiness leading to mental satisfaction.

Feeling the Burn

This category was specifically added because it represented a physiological characteristic contributing to the feeling of accomplishment as opposed to mental satisfaction. This category emerged only in Group 1 and 2.

Responses from participants from Group 1 indicated above were in relation to the app exercise routines. While the app exercise routines were low intensity at the start, these responses showed that the intensity of the exercises did increase through the 8-weeks.

Feeling Validated for Efforts

This category was evident only in Group 1. A few participants indicated that commitment forced effort on their part, and the result of the effort was visible in the form of increased energy to do more, measuring progress, and tracking improvements in body condition. The app indicated progression and provided a feeling of accomplishment which helped to validate their efforts in doing PA. Being validated for efforts contributed to a mind-body feeling of accomplishment, and wanting to continue with the 8-week program.

Improving Body Conditioning

Participants in Group 1 were keen on achieving tighter muscles, toning the body, maintaining weight, improving posture to mention a few concepts that emerged from the coding. The usage of the app exercises did give participants the realization that they could do certain types of exercises, which was considered to be difficult. The usage also helped participants to realize that they were not out of breath when climbing stairs as before (P08), provided the feeling of being able to do wall push-ups (P31), and, gave the awareness to improve on posture. The app also indicated progression and maintenance of these routines enabled participants to understand their weakness and work toward improving them.

In Group 2, feeling relaxed and maintaining body weight, feeling better due to "muscle tiredness" contributed to this category leading to a feeling of accomplishment. Participants indicated that quantifying their efforts on the pedometer helped with wanting to do more.

In Group 3, increased stamina, and the feeling of relaxation due to exercises were concepts in this category leading to a feeling of accomplishment.

Improving Confidence

In Group 1, certain app exercise routines like wall push ups, stretch band, and exercise ball seated exercises, helped provide participants with increased confidence in their abilities to do certain routines. The feeling of exercise routines being easy to do with increasing intensities along the weeks also provided them with an understanding of their capabilities.

Participants in Group 2 indicated that the ability to do more steps was a point of discovery, and gave them improved confidence leading to a feeling of accomplishment. For one participant, going from 500 steps to more than 10,000 steps per day (P23) was something that was so surprising. The pedometer helped participants to improve prior targets.

In Group 3, participants were keen on being able to do their routine programs and completing them whenever they had the time.

Improving Health Condition

From the perspective of having a goal and the feeling of accomplishment, participants in Group 1 were interested in overcoming their health conditions like sore hip, back issues, diet control, and improve their posture by doing PA. The flexibility exercise routines from the app helped to relieve stress from their body, and the easy recall of these routines helped them use these routines repeatedly.

Group 3 participants had a feeling of accomplishment when they were able to control certain physiological aspects of their body such as: were only keen on walking to control their sugar level without medication or be physically healthy, monitor heart rate and lowering blood pressure.

Improving Ability

Participants in Group 1 indicated various characteristics of being able to lift higher weights, increased intensities, and ability to do new routines that were not tried prior, contributed to the feeling of accomplishment in this category. Participants indicated that the app showed them new exercise routines, with newer intensities, and the reps and steps showed their progression on the screen. This indicated their increased ability as the weeks progressed.

Participants in Group 2 indicated increased step-counts as they progressed through the weeks (P13, P18, P21).

In Group 3, participants indicated their ability to do the activities that they have been used to doing.

Increasing Independence

While many participants indicated that the feeling of independence stemmed from the ability to navigate routines on one's own without help, and ability to do increased intensity exercises, there was no indication of the app facilitating independence from the point of enabling ability. The only independence aspect that was provided by the app was it afforded the freedom to use the app anywhere and anytime.

Inspiring Motivational Affordances

The resulting codes in this category provided insights into the feeling of accomplishments of participants leveraged through motivational affordances facilitated through technology. Participants from Group 1 indicated that compliments and feedback provided reassurance of their efforts to contribute to the feeling of accomplishment. Compliments were provided in textual format as a visual check mark for completion with a graphical representation of an icon of a person with raised hands indicating a *"hurrah"* for completion of the task. Validation for their efforts was recognizable in their positive attitude toward elements like points, stars, progression information, and graphic representation of their effort and completion of tasks. Furthermore, feedback metrics such as reps and steps and timers for exercise completion contributed to a sense of completion leading to a feeling of accomplishment. These findings suggested that reassurance of efforts and validation of work done are key attributes that contributed to the feeling of accomplishment.

For participants from Group 2, the daily step count and increase in step count contributed to the feeling of accomplishment.

Inspiring Performance

This category emerged in Group 1 which indicated concepts such as attempting to do exercise routines correctly, pushing oneself to do more and trying to reach peak ability, and, doing more than specified in the app routine. This category was also facilitated by the availability of videos and visual feedback provided through the app interface indicating that inspired performance by the participants led to a feeling of accomplishment.

Progressing Through Activities

This category was seen across the three groups of participants. While Group 1 participants relied on the app to showcase their progression through the 8-week study, greater emphasis was placed on adding new challenges to existing routines provided by the app. This indicated that participants' self-measures for progressing through activities also indicated the need to be challenged. The combination of exercise intensities, types of exercises (endurance, flexibility, and strengthening), and interplaying outdoor and indoor activities provided a feeling of accomplishment.

While completing daily routines, and doing more steps in addition to meeting step targets (pedometer) represented concepts for Group 2 in the progressing through activities category.

Group 3 participants indicated that the completion of outdoor household projects and ability to complete asks to be more important measures for progressing through the activity to give them a feeling of accomplishment.

Seeking External Resources

This category emerged in Group 1 where a feeling of accomplishment was indicated by following recommendations from fitness instructors, virtual coaches, yoga instructors, physiotherapists, and online videos.

Social Interaction

There was a mixed response from participants in Group 1 regarding the feeling of accomplishment engendered by social interaction with others. While participants liked to have the presence of people around them in a gym environment, they kept to themselves and went an about doing their own routines. However, the presence of people around them provided the feeling that they were not alone in the battle for fitness and overcoming weight challenges and the feeling of sedentariness.

There were many suggestions for the app to provide virtual connection with other friends and family in an online mode to facilitate the feeling of overcoming loneliness when working out with the app at home or in other places. The presence of people around them had to be acknowledged (physically or virtually); however, interactivity between individuals were kept to a bare minimum level. Feeling of accomplishment was also fostered by working out together but to independent goals as opposed to a combined goal.

For Group 2, the feeling of accomplishment was in comparing workout done by others on the same routine and seeing their own progression. Walking in groups and comparing step counts provided a social interaction and a feeling of accomplishment.

In Group 3, intergenerational play and continuing to walk with friends provided a feeling of accomplishment on a daily basis.

Fears and Barriers

While fears and barriers prevented participation in PA, codes that emerged indicated challenging health conditions, fearing inability, appearance issues, psychological challenges, limitations of resources, lack of confidence, lack of performance, and being afraid of social interactions to be primary categories for this question. However, with regards to technology facilitation, in Group 1, the challenges with smartphones and apps taking on a "Big Brother" (P11) approach leading to invasion of privacy was a major concern. This participant refused to purchase a cellphone due to this barrier and did all the exercise routines on his desktop. Challenges of not having access to technology (desktop, app, or smartphones) at the desired time of wanting to do exercise routines were other barriers related to technology and PA (P04, P08). Compared to the ease of using a mobile device, one participant considered the challenges of having to operate a desktop to be a barrier (P04).

In Group 2, while many embraced technologies and the pedometer, one participant was afraid that monitoring would trigger her past nature of obsessively measuring weight, calorie intake, calories burnt, and even measuring the amount of wine consumed each week (P23).

While the control group participants did not allay any fears with technology; the properties of their fears and barriers to exercise are shown in **Supplementary Material**.

Rewards and PA

From the cognitive evaluation theory, the type of rewards can also help to foster intrinsic motivation and behavior (Deci et al., 1975). Participants indicated a lot of interest in the types of rewards associated with PA. While many rewards expressed by the participants were intrinsically driven, there were many instances of being encouraged for PA due to the presence of extrinsic rewards which included: progression bar, completion of tasks, points, and stars. We discuss these in relation to technology facilitation and PA. In Group 1, many participants were encouraged by immediate rewards, long-term rewards, intangible rewards and tangible rewards (**Figure 3**). While immediate and long-term rewards could be either intrinsic or extrinsic, there were indication of further segregation of these into tangible and intangible rewards.

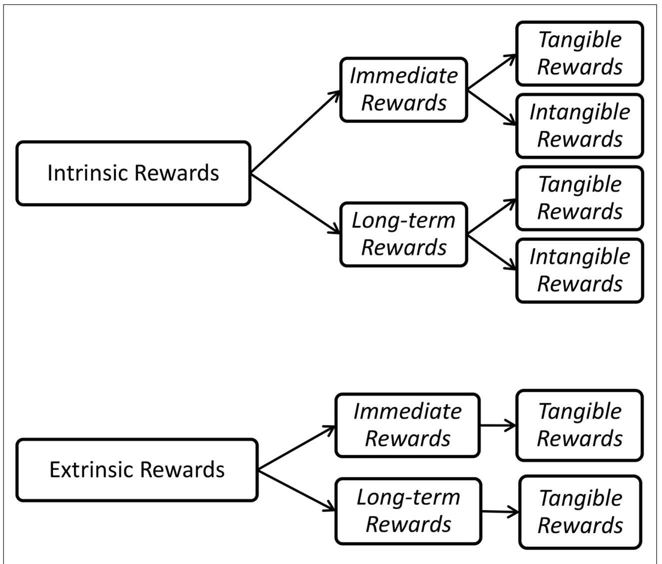


FIGURE 3 | Reward categorization.

Completing an Activity

Within this category, completing an activity was a type of immediate, intangible intrinsic reward. The satisfaction of completion, feeling good after a workout, feeling relieved that the work was done, enjoyment of tiredness and being conscious (aware) of benefits were examples of immediate, intangible, intrinsic rewards. From a technology facilitation perspective, many participants indicated that receiving points and stars, seeing their progression across the top of the app along the course of the 8-weeks (immediate, tangible, extrinsic rewards) provided reassurance (confident of ability as a reward) for the work done and validation of their efforts. The app fostered the possibility of replaying the video so that the exercises could be done correctly.

Participants in Group 2 also indicated that steps tracking with the pedometer provided a sense of relief in completing the task and was a good feeling to see numbers increase every day.

Participants in Group 3 were happy in getting the task done.

Having Freedom of Usage

The feeling of having the freedom to use the app anywhere and anytime was a benefit and a reward because the app afforded the possibility of feeling good when using the app. *"…you are able to do the exercises…in your pyjamas…while waiting for your eggs to boil…" (P31)*. This participant was able to use the app on her iPad. While the app in its current format was usable only on desktops, the technology provided a fitness-on-the-go approach, a concept that was appreciated by many.

Participants from Group 2 expressed the freedom to walk anytime anywhere with tracking metrics.

Having Intangible Rewards

Feeling important due to personalization of the app and, feeling good at receiving checkmarks for completing a routine were key properties of this category leading to rewards. Sincere praise from a PT or feedback from virtual trainers for task completion were also intangible rewards indicated to be fostered by technology.

In Group 2, the satisfaction of reaching pedometer steps and feeling energetic and vibrant after an exercise routine were key properties of this category leading to intangible rewards.

Group 3 participants felt getting complimented by others and doing the routines correctly contributed to intangible rewards.

Having Tangible Rewards

Many participants indicated rewarding themselves with cake, beer, wine, sweets and bread, on completion of the PA routines. In Group 1, from a technology facilitation perspective, gaining points doing the flexibility routines and aiming for smaller sized pants/trousers, weight loss, and unchanged belt buckle position were a few properties of this category. Motivation to participate in PA to earn these rewards was noted in a few participants (P10, P24, P31). The tangible rewards provided validation of their efforts (P10).

Group 2 participants were keen on seeing an increase in number of steps on the pedometer (P02, P13, P18, P21). This milestone led to them rewarding themselves with sweets and ice cream occasionally (P02, P21).

Group 3 participants sometimes rewarded themselves with snacks, coffee pizza and beer.

Feeling of Mental Satisfaction

This category represented the enjoyment and engagement experiences of participants from a PA perspective. When comparing the properties of the open codes and themes most participants from all the three groups expressed that the feeling of mental satisfaction stemmed from being less tired in doing the same routines on a weekly basis, seeing a visual improvement in their appearance, feeling energetic and meeting daily targets. Completion of PA tasks was a desirable characteristic among participants in all the three groups.

From a technology facilitation perspective, participants in Group 1 indicated that feedback elements, seeing progression in daily activities and seeing the point's number increase contributed to mental satisfaction leading to rewards.

Satisfaction from reaching daily targets on the pedometer provided participants from Group 2 a sense of reward: *"... often (as a reward) the feeling of energy and completion is good..." (P13); "... just hard work makes me feel good... and is a reward in itself..." (P18).*

Participants from Group 3 felt satisfied because of the feeling good aspect after completing the PA: *"... there were no specific rewards just a good feeling that I had a good week..." (P27).*

Having Self-Awareness

Another type of reward that emerged was the feeling of self-awareness, which was contributed to by understanding their deficiencies and overcoming laziness. This category emerged in Group 1 and Group 2. From a technology facilitation perspective, the app helped participants to be aware of lower upper-body strength, lower arm strength, and the need for a sense of discipline to be able to meet set daily challenges or targets posted by the app.

Experiencing a feeling of discipline and the desire to do more were a few properties that emerged from Group 2. This theme was not evident in Group 3.

Having Sense of Accomplishment

While participants from Group 1 maintained that getting things done, increasing challenges or exercise intensities, many participants did more than what was asked for in the app (P10, P11, P31).

Reaching daily targets (pedometer) was a key characteristic of having a sense of accomplishment leading to rewarding PA (P02, P13, P21).

Participating and competing PA like a half-marathon (P25) led to a sense of accomplishment leading to a feeling of reward.

Improving Confidence

Interestingly this category emerged from Group 1 and Group 2. Participants from Group 1 attributed the app to making them feel competent leading to a feeling of reward. *"... the app showed me that I could do wall push-ups, I could never do push-ups, and now I am bragging to my friends in my age group that I can do this" (P31). "... it is like an inspiration to do more activity or exercises" (P04).* Simplified age-centric exercise routines fostered this feeling of confidence in themselves. The points helped reassure that they completed the routines and were able to move on to the next exercise routine or the next level (day).

In Group 2, being competent was critical to participants to overcome the stereotypical notion that older persons were sedentary. *"... and the feeling of being competent ... being competent is my reward and at my age, I first need to be active and then comes the health..." (P23); "I started out ... you know at 500 steps per day... and now I am doing more than 10K per day... (P23)."*

This contributed to feeling confident and competent to do PA, leading to a feeling of reward in both groups.

Improving Health Condition

Rewards were also about achievement of better health due to PA. For participants in Group 1, improved breathing due to the app exercises, improvement in posture, lower blood sugar levels due to exercise and not being out of breath were some of the properties of this category. While these could be attributed to any exercise routine, the technology facilitation for participants from this group was that the app indicated their progression though the 8-week program, and when they put themselves to reality challenges, they were able to perform at a better level.

Participants in Group 2 and Group 3 were keen on lowering their blood-sugar level through exercises. Improving health conditions was an intangible, long term reward and inspired a feeling of attainable goal.

Inspiring Motivational Affordances

Participants in Group1 indicated that technology facilitation of PA through motivational affordances provided them with a feeling of immediate, tangible, and extrinsic rewards. These rewards fostered a sense of competence, a sense of accomplishment, a sense of being validated for their efforts, a sense of reassurance that they could do the routines and progress through the 8-week study.

Frustration also stemmed from the fact that the progress bar moved quite slowly (P26) in relation to the workouts remaining to be done in the 8-week program. Participants (P08, P31) also wanted the opportunity to do more exercises on the same day to gain more points and add to the visual progression or achieve higher challenge levels.

Participants from Group 2 were keen on step count and receiving badges for completing certain challenges like adding more difficulty levels like climbing hills and seeing progress through the week.

Participants in Group 3 were keen on completing their routine tasks.

Seeing Results of Efforts

Participants in all the three groups indicated that seeing the results of their efforts led to the feeling of being rewarded. In Group 1, in addition to their daily activities, the progression bar, the checkmark for having done the exercise routines, indication of completion of the reps and steps provided a result-oriented interface for the participants. Additionally, participants also started to include a heart monitor (P10, P11) to identify their heart rates pre-and post-test situation and also indicated that a calorie burn indicator synced with the app would be a good addition.

Participants in Group 2 noted pedometer reading each day and expressed that reaching daily targets was reward in itself.

In Group 3, time taken for the daily routine, and feeling of tiredness was a measure of the result of one's effort leading to a feeling of reward in doing PA.

Social Activity

Comparing one's progress with a running partner of a spouse were key rewards attributes for participants from all the three groups. While the app and the pedometer did not have features to enable social comparison of progression, participants voiced their interest in being able to do such a comparison.

In Group 2, running with a partner, comparing activities done (P02, P18) on specific routines and step count provides a sense of rewarding social activity.

Running for companionship and sharing experiences and life's challenges were key properties of rewarding social activity for Group 3 participants.

Tracking and PA

While there were many positive attributes to tracking and PA, there participants did indicate the negative aspects of tracking. Engagement in PA can be affected by these negative attributes; I discuss these positive and negative properties of the themes that emerged from the qualitative data. All of these themes relate distinctively to technology facilitation of PA.

Challenging Tracking Issues

A few participants in Group 1 were concerned that technology tracking of their movements and locations would be an invasion of privacy (P08, P11), and voiced concern that monitoring numbers would be a *botheration* (P08) and would make them *obsessive* (P11) and did not want to *punish themselves* (P08). Additionally, one participant indicated that the *inclination would be low* if the data had to be input each time (P29).

A few participants in Group 2 indicated that the tracking of inactive time would be good and would be a trigger to do more PA (P13, P23). Additionally, low numbers and not being fixated by numbers were additional tracking challenges from this group.

Group 3 participants did not give any feedback regarding challenges in tracking for PA.

Indicating Completion Status

A few participants from Group 1 were happy to see a completion status (P01), a checkmark (P08), or striking off from a list (P08, P11).

Participants from Group 2 wanted to see a comparison of steps done daily with the steps done in the past.

Participants from Group 3 wanted to see the amount of time taken to do a regular routine like walking (P15), time to destination (P25), and distance to go (P20).

Improving Body Form

Most participants from Group 1 indicated that improving body shape, form, and posture were very important for them. The app indicated reps and steps for each exercise routine which was used to do more PA. Many participants suggested the advantages of a automatic form checker and gait/posture improvement possibility in the app.

Participants in Group 2 indicated that in addition to step count they would welcome a feature that would help them improve their posture and gait when walking.

In Group 3 participants were more interested in reviewing their body improvements in the mirror and measuring weigh loss on a daily basis.

Indicating Motivational Affordances

Participants (Group 1) indicated technology facilitation of PA was fostered in this category by recognizing the value of recording progression, achievements check marked on a list of routines, time duration of routines and increasing challenges provided on a weekly basis. While a few participants did not notice the presence of points and stars, in the beginning, many participants did acknowledge that the points and stars served to validate their efforts and indicated their progression through the app program.

Participants from Group 2 clearly indicate the desire to maintain their daily walking targets and also trying to do better on each occasion.

Participants from Group 3 did not indicate any motivational affordance.

Making Social Connections

Participants from Group 1 indicated that comparing progress with a spouse or a partner gave them confidence to do more. However, the app did not have this feature enabled in the present format to acknowledge any social comparison. One participant (P24) indicated the possibility of overcoming loneliness through the app by the comment *"…it would be good to connect with other participants working on similar routines so that they would not feel that they are doing this alone at 10 pm on a Sunday evening…"*.

In Group 2, participants (P02, P18, P21) were keen on comparing with persons of the same age, which helped to reassure them that their efforts were in the right direction.

Group 3 participants (P07, P09, P14, P20) were more interested in the social aspects of walking and doing exercises together.

Needing Feedback

From a tracking perspective, participants in Group 1 indicated that the app provided feedback on progression (daily and weekly), number of reps and steps done per day and provided video information on the correct method of doing the exercise routines. Participants' suggestions were to use the app as a reminder to initiate the process of doing the exercise and provide notes of encouragement and progression on a daily basis. Presence of a virtual coach as indicated in the app helped reaffirm their desire to do the app routines, however, participants suggested that getting real-time feedback from the virtual coach would be beneficial to their improvement and confidence the correctness of their routines.

Participants in Group 2 and 3 needed feedback on weight loss and improvement in their daily activities.

Measuring PA

Participants from Group 1 and 2 indicated measuring of physical metrics such as weight loss on a daily basis, calories burnt, calorie intake, and heart rate on an intermittent basis. Quantifying PA for regular activities was defined by time duration of activity, distance walked/run, and steps done. Furthermore, in Group 1, app elements such as progression points earned, challenge levels on a weekly basis and, score, and stars contributed to the validation of effort done in a PA.

Participants in Group 3 indicated that measuring time duration for PA and weigh loss reassured them about their progress in daily tasks and fitness goals.

Technology Facilitation of PA

While terminologies such as gamification and motivational affordances may be new to older adults, they do understand aspects of setting up of attainable goals, on-the-spur of the moment challenges and pushing themselves to do more PA (Tabak et al., 2015; Kappen et al., 2016). The gamified PA system provided older adults with advantages such as monitoring PA, recognize their limitations with exercise intensities, increase challenges, feel validated for their efforts, and be rewarded for their task completion stages. Additionally, qualitative analysis indicated that the quantification of PA using tracking metrics and pedometers fostered the drive to walk more, explore the addition of new challenges and exercise intensities. These findings extend the results of prior studies on PA and older adults using gaming technologies (Far et al., 2015; Brox et al., 2017). Furthermore, older adults had the choice of monitoring their progression, tracking achievements, and improved sense of control of their efforts for PA with gamification elements. Additionally, these findings extend the results of prior studies on improving PA in older adults through mail-based interventions (Martinson et al., 2008; Peels et al., 2013) to the context of gamified PA interventions.

TA mapping of the evidential chain (Miles et al., 2014) indicated the justification of gamified PA for older adults is illustrated (**Figure 2**) (**Supplementary Table 11**). The first column represents the main research question, the second column represents the categories of deductive generalizable categories, the third column represents the motivational affordances that were findings from the analyses. The category on *fears and barriers for PA* was not included this illustration because the themes from this category were not gamification elements. The last column represents properties of motivational affordances as evident from participant responses.

To understand older adults' enjoyment and experiential aspects of using technology for PA, it was critical to examine the relevance of technology in the context of PA motivation, setting up goals, feeling of accomplishments, fears and barriers, and rewards, and tracking. TA in this paper extends prior work that listed axial codes and adaptive gamification guidelines (Kappen et al., 2018) by evolving a detailed list of gamification elements (motivational affordances) as illustrated in the evidential chain (**Figure 2**). These serve as valuable gamification elements that could help designers and researchers extend their PA technology for older adults within specific themes (axial codes). Furthermore, by investigating the influence of gamification elements in PA technology, this paper extends prior work of using web-based interventions to promote PA by sedentary older adults (55+) (Irvine et al., 2013). These findings also support improved behavioral changes and effective changes in PA of older adults (50+) (Peels et al., 2012) due to computer-tailored interventions and justifying the need for improved web-based interventions for older adults (50+) for better sustainability (van Stralen et al., 2010).

The evidential chain (**Figure 2**) indicated a mapping of themes related to the question categories and older adults PA from the gamification contexts. The gamified group showed important interest and enjoyment by the following: improving on their deficiencies, increasing challenges progressively, indicated perceived competence through increasing challenges progressively, feeling of the ability to do more and increasing difficulty levels, feeling importance of effort/importance by feeling validated for their efforts, measuring progress, and improvement in body conditioning. Perceived choice was afforded by the ability to select goals and challenges, self-regulation of routines, and flexibility of usage. Furthermore, value/usefulness was afforded by feeling energetic, wanting to do more, improved confidence, and improving ability. The themes emerging from the qualitative analysis also indicated that participants in the gamified group felt that a scheduled program with daily achievements and challenges with motivational affordances like points and stars (rewards) helped them feel that there was validation of their efforts, and provided constant monitoring of their progress. Furthermore, the emergent themes indicated the possibility of tailoring and customizing PA routines for older adults leading to adaptive engagement (Kappen, 2017; Kappen et al., 2018).

Design Guidelines

Key design guidelines for researchers and designers of gamified PA technology evidenced from the TA mapping (**Figure 2**) are summarized as follows:

Motivation—Autonomy

The gamified system should foster autonomy by incorporating short-term goals leading to long-term goals, indicate progression based on effort, a gradual increase in the intensity of the activity leading to harder and more complex tasks. This scaffolding will enable older adults to have the choice to commit and overcome their fear of inability.

Motivation—Competence

Incorporating new challenges into gamified technology in the form of higher intensities, a variety of exercise routines that could be performed anywhere-anytime, will help older adults feel reassured and find value in simple exercises and overcome their deficiencies. Additionally, the system should incorporate new, spontaneous, and surprise PA that will help overcome the routine and potential monotony. This will help them see value in doing PA that leads to improving strength, endurance and flexibility.

Motivation—Relatedness

The presence of virtual coaching and sharing of efforts of PA through the gamified PA system will help older adults overcome loneliness through virtual connections. Additionally, the system should incorporate PA routines that could be done with grandchildren, family members and other older adults to help support intergenerational activities and community building.

Setting Up Goals

The system should allow the flexibility of combining exercise types and intensities cognisant of a variety of health challenges related to aging. The ability of the system to progressively increase challenges, improve on deficiencies, and commitment to a schedule will foster self-regulation of PA activities.

Accomplishments

The system should be responsive to focus on the quality of the effort (doing it better) in doing the PA as opposed to quantity and time spent. The app should measure progress, track improvements while indicating the completion status of PA. This will foster mental satisfaction and confidence in the ability to overcome deficiencies and engender a sense of accomplishment.

Rewards

While points and badges are important to validate efforts and progression, the app should incorporate PA that enables intangible rewards such as freedom of usage, feeling good, energetic, better on the completion of a task. Attributes such as form checking, posture correction and encouragement through auditory feedback are attributes that would lead to tangible rewards such as weight loss and slim figure including waist size reduction.

Tracking

The app should incorporate simple tracking such as challenge levels, points earned, progression, sharing tips, and status with others and badges for effort and completion.

The above guidelines will be beneficial to researchers and designers of gamified PA technology when designing playful systems specific to the older adult demographic.

Implications of Practical Applications

While many of the motivational affordances indicated (**Figure 2**) can be specifically used for developing gamified PA technology for older adults, researchers, and designers must take into consideration age-related impairments, current health conditions and barriers to exercise (Dacey et al., 2008; Bamidis et al., 2014; Kappen et al., 2016). Older adult's challenges with acceptance of technology contrasted with analog methods must be taken into consideration when developing gamified PA technology. The gamified applications must provide for ease of onboarding, learnability, and foster increased agency.

CONCLUSION

While research on older adults in the research space of motivation and physical activity (PA) exists, to the best of our knowledge, limited research is available in the intersection of motivation, and activity facilitated by gamified technology over longer durations. Additionally, to the best of our knowledge, current gamification studies are limited to single-use-testing study design protocols where the effectiveness of such interventions cannot be gauged over a longer period of time. Our paper overcomes this gap and studied the usage of gamification elements over an 8-week period in the context of older adults PA using gamification. TA of the interview data showed distinct variations in emergent themes for the three groups over an 8-week period. This further indicated that gamification elements can be customized to participants for the 50+ age group and tailored to suit their current health conditions and prevalent barriers thus facilitating adaptive engagement in PA. The emergent themes, motivational affordances and design guidelines are valuable for researchers and designers of PA technology and will help to advance the development of gamified PA technology for older adults.

DATA AVAILABILITY STATEMENT

The data generated for this study are available on request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Research Ethics Board of the University of Ontario Institute of Technology (UOIT, now Ontario Tech University) and the Research Ethics Board of the Humber Institute of Technology and Advanced Learning. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

DK contributed to the study design, analysis, and writing of the contents of this article which was also a major component of the DK's Ph.D. dissertation: *Adaptive Engagement of Older Adults' Fitness through Gamification*. This dissertation is also referenced in this article. PM-B and LN supervised DK's Ph.D. and contributed to guiding with the study design, analysis, reviewing, and editing the final writing.

FUNDING

This work was made possible due to the following: Lennart Nacke's NSERC Discovery Grant 2018-06576, the Canada Foundation for Innovation Infrastructure Fund 35819 "'SURGE' - The Stratford User Research and Gameful Experiences Lab," the NSERC CREATE SWaGUR grant, Mitacs Accelerate, and the Social Sciences and Humanities Research Council Canada Grant 895-2011-1014 (IMMERSe); and Pejman Mirza-Babaei's NSERC Discovery Grant: GPIN-2014-05763. Additionally, Humber College of Technology and Advanced Learning, Faculty of Applied Sciences and Technology, Humber Employees Scholarship Fund and Applied Research and Innovation contributed to supporting the first author's Ph.D. research and therefore this experimental study.

ACKNOWLEDGMENTS

Authors are thankful to all participants who contributed their time and resources over an 8-week timeframe.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomp.2020.530309/full#supplementary-material

REFERENCES

Aarhus, R., Grönvall, E., Larsen, S. B., and Wollsen, S. (2011). Turning training into play: embodied gaming, seniors, physical training and motivation. *Gerontechnology* 10, 110–120. doi: 10.4017/gt.2011.10.2.005.00

Albaina, I. M., Visser, T., van der Mast, C. P. G., and Vastenburg, M. H. (2009). "Flowie: a persuasive virtual coach to motivate elderly individuals to walk," in *Proc. of the 3d International ICST Conference on Pervasive Computing Technologies for Healthcare*, 1–7. ICST. doi: 10.4108/ICST.PERVASIVEHEALTH2009.5949

Alhojailan, M. I., and Ibrahim, M. (2012). Thematic analysis: a critical review of its process and evaluation. *WEI Int. Eur. Acad. Conf. Proc.* 1, 8–21. Available online at: https://fac.ksu.edu.sa/sites/default/files/ta_thematic_analysis_dr_mohammed_alhojailan.pdf

Bamidis, P. D., Vivas, B., Styliadis, C., Frantzidis, C., Klados, M., Schlee, W., et al. (2014). A review of physical and cognitive interventions in aging. *Neurosci. Biobehav. Rev.* 44, 206–220. doi: 10.1016/j.neubiorev.2014.03.019

Bethancourt, H. J., Rosenberg, D. E., Beatty, T., and Arterburn, D. E. (2014). Barriers to and facilitators of physical activity program use among older adults. *Clin. Med. Res.* 12, 10–20. doi: 10.3121/cmr.2013.1171

Bleakley, C. M., Charles, D., Porter-Armstrong, A., McNeill, M. D. J., McDonough, S. M., and McCormack, B. (2015). Gaming for health: a systematic review of the physical and cognitive effects of interactive computer games in older adults. *J. Appl. Gerontol.* 34:NP166–89. doi: 10.1177/0733464812470747

Boyatzis, R. E. (1998). *Transforming Qualitative Information: Thematic Analysis and Code Development.* SAGE Publications Ltd.

Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitat. Res. Psychol.* 3, 77–101. doi: 10.1191/1478088706qp063oa

Brawley, R. L. W., Rejeski, J., and King, A. C. (2003). Promoting physical activity for older adults: the challenges for changing behavior. *Am. J. Prevent. Med.* 25, 172–183. doi: 10.1016/S0749-3797(03)00182-X

Brox, E., Th Konstantinidis, S., and Evertsen, G. (2017). User-centered design of serious games for older adults following 3 years of experience with exergames for seniors: a study design. *JMIR Serious Games* 5:e2. doi: 10.2196/games.6254

Cantor, A. B., and Lee, H. (1996). Sample-size calculations for cohen's kappa. *Psychol. Methods* l:150–153. doi: 10.1037/1082-989X.1.2.150

Charmaz, K. (2006). *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis.* Vol. 10. Thousand Oaks, CA: SAGE Publications.

Chase, J. D. (2013). Physical activity interventions among older adults: a literature review. *Res. Theory Nursing Pract.* 27, 53–80. doi: 10.1891/1541-6577.27.1.53

Consolvo, S., Everitt, K., Smith, I., and Landay, J. (2006). "Design requirements for technologies that encourage physical activity." in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI*' 06:457. doi: 10.1145/1124772.1124840

Corbin, J., and Strauss, A. (2015). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Fourth Edi. Thousand Oaks, CA: SAGE Publications, Inc.

Dacey, M., Baltzell, A., and Zaichkowsky, L. (2008). Older Adults' intrinsic and extrinsic motivation toward physical activity. *Am. J. Health Behav.* 32, 570–582. doi: 10.5993/AJHB.32.6.2

De Swert, K. (2012). *Calculating Inter-coder Reliability in Media Content Analysis Using Krippendorff's Alpha* Available online at: http://www.polcomm.org/wp-content/uploads/ICR01022012.pdf

Deci, E. L., Cascio, W. F., and Krusell, J. (1975). Cognitive evaluation theory and some comments on the calder and staw critique. *J. Personal. Soc. Psychol.* 31, 81–85. doi: 10.1037/h0076168

Deterding, S., Dixon, D., Khaled, R., and Nacke, L. E. (2011). From game design elements to gamefulness: defining 'gamification. *Proceedings of MindTrek 2011*, 9–15. Tampere: ACM. doi: 10.1145/2181037.2181040

Fan, C., Forlizzi, J., and Dey, A. (2012). "Considerations for technology that support physical activity by older adults." in *Proc. of ASSETS'12* (New York, NY: ACM Press), 33–40. doi: 10.1145/2384916.2384923

Far, I. K., Ferron, M., Ibarra, F., Báez, M., Tranquillini, S., Casati, F., et al. (2015). The interplay of physical and social wellbeing in older adults: investigating the relationship between physical training and social interactions with virtual social environments. *PeerJ Comput. Sci.* 1–25. doi: 10.7717/peerj-cs.30

Fereday, J., and Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis : a hybrid approach of inductive and deductive coding and theme development. *Int. J. Qual. Methods* 5, 1–11. doi: 10.1177/160940690600500107

Fife, D. L. (2008). *Reasons for Physical Activity and Exercise Participation in Senior Athletes*. Brigham Young University - Provo.

"FitBit" (2015). Available online at: www.fitbit.com

"Google Fit" (2015). Availabe online at: https://fit.google.com/

Gerling, K. M., Schild, J., and Masuch, M. (2010). "Exergame design for elderly users: the case study of SilverBalance," in *Proceedings of the 7th International Conference on Advances in Computer Entertainment Technology (ACE '10)* (New York, NY: Association for Computing Machinery), 66–69. doi: 10.1145/1971630.1971650

Gerling, K. M., Schulte, F. P., and Masuch, M. (2011). Designing and evaluating digital games for frail elderly persons," in *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology - ACE'11*, 1. New York, NY: ACM Press. doi: 10.1145/2071423.2071501

Gerling, K. M., Schulte, F. P., Smeddinck, J., and Masuch, M. (2012). "Game design for older adults: effects of age-related changes on structural elements of digital games." in *ICEC 2012, LNCS* 7522, 235–242. doi: 10.1007/978-3-642-33542-6_20

Geurts, L., Vanden Abeele, V., Husson, J., Windey, F., Van Overveldt, M., Annema, J., et al. (2011). "Digital games for physical therapy: fulfilling the need for calibration and adaptation," in *Proceedings of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction*, 117–24. doi: 10.1145/1935701.1935725

Guest, G., MacQueen, K. M., and Namey, E. E. (2012). "Validity and Reliability (Credibility and Dependability) in Qualitative Research and Data Analysis." in *Applied Thematic Analysis*, 11. doi: 10.4135/978148338 4436.n4

Hagstromer, M., Oja, P., and Sjostrom, M. (2006). The international physical activity questionnaire (IPAQ): a study of concurrent and construct validity. *Public Health Nutrition* 9, 755–762. doi: 10.1079/PHN2005898

Hamari, J., and Koivisto, J. (2015). Working out for likes': an empirical study on social influence in exercise gamification. *Comput. Human Behav.* 50, 333–347. doi: 10.1016/j.chb.2015.04.018

IPAQ Group (2005). *Guidelines for Data Processing and Analysis of the International Physical Activity Questionnaire (IPAQ) – Short and Long Forms*. Ipaq, no.

Irvine, A. B., Gelatt, V. A., Seeley, J. R., Macfarlane, P., and Gau, J. M. (2013). Web-based intervention to promote physical activity by sedentary older adults: randomized controlled trial. *J. Med. Int. Res.* 15:e19. doi: 10.2196/jmir.2158

Jancey, J. M., Clarke, P. H., Maycock, B., and Lee, H. (2009). Perceptions of physical activity by older adults: a qualitative study. *Health Educ. J.* 68, 196–206. doi: 10.1177/0017896909339531

Kappen, D., Mirza-Babae, P., and Lennar, P. N. (2018). "Gamification of older adults' physical activity : an eight-week study," in *Proceedings of the Annual Hawaii International Conference on System Sciences*-51, 1–12. doi: 10.24251/HICSS.2018.149

Kappen, D. L. (2017). *Adaptive Engagement of Older Adults' Fitness through Gamification.* University of Ontario Institute of Technology. Available online at: http://hdl.handle.net/10155/881.

Kappen, D. L., Mirza-Babaei, P., and Nacke, L. E. (2017). Gamification through the application of motivational affordances for physical activity technology. *In Proc. of CHIPLAY* 17, 5–18. doi: 10.1145/3116595.3116604

Kappen, D. L., Mirza-Babaei, P., and Nacke, P. L. E. (2019). "Motivational affordances for older adults' physical activity technology : an expert evaluation," in *Human Aspects of IT for the Aged Population. Design for the Elderly and Technology Acceptance. HCII 2019. Lecture Notes in Computer Science, Vol 11592*, edited by Zhou J. and Salvendy G (Springer International Publishing AG), 1–19. doi: 10.1007/978-3-030-22012-9_28

Kappen, D. L., and Nacke, L. E. (2013). "The kaleidoscope of effective gamification: deconstructing gamification in business applications," in *Proceedings of the First International Conference on Gameful Design, Research, and Applications - Gamification* 13, 119–122. doi: 10.1145/2583008.2583029

Kappen, D. L., Nacke, L. E., Gerling, K. M., and Tsotsos, L. E. (2016). Design strategies for gamified physical activity applications for older adults." in *Proceedings of the Annual Hawaii International Conference on System Sciences-49, IEEE*, 1309–18. *IEEE Computer Society*. doi: 10.1109/HICSS. 2016.166

Kaushal, N., and Rhodes, R. E. (2015). Exercise habit formation in new gym members: a longitudinal study. *J. Behav. Med.* 38, 652–663. doi: 10.1007/s10865-015-9640-7

Keenan, T. (2009). Internet use among midlife and older adults: an aarp bulletin poll. *AARP Res.* 1–14.

King, A. C. (2001). Interventions to promote physical activity by older adults. *J. Gerentol.* 56A, 36–46. doi: 10.1093/gerona/56.suppl_2.36

King, A. C., Jack Rejeski, W., and Buchner, D. M. (1998). Physical activity interventions targeting older adults: a critical review and recommendations. *Am. J. Prevent. Med.* 15, 316–333. doi: 10.1016/S0749-3797(98)00085-3

King, A. C., and King, D. K. (2010). Physical activity for an aging population. *Public Health Rev.* 32, 1–19. doi: 10.1007/BF03391609

Kuroda, Y., Sato, Y., Ishizaka, Y., Minoru, Y., and Yamaguchi, N. (2012). Exercise motivation, self-efficacy, and enjoyment as indicators of adult exercise behavior among the transtheoretical model stages. *Global Health Promotion* 19, 14–22. doi: 10.1177/1757975911423073

Lister, C., West, J. H., Cannon, B., Sax, T., and Brodegard, D. (2014). Just a Fad? Gamification in health and fitness apps. *JMIR Serious Games* 2, 1–12. doi: 10.2196/games.3413

Mahmud, A., Mubin, O., Shahid, S., and Martens, J. B. (2010). Designing social games for children and older adults: two related case studies. *Entertainment Computing* 1, 147–56. doi: 10.1016/j.entcom.2010.09.001

Marston, H. R. (2013). Digital gaming perspectives of older adults: content vs. interaction. *Educ. Gerontol.* 39, 194–208. doi: 10.1080/03601277.2012.700817

Martinson, B. C., Crain, L., Sherwood, N. E., Hayes, M., Pronk, N. P., and O'Connor, P. J. (2008). Maintaining physical activity among older adults: six-month outcomes of the keep active minnesota randomized controlled trial. *Prevent. Med.* 46, 111–119. doi: 10.1016/j.ypmed.2007.08.007

Martinson, B. C., Sherwood, N. E., Crain, L., Hayes, M. G., King, A. C., Pronk, N. P., et al. (2010). Maintaining physical activity among older adults: 24-month outcomes of the keep active minnesota randomized controlled trial. *Prevent. Med.* 51, 37–44. doi: 10.1016/j.ypmed.2010.04.002

Miles, M. B., Huberman, M., and Saldana, J. (2014). "Drawing and verifying conclusions," in *Qualitative Data Analysis: A Methods Sourcebook*, 275–322.

Mouter, N., Vonk Noordegraaf, D., and Vonk Noordegraaf, D. M. (2012). *Intercoder Reliability for Qualitative Research*. October.

Mouton, A., and Cloes, M. (2013). Web-based interventions to promote physical activity by older adults: promising perspectives for a public health challenge. *Arch. Public Health* 71:16. doi: 10.1186/0778-7367-71-16

Mullen, S. P., Olson, E. A., Phillips, S. M., Szabo, A. N., Wojcicki, T. R., Mailey, E. L., et al. (2011). Measuring enjoyment of physical activity in older adults: invariance of the physical activity enjoyment scale (Paces) across groups and time. *Int. J. Behav. Nutr. Phys. Activity* 8:103. doi: 10.1186/1479-5868-8-103

Peels, D. A., Bolman, C., Johanna Golsteijn, R. H., Vries, H., Mudde, A. N., Marieke van Stralen, M., et al. (2013). Long-term efficacy of a printed or a web-based tailored physical activity intervention among older adults. *Int. J. Behav. Nutr. Phys. Activity* 10:104. doi: 10.1186/1479-5868-10-104

Peels, D. A., Van Stralen, M. M., Bolman, C., Golsteijn, R. H. J., De Vries, H., Mudde, A. N., et al. (2012). Development of web-based computer-tailored advice to promote physical activity among people older than 50 years. *J. Med. Internet Res.* 14, 15–27. doi: 10.2196/jmir.1742

Romero, N., Sturm, J., Bekker, T., de Valk, L., and Kruitwagen, S. (2010). Playful persuasion to support older adults' social and physical activities. *Interact. Comput.* 22, 485–495. doi: 10.1016/j.intcom.2010.08.006

Ryan, G. W., and Bernard, H. R. (2000). "Data management and analysis methods," in *Handbook of Qualitative Research.,* eds D. Norman and Y. Lincoln (Sage Publications), 769–802.

Schmidt, L., Rempel, G., Murray, Tara-leigh Mchugh, T. C., and Vallance, J. K. (2016). *Exploring Beliefs Around Physical Activity Among Older Adults in Rural Canada* 1:1–8. doi: 10.3402/qhw.v11.32914

Schutzer, K. A., and Graves, B. S. (2004). Barriers and motivations to exercise in older adults. *Prevent. Med.* 39, 1056–1061. doi: 10.1016/j.ypmed.2004.04.003

Stathi, A., Fox, K. R., Withall, J., Bentley, G., and Thompson, J. L. (2014). "Promoting physical activity in older adults: a guide for local decision makers," in *Avon Network for the Promotion of Active Ageing in the Community*, 1–72.

Tabak, M., Dekker-van Weering, M., van Dijk, H., and Vollenbroek-Hutten, M. (2015). Promoting daily physical activity by means of mobile gaming: a review of the state of the art. *Games Health J.* 4, 460–469. doi: 10.1089/g4h.2015.0010

Vaismoradi, M., Jones, J., Turunen, H., and Snelgrove, S. (2016). Theme development in qualitative content analysis and thematic analysis. *J. Nursing Educ. Pract.* 6, 100–110. doi: 10.5430/jnep.v6n5p100

van der Bij, A., Laurant, M. G. H., and Wensing, M. (2002). Effectiveness of physical activity a review. *Am. J. Prevent. Med.* 22, 120–133. doi: 10.1016/S0749-3797(01)00413-5

van Stralen, M., Hein de Vries, M., Bolman, C., Mudde, A. N., and Lechner, L. (2010). Exploring the efficacy and moderators of two computer-tailored physical activity interventions for older adults: a randomized controlled trial. *Annals Behav. Med.* 39, 139–150. doi: 10.1007/s12160-010-9166-8

van Stralen, M. M., de Vries, H., Mudde, A. N., Bolman, C., and Lechner, L. (2011). The long-term efficacy of two computer-tailored physical activity interventions for older adults: main effects and mediators. *Health Psychol.* 30, 442–452. doi: 10.1037/a0023579

Weber, A. S., and Sharma, M. (2011). Enhancing effectiveness of physical activity internventions among older adults. *Am. J. Health Stud.* 26, 25–36. doi: 10.1016/s0749-3797(01)00413-5

Zelinski, E. M., and Reyes, R. (2009). Cognitive benefits of computer games for older adults. *Gerontechnology* 8, 220–235. doi: 10.4017/gt.2009.08.04.004.00

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Computer Science

ORIGINAL RESEARCH
published: 23 October 2020
doi: 10.3389/fcomp.2020.00033

Check for updates

"HIIT" the ExerCube: Comparing the Effectiveness of Functional High-Intensity Interval Training in Conventional vs. Exergame-Based Training

Anna Lisa Martin-Niedecken[1], Andrea Mahrer[2], Katja Rogers[3,4], Eling D. de Bruin[2,5] and Alexandra Schättin[2]*

[1] Subject Area in Game Design, Department of Design, Zurich University of the Arts, Zurich, Switzerland, [2] Motor Control and Learning, Department of Health Sciences and Technology, Institute of Human Movement Sciences and Sport, ETH Zurich, Zurich, Switzerland, [3] Institute of Media Informatics, University of Ulm, Ulm, Germany, [4] HCI Games Group, University of Waterloo, Waterloo, ON, Canada, [5] Division of Physiotherapy, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden

OPEN ACCESS

Edited by:
Kathrin Gerling,
University of Lincoln, United Kingdom

Reviewed by:
Jussi Holopainen,
University of Lincoln, United Kingdom
Rainer Malaka,
University of Bremen, Germany

***Correspondence:**
Anna Lisa Martin-Niedecken
anna.martin@zhdk.ch

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Computer Science

Received: *31 January 2020*
Accepted: *10 July 2020*
Published: *23 October 2020*

Citation:
Martin-Niedecken AL, Mahrer A, Rogers K, de Bruin ED and Schättin A (2020) "HIIT" the ExerCube: Comparing the Effectiveness of Functional High-Intensity Interval Training in Conventional vs. Exergame-Based Training. Front. Comput. Sci. 2:33. doi: 10.3389/fcomp.2020.00033

Regular physical activity is crucial for a physically and mentally healthy lifestyle. Training methods such as high-intensity interval training (HIIT) have become increasingly popular as they enable substantial training effects in little time. HIIT typically involves recurring short phases of close-to-maximal exercise intensity, interspersed with low-intensity recovery phases. Originally mainly practiced via uniformly repetitive movements, newer variations include varied functional and holistic exercises (fHIIT). While HIIT facilitates many health advantages, fHIIT is considered more beneficial since it activates more muscles, requires more coordination, strength and balance, and mimics more natural movements which transfer well to daily life. However, fHIIT is a very intense training approach; it requires strong focus and intrinsic motivation to frequently push beyond perceived physical and mental limits. This is a common barrier to exploiting the full potential of this efficient training method. Exergames may facilitate this kind of training due to their playful, immersive, motivating nature. Yet so far, few studies have investigated HIIT-exergames – no fHIIT-exergames. This is possibly because few exergames featured both (1) an effective training concept that is comparable to HIIT, and (2) an attractive and motivating game design. We believe that this lack of holistic integration of both aspects is partly why there is currently little evidence for long-term motivation and training effects in exergame-based training. Our work addresses this gap through the design of an adaptive fHIIT protocol for the ExerCube fitness game system, creating a HIIT-level functional exergame. We conducted a within-subjects study to compare objective and subjective training intensity induced by the ExerCube against a conventional fHIIT session with healthy young adults. Furthermore, we evaluated participants' subjective experience with regards to motivation, flow, and enjoyment during both conditions. Our results contribute empirical evidence that exergames can induce HIIT-level intensity. While perceived physical exertion was slightly lower in the ExerCube condition, it yielded significantly better results for flow, enjoyment, and motivation. Moreover, the ExerCube seemed to enable a dual-domain training (higher cognitive load). We discuss these results in the context of exergame design for fHIIT, and provide practical suggestions covering topics such as safety precautions and physical-cognitive load balancing.

Keywords: exergame, high-intensity interval training, functional training, effective, attractive, heart rate, ExerCube

INTRODUCTION

Regular physical activity is crucial for a physically and mentally healthy lifestyle at all ages, as it protects against cardiovascular diseases and diabetes (Folsom et al., 2000; Steinbeck, 2001) and mental disorders such as depression (Biddle and Asare, 2011). However, especially in young adults, a lack of motivation is a common barrier to participating in regular physical activity (Trost et al., 2002; Teixeira et al., 2012). Therefore, training methods such as high-intensity interval training (HIIT) have become increasingly popular due to their good dose-effect relationship, i.e., substantial training effects in little time (MacInnis and Gibala, 2017). HIIT typically involves recurring short phases of close-to-maximal exercise intensity [beyond 80% of the maximum heart rate (HR)], interspersed with low-intensity recovery phases (Gibala et al., 2012). Originally, conventional HIIT (cHIIT) was mainly practiced via uniformly repetitive movements such as cycling on an ergometer, rowing on a rowing machine or running on a treadmill. Only 20 min of cHIIT three times a week can result in significant health benefits (Kilpatrick et al., 2014; Weston et al., 2014). Newer HIIT variations follow the same structure, but include functional multi-joint exercises such as squats, lunges, and burpees (Feito et al., 2018). Although cHIIT positively affects aerobic fitness, body composition, insulin sensitivity, blood lipid profile, blood pressure as well as cardiovascular functions (Burgomaster et al., 2008; Babraj et al., 2009; Kemi and Wisløff, 2010), functional HIIT (fHIIT) is considered more beneficial (McRae et al., 2012; Buckley et al., 2015; Kliszczewicz et al., 2019; Menz et al., 2019). It activates more muscles activity (Folsom et al., 2000), requires more coordination (Wilke et al., 2019), positively affects motor functions such as strength and balance (Weiss et al., 2010; Wilke et al., 2019), and mimics more natural daily movements which transfer well to daily life (Weiss et al., 2010). Overall, HIIT has been shown to have a more beneficial impact on fitness and cardiovascular health than other exercise methods (Weston et al., 2014).

However, HIIT—in all its variations—is a very intense training approach; it requires strong focus and intrinsic motivation (Teixeira et al., 2012), as participants have to frequently reach for or push beyond their perceived physical and mental limits. It has been shown that while the high intensity component of HIIT is useful for health benefits, the motivation to continue exercising decreases as the intensity of the exercise increases (Peng et al., 2011; Foster et al., 2015). Therefore, many people lose motivation when doing HIIT (Haller et al., 2019). These concerns are supported by a recent study which found that following a HIIT intervention with overweight and obese adults, only 40% adhered to the program 12 months later (Roy et al., 2018).

Exergames, if designed properly with regards to effectiveness and attractiveness (Sinclair et al., 2009), appear to be a suitable and appealing tool to facilitate this kind of training due to their playful, immersive, and motivating nature (Oh and Yang, 2010; Farrow et al., 2019). So far, only few studies have investigated exergames in the context of cHIIT (de Bruin et al., 2019; Farrow et al., 2019; Haller et al., 2019; Keesing et al., 2019) and none in fHIIT. This is possibly because few exergames feature both (1) an effective training concept that is comparable to cHIIT, and (2) an attractive game design to sustain players' motivation (Martin-Niedecken et al., 2019). We believe that this lack of holistic integration of both aspects in exergames is partly why currently little evidence exists for long-term motivation and training effects in exergame-based training (Best, 2015). Prior to long-term investigation, the research and development community needs to design suitable exergames and investigate their feasibility. Thus, there is a need to design and evaluate exergames that combine the best of gaming and fitness; i.e., developing training tools that are both motivating and effective, while following more holistic HIIT variations.

One exergame specifically designed in terms of this holistic approach is the ExerCube: a commercial immersive fitness game setting by Sphery Ltd. (Martin-Niedecken and Mekler, 2018; Martin-Niedecken et al., 2019). The company is open to making the ExerCube available to researchers as a research platform. Thus, the first early stage functional fitness game prototype designed for this system was found to be on par with personal training regarding immersion, motivation, and flow as shown in a previous empirical study (Martin-Niedecken and Mekler, 2018; Martin-Niedecken et al., 2019). This previous study with the ExerCube (Martin-Niedecken and Mekler, 2018; Martin-Niedecken et al., 2019) largely employed self-reported and subjective measures, leading to a research gap with regards to its objective training intensity. An objectively high training intensity is necessary to achieve benefits of HIIT. This leads us to explore the following main research question: How does objective and subjective physiological training intensity in the ExerCube compare to that induced by conventional fHIIT (cfHIIT)?

Our work explores this research gap, with the goal of better understanding the design requirements of and potential for holistic HIIT in attractive and effective exergames. Therefore, we provide both design and research contributions: First, we

Abbreviations: HIIT, high-intensity interval training; fHIIT, functional high-intensity interval training; cHIIT, conventional high-intensity interval training; cfHIIT, conventional functional high-intensity interval training; HIFT, high-intensity functional training; HR, heart rate; HR_{max}, maximum heart rate; VR, Virtual Reality; CHR_{max}, calculated maximum heart rate; HR_{avg}, average heart rate.

designed a fHIIT-protocol with physiological and cognitive measures for the ExerCube system to create a HIIT-level functional exergame as well as a comparable cfHIIT protocol. We conducted a within-subjects study to compare the subjective and objective training intensity induced by a single ExerCube session and a single cfHIIT session (best practice in the fitness market) with young healthy adults. Furthermore, we evaluated participant's subjective experience including motivation, flow experience, and enjoyment during both types of training.

Our results show that the employed exergame is a feasible tool for inducing HIIT-level intensity. While perceived physical exertion was lower than in the cfHIIT condition, the interquartile range of the ExerCube condition's average HR reached the HIIT threshold (moderate to high-intensity). The ExerCube condition also yielded significantly better results for flow, enjoyment, and motivation. It also seemed to trigger higher cognitive load, i.e., it achieved a dual-domain training. We present a comparison with high external validity and applicability within the fitness industry; our results thus contribute empirical evidence that an exergame can be used to induce HIIT-level intensity in addition to positive effects on motivation. Based on the results, we discuss how effective and motivating exergames should be designed to implement fHIIT, and inform future explorations of their effects in terms of associated health benefits and long-term motivation.

RELATED WORK

HIIT is an extremely time-efficient and beneficial training method, originally often performed with an ergometer, rowing machine or by running (Feito et al., 2018). Back in 1996, Tabata et al. (1996) were the first to demonstrate that a 4-min high-intensity workout (consisting of eight 20-s bouts of all-out performance with 10-s breaks in-between) is more effective than exercising for 1 h at moderate intensity. While both methods increased oxygen consumption (VO_2max), only the high-intensity training enhanced anaerobic capacity. Other studies have confirmed this finding, as also covered by a more recent systematic review (Milanović et al., 2015). Today, there are many different ways to perform HIIT. What all programs have in common is that they are characterized by periods of very heavy effort combined with periods of either complete rest or low-intensity recovery. HIIT variations such as spinning classes have been extremely popular for years; by allowing for social interaction and group dynamics, they increase or maintain motivation and help people to stay with this intensive training approach long-term (Caria et al., 2007). This is also reflected in the "Worldwide survey of fitness trends 2019" (Thompson, 2018), where HIIT stook third place. Parallel to this trend, the survey reports functional fitness training in ninth place, and first place for wearable technologies such as HR sensors. This tendency clearly indicates a combination of certain training approaches (frequent endurance training with additional regular strength training and neuro-motor exercise) that most attract today's young adults, and are recommended in this combination by international guidelines on physical activity (Thompson et al., 2010; World Health Organization, 2010). However, cHIIT does not necessarily incorporate major stimuli improving strength, coordination, and motor control (Wilke et al., 2019).

From HIIT to HIFT to fHIIT

A newer HIIT-related variation is high intensity functional training (HIFT) which is a combination of functional multi-joint movements. These movements are adjustable to any fitness level and elicit greater muscle recruitment than more traditional exercises (Feito et al., 2018). These functional training elements, i.e., exercises that mimic movements of daily living (e.g., squats and lunges), have been shown to simultaneously improve strength and balance (Weiss et al., 2010). While HIIT exercise is characterized by relatively short bursts of repeated vigorous activity, interspersed by periods of rest or low-intensity exercise for recovery, HIFT utilizes constantly varied functional exercises and various activity durations that may or may not incorporate rest (Feito et al., 2018). The commonly practiced combination of both approaches is fHIIT.

A recent study compared effects of moderate aerobic exercise and circuit-based fHIIT on motor performance and exercise motivation in untrained adults (Wilke et al., 2019). The circuit-based fHIIT enhanced physical functions (strength and endurance) and motivation to exercise more effectively than the moderate condition. Another study examined the physiological effects of an fHIIT program on endurance and strength of physically active adults over a 4-week period and found rapid physiological improvements in strength as well as in aerobic and anaerobic capacity (Kliszczewicz et al., 2019). fHIIT seems to be a beneficial variation of HIIT as its protocols allow for multiple performance and physiological adaptations that are not observed by training using unimodal HIIT methodology (Feito et al., 2018). fHIIT combines the best of HIIT and HIFT, benefits the whole body (endurance, strength, coordination, flexibility, etc.) and transfers more to daily life activities (McRae et al., 2012; Buckley et al., 2015; Feito et al., 2018; Menz et al., 2019).

Today's fitness market is reacting to this and provides special fHIIT classes with different foci (e.g., BodyAttack®) which—similar to spinning classes—enable an intense and socially motivating group workout on a holistic level. Mobile fitness apps such as Freeletics[1] further provide options for digital fHIIT-like training for users "on the go" and allow them to share, compete, and cooperate with one another. Although fitness providers frame HIIT and fHIIT as motivating as possible, it remains an extremely challenging training approach.

Exergames: A Promising Training Tool

In today's digital age, exergames (Oh and Yang, 2010)—games that are controlled by physical exercises and provide an additional cognitive challenge for the player—are being explored as a suitable tool to introduce more people to effective training approaches and motivate them to keep on track.

Studies on exergame training in different target populations such as older adults, children, adolescents or patients indicate effects on cognitive (e.g., executive function, attention, and visual-spatial skills) (Li et al., 2016; Joronen et al., 2017; Lee

[1] freeletics.com

et al., 2017; Byrne and Kim, 2019; Kappen et al., 2019; Stojan and Voelcker-Rehage, 2019), physical (e.g., energy expenditure, HR, and physical activity) (Lu et al., 2013; Lee et al., 2017; Byrne and Kim, 2019; Tondello et al., 2019), and mental (e.g., social interaction, self-esteem, motivation, and mood) (Macvean and Robertson, 2013; Lyons, 2015; Lee et al., 2017; Martin-Niedecken and Götz, 2017; Valenzuela et al., 2018) aspects. Generally, exergames are well-known for their playful combination of physically and cognitively challenging tasks and thus provide dual-domain training which promises greater effects compared to traditional single-task training approaches (Huang et al., 2014; Hardy et al., 2015; Benzing and Schmidt, 2018; Kappen et al., 2019).

Besides specific effects of exergame training, it is further known for its appealing and motivating impact, especially in physically inactive populations (Wüest et al., 2014; Hoffmann et al., 2016). By providing different players [of different motivational types (Isbister, 2016)] with audio-visually, narratively appealing, and immersive game scenarios, exergames shift players' (cognitive) focus onto the playful experience. This makes it easier to engage with a physically challenging workout (Xiong et al., 2019). Therefore, exergames have successfully been shown to increase training adherence (Kajastila and Hämäläinen, 2015), long-term motivation (Márquez Segura et al., 2013), engagement (Mueller and Isbister, 2014), immersion (Wüest et al., 2014), and flow (Sinclair et al., 2007) in players from different populations.

Exergame-Based HIIT

In the context of cHIIT, only few studies exists that investigated feasibility of exergames specifically designed for cHIIT with regards to physiological training outcomes or qualitative factors such as motivation and enjoyment. To the best of our knowledge, no exergames have been specifically designed and evaluated for fHIIT as of yet.

de Bruin et al. (2019) investigated the feasibility and effects on cardiovascular fitness of an exergame-based HIIT program in untrained elderly people. The 4-week training included a cognitively-simple game which required fast steps for the intense training phases, and games that were cognitively more but physically less challenging for the low-intensity phases. In the low-intensity phase, participants ranged within 50–70% of their maximum heart rate (HR_{max}). Both used a step-based platform as game controller. They found that the exergame-based HIIT was a feasible and well-accepted approach and led to the intended physical intensity (70–90% of HR_{max}). Furthermore, their collected qualitative feedback identified certain aspects which could increase study outcomes in future iterations (e.g., game music, more audiovisual feedback, and increased challenge).

Farrow et al. (2019) compared different in-game conditions (allowing participants to race against their own performance or by increasing the resistance) in a head-mounted virtual reality (VR) exergame-based HIIT on an ergometer against traditional ergometer-based HIIT in physically inactive young adults. They found that VR exergaming increased enjoyment during a single bout of HIIT and led to an average of 74–89% of HR_{max} over all tested conditions in untrained individuals. Furthermore, the presence of a virtual ghost to compete with appeared to be an effective method to increase exercise intensity of VR-based HIIT.

Barathi et al. (2018) proposed and evaluated an interactive feedforward approach (a method based on competition with oneself, i.e., against an improved self-model of the player) to rapidly improve performance in a HIIT cycling VR exergame. They found that the interactive feedforward method led to improved performance (participants' average HR was above 80% of HR_{max}) while maintaining intrinsic motivation and was superior to competing against a virtual competitor.

A different VR-HIIT exergame was developed for a rowing machine by Keesing et al. (2019). They utilized gameplay mechanics and the synchronization of rowing rhythm with music rhythm to automatically induce HIIT without the need for a physical instructor. They reported that gameplay and music were both effective at inducing HIIT, but music had a stronger effect on both performance and enjoyment.

Haller et al. (2019) investigated the effects of virtual spectators (and their rhythmic clapping based on participants' ergometer speed) on motivation during a HIIT-exergame. They found that virtual crowd feedback increased cycling speed and participants' HR [to around 171 beats per minute (bpm); percentages of HR_{max} were not reported].

Finally, Moholdt et al. (2017) compared HIIT with an online multiplayer ergometer-based exergame to walking in male students. Their exergame elicited an average intensity of 73–83% of HR_{max} and a higher enjoyment than walking.

In summary, the evaluated exergames did not feature full-body functional exercises, nor necessarily a comprehensive, meaningful, audio-visually appealing, and adaptive game design. By this, we mean that—besides different training approaches—these exergames did not follow a holistic design approach covering all design levels of an exergame (Martin-Niedecken and Mekler, 2018; Martin-Niedecken et al., 2019) as well as taking into account potential interdependencies and interaction effects between these, which can affect the targeted game experience. An attractive and effective exergame design encloses the player's moving and sensing body (Bianchi-Berthouze et al., 2007; Mueller et al., 2011) and allows for effective and playful exercises (Marshall et al., 2016). These exercises in turn are mediated by the game controller technology which should be easily and naturally embedded into the moving player's body scheme (Pasch et al., 2009; Kim et al., 2014; Shafer et al., 2014). Moreover, the virtual game scenario represents the player's bodily input in the virtual environment and provides audio-visual as well as haptic or tactile feedback for the player and their reacting body (Shaw et al., 2017). Furthermore, the aforementioned exergames did not necessarily feature individually adjustable cognitive and physical challenges (Sinclair et al., 2009). Thus, to the best of our knowledge, while there are HIIT exergames, there are no exergames specifically designed for fHIIT, nor studies investigating them.

MATERIALS AND METHODS

Based on this gap, we aimed to explore whether an fHIIT exergame can induce the same exercise intensity as a cfHIIT in our primary research question. For this comparison, we designed an fHIIT exergame that leverages the full potential of exergames: (1) targeting whole body exercise (holistic) while also providing challenges for coordination and cognition, (2) with attractive audio-visual design to increase motivation, and (3) automatic challenge adaptation (physical and cognitive) by the system's algorithm.

The baseline for this kind of exergame was cfHIIT as it is currently practiced on the fitness market, i.e., also at its full potential: (1) targeting whole body exercise, (2) with music to ensure roughly equivalent auditory appeal, and (3) with physical-challenge adaptation by the instructor, plus to a degree self-chosen adaptation. Further, cfHIIT is often offered in small groups, i.e., leveraging social factors for motivation. Given a secondary research question on whether an fHIIT exergame can compare to cfHIIT in eliciting motivation, we considered it a not field-compatible comparison if the cfHIIT would have been practiced in individual sessions.

Stimuli: The ExerCube

The ExerCube (Martin-Niedecken and Mekler, 2018; Martin-Niedecken et al., 2019) is an immersive mixed-reality fitness game. Players are surrounded by three walls, which serve as projection screens and a haptic interface for energetic bodily interactions. A customized motion tracking system tracks players' movements via HTC Vive trackers (attached to their wrists). To ensure an ideal workout experience [in terms of attractive design and effective exercises (Sinclair et al., 2009)] for a wide spectrum of players with different skill sets, the ExerCube continuously adapts game difficulty to players' individual fitness and cognitive skills. Training intensity is measured via continuous HR tracking (i.e., players wear a HR-sensor chest strap) and set to an individual pre-defined HR training range. Cognitive skills are measured via in-game performance (reacting to visual stimuli at the right time).

The Sphery Racer (see **Figure 1**) is a single-player game experience designed for the ExerCube setting. It was developed in several iterations based on the prototype presented and evaluated by Martin-Niedecken and Mekler (2018) and Martin-Niedecken et al. (2019), and it is now being employed as a research object by several research groups. In collaboration with the ExerCube's development team, we modified the game design (game mechanics and audio-visual design), the level structure and the HR-based game adaption algorithm to be comparable with a cfHIIT.

Like its prototypical predecessor, Sphery Racer asks players to progress along a fast-paced race track via an avatar on a hoverboard. The motion tracking system transfers player movements (based on a functional workout) onto this avatar and thus on the virtual racing track. Along the race, players are challenged by obstacles that require physical exercises (e.g., squats, lunges, and burpees) and by an additional cognitive challenge including quick information processing, which exercise has to be performed when (i.e., reaction, time, and coordination challenges).

The game starts with an on-boarding tutorial scene during which the game is calibrated to the exact height of the player. After successful calibration, the player's avatar drives onto the racing track and to the first instructional pitstop sequence (see **Figure 2**). The game contains five training pitstops (∼0.5–2 min. each, see **Table 1**), which serve as tutorials to become familiar with the respective steering movements. All exercises are instructed audio-visually (i.e., the avatar shows the exercise, written instructions are added, and a voice provides additional hints). The exercises start with low-to-moderate intensity (in terms of both physical and cognitive demand) and over time gradually increase until reaching high-intensity exercises (e.g., skippings and burpees).

After each pitstop, players return to the racing track, where they perform all thus-far learned movements in five racing sections. To integrate a warm-up phase followed by gradually intensifying level design, the racing section durations range from 2.5 min (first and second sections), to 5 min (third and fourth), and finally 10 min (last section). A complete workout session in the ExerCube takes 26–28 min.

The physical and cognitive game difficulty adjustments are gradually adapted independently over all training levels on a 10-point difficulty scale, where one level is defined as one step on the 10-point scale (e.g., from 5 to 6). Like previously, the algorithm determines players' individual calculated HR_{max} (CHR_{max}) based on the following formula (Nes et al., 2013):

$$CHRmax = 211 - 0.64 \times age \tag{1}$$

A comprehensive fitness study proved that this formula adequately explained HRmax by considering an age range of 19–89 years (Nes et al., 2013). Previously the ExerCube's algorithm also aimed toward reaching a high intensity training level (80–90% of HR_{max}), by a less finely tuned algorithm (HR <150 bpm for 0.5 min: increase speed slightly by one level; HR >175 bpm for 1 min: decrease speed slightly by one level; HR >190 bpm: decrease speed strongly by two level). However, the algorithm was not found to reach this training intensity. For the purpose of the presented study, we refined the algorithm: During the racing sections, the game aimed to get players to a specific HR range (70–90% of CHR_{max}) and then kept them at this level. Outside of this range, a lower HR lead to an increase in physical challenge, i.e., speed, exercise frequency (one level per check), while a higher HR lead to a decrease (once 100% of CHR_{max} was reached, this decrease was sped up by three levels to ensure players' safety). For the first two racing sections, the system employed a strategy for increasing players' HR, i.e., when 70% of CHR_{max} has not been reached, it checked actual HR_{max} every 30 s (whereas every 60 s otherwise). For the subsequent racing sections (3–5), the system checked more often (every 20 s in the increasing phase, and 10 s when above 90% of CHR_{max}).

Since the focus of the presented study was to compare the physical training intensity, we employed the same algorithm for cognitive game difficulty adjustment as used by Martin-Niedecken et al. (2019). The main cognitive challenge of the game related to how early players were visually instructed about the direction (right or left) of the upcoming exercise (e.g., a yellow gate rotates to the right for a high touch). If a player performs error-free for 20 s, the cognitive difficulty increased

FIGURE 1 | The ExerCube training (on right) reaches high-intensity training thresholds and is perceived as more motivating, enjoyable, and offering better flow than a conventional functional high-intensity interval training (on left).

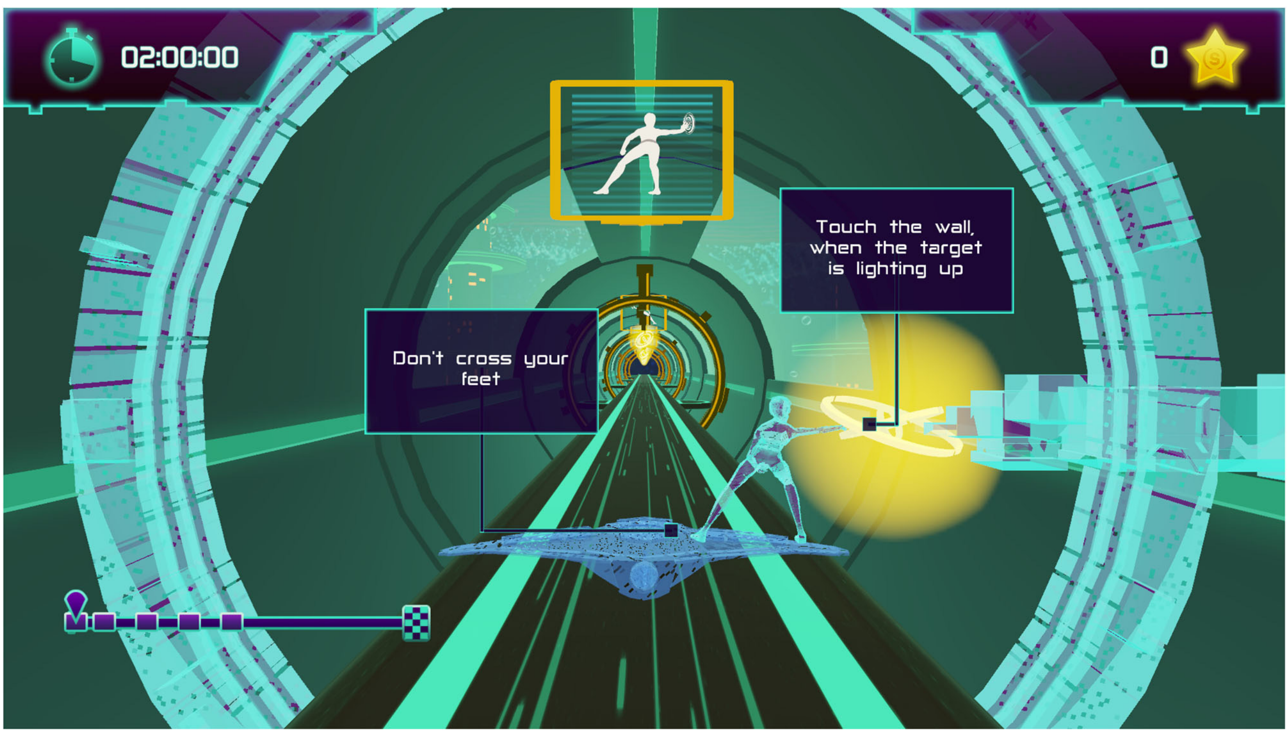


FIGURE 2 | Pitstop tutorial in the ExerCube.

by one level (resulting in a delayed display of the direction of the next exercise) until they made three mistakes within 20 s, inducing a difficulty decrease by one level (resulting in an earlier display of the direction of the next exercise).

However, since the first ExerCube iteration (Martin-Niedecken et al., 2019), a new physical-cognitive challenge was added to the game scenario: players are rewarded with up to three stars depending on their timing. The audio design was

TABLE 1 | Exercise protocols of ExerCube and cfHIIT condition.

	Exercube	cfHIIT
Duration	26–28 min	28 min
Exercises	**Level 1:** Touch, Touch Low, Touch High (L/R)	**Block 1:** Warm-up Routine
	Level 2: + Squat, Jump, Punch (L/R)	**Block 2:** Suicide Drills and Jump Squat
	Level 3: + Lunge (L/R)	**Block 3:** High Knees to Toes and Sumo Squat with Punches
	Level 4: + Skippings	**Block 4:** Mountain Climber and Lunge Jumps
	Level 5: +Burpee	**Block 5:** Burpee with 180° Jump and Skippings and Skater Plyos
Intervals	**Racing:** 2.5–10 min	**Workout time per block:** 4–6 min (8–12 times: 20 s workout and 10 s break)
	Pit stops (breaks): 30 s–2 min	**Break between blocks:** 1.5–2 min
Instructor	Avatar	Coach
Difficulty and intensity	Automatically and individually adapted	Self-regulated

cfHIIT, conventional functional high intensity interval training.

developed further to emphasize the background music's rhythm, emphasize feedback via sound effects, and incorporate audio instructions. Finally, the visual feedback system was iterated for clarity.

Stimuli: Conventional fHIIT

To provide a comparable training protocol, we created a specific cfHIIT (see **Figure 1**) that was as close to actually practiced fHIIT as possible, still comparable to the ExerCube's training protocol. The ExerCube's exercises and intervals thus served as a basis to ensure a similar physical load in both conditions (**Table 1**).

The fHIIT protocol consists of five blocks. It started with a short warm-up block (block 1: 5 min stretching and toning) followed by four interval training blocks, whereby blocks 2–4 included two different exercises (e.g., jump squats and lunge jumps) and block 5 included three different exercises (e.g., skipping). This ensured an increase in physical load toward the end of the training session and matched the ExerCube's last interval. Each (non-warm-up) exercise interval consisted of 20 s of workout (alternately performing the respective exercises) and 10 s of rest. These 30 s workout-rest phases were repeated 8 times (blocks 2–4) or 12 times (block 5), leading to a total duration of 4 min (blocks 2–4) or 6 min (block 5). Interval blocks were separated by short breaks of 1.5 min (blocks 2–4) or 2 min (between blocks 4 and 5). Overall, the cfHIIT lasted about 28 min.

Participants were instructed by the coach that they could individually adapt exercise intensity themselves by choosing a lower or higher level of the initial exercise (e.g., lunges instead of lung jumps) based on their subjective experience of their physical exertion. Additionally, participants were offered a mobile device positioned on the floor nearby the participants that showed their current HR in real time to allow them to keep track of their individual training intensity.

The cfHIIT session was accompanied by music that also functioned as a timer for the intervals. The selected music was specifically composed for HIIT with a pace of 128 beats per minute (bpm), while slower songs were chosen for the breaks in-between intervals to support recovery. The music was played to enhance participants' motivation, to facilitate similar conditions as the ExerCube training (accompanied by a specifically developed and adaptive sound design), and to present a realistic scenario.

Study Design

Two study objectives were determined to investigate the manifestation of objective and subjective components of a single fHIIT exergame session in comparison to a single cfHIIT session. The primary objective of this study was to evaluate the objectively and subjectively experienced training intensity of a single fHIIT exergame session in comparison to a single cfHIIT session. The secondary objective was to assess motivation, flow, and enjoyment during the two training approaches.

The comparative study was set up as a within-subjects design allowing the comparison of two different training methods: an ExerCube vs. cfHIIT session. Whereas, the ExerCube session was performed as a single-player session and controlled by a certified fitness coach, the cfHIIT was performed in small groups of 2–3 participants and instructed by the same coach. Although the ExerCube was mainly self-explanatory, the coach instructed participants and supervised them throughout the session. In the cfHIIT condition, the coach directly instructed the exercises and performed the training session together with the participants. In both sessions, the coach provided corrections, verbal support and cheers, if needed. Moreover, participants did not interact (physically or verbally) with each other in the cfHIIT session.

Participants

The sample size was calculated a priori based on a previous study comparing two HIIT protocols regarding cardiac responses (Schaun and Del Vecchio, 2018). Their study showed the following during exercise: average HR (HR_{avg}) 144.2 ± 11.9 bpm and 130.6 ± 10.4 bpm. Considering an 80% power and 5% significance level, a sample size of 11 subjects would have been necessary. To account for a potentially smaller difference in HR_{avg} between the two training types of our study and regarding possible losses or refusals, the final sample size was set to 16–20 participants.

Twenty participants (10 male, 10 female) were recruited by word-of-mouth and by emailing, without offering any financial compensation for the attendance. The selected study population included healthy young adults (self-reported using a health questionnaire) aged 18–35 years (M = 23.8 years, SD = 3.2). Fifteen participants had experience with exergames, five did not. Participants were excluded from the study if one of the following exclusion criteria was met: (1) history of cardiovascular issues or musculoskeletal injuries that would prevent training participation, (2) asthma (not controllable), (3) pain that would be reinforced by sportive activities, (4) pregnancy.

The recruited participants reported an average exercising time of $M = 300.3$ min/week ($SD = 167$), and reported their subjective fitness as an average of $M = 3.9$ ($SD = 0.9$) on a 6-point scale (1 = poor, 2 = satisfactory, 3 = average, 4 = good, 5 = very good, 6 = competitive). Their resting HR measured $M = 70.3$ bpm ($SD = 9.9$)—we thus calculated their CHR_{max} at $M = 195.8$ ($SD = 2.0$).

Measures

We distinguish between primary outcomes—relating to training intensity—and secondary outcomes, which relate to the qualitative experience of the two training types.

Primary Measures: Training Intensity

HR data were used as an objective measurement of training intensity. The HR recording was assessed during the training session, measuring average and maximal HR (HR_{avg} and HR_{max}). To enable HR data collection, participants wore a HR receiving chest belt of the brand Wahoo (Wahoo Fitness 2014, Atlanta, Georgia, USA)—either connected to and recorded with the ExerCube (log files) or with the compatible "Wahoo RunFit" App, which was installed on an Apple mobile device (.csv files).

The Borg 10-point rating scale was selected as a subjective measurement of training intensity (where 1 = very weak and 10 = very, very strong) (Borg, 1982). This scale was used to assess both physical ($Borg_{physical}$) and cognitive ($Borg_{cognitive}$) perceived exertion.

Secondary Measures: Motivation, Flow, and Enjoyment

We employed the Situational Motivation Scale (SIMS) to assess participants' intrinsic and extrinsic motivation by 16 items (Guay et al., 2000). The SIMS questionnaire comprises four factors: intrinsic motivation, identified regulation, external regulation, and amotivation. The Flow Short Scale (FSS) was used to evaluate participants' flow experience by 13 items (Rheinberg et al., 2003). The flow experience is measured overall and as three factors: fluency of performance, absorption by activity, and perceived importance. Further, we assessed participants' enjoyment of the training via the Physical Activity Enjoyment Scale (PACES), consisting of 18 items (Kendzierski and DeCarlo, 1991; Motl et al., 2001). All three questionnaires were rated on a 7-point Likert scale (SIMS: 1 = corresponds not all, 7 = corresponds exactly; FSS: 1 = not at all, 7 = very much; PACES: (1 = disagree a lot, 7 = agree a lot). These standardized questionnaires were implemented as they are widely used in the area of physical exercising and exergaming and therefore allowed quantifiable comparisons.

Procedure

After study explanation, each participant gave written informed consent. Afterwards, participants filled out a demographic questionnaire to screen for inclusion and exclusion criteria and to assess baseline characteristics such as gender, age, physical activity time, fitness status, and exergame experience. Participants were randomly assigned to one of the two trainings. Each training session lasted 26–28 min. After the training, participants rated their perceived physical and cognitive exertion using the Borg Scale and answered questionnaires covering motivation, flow, and enjoyment. A training session with subsequent questionnaires was then repeated with the other type of training on a different day (but same time of day), after a minimum of 4 days and a maximum of 14 days in-between. The study procedure is illustrated in **Figure 3**.

Analysis

Statistical analysis was conducted in SPSS (IBM SPSS 26). The level of significance was set at $p < 0.05$. The comparison of the data was performed using Wilcoxon signed-rank tests. Wilcoxon signed-rank was used because the assumptions for parametric statistics were not fulfilled. Correlations were calculated using the Spearman correlation coefficient. See Cohen (2013) for an overview of thresholds for correlation coefficients and effect sizes.

RESULTS

Each participant successfully completed both training sessions and all data were considered for further analysis.

Primary Outcomes

Table 2 presents the results from the comparison of the average and maximal measured HR and Borg values between the ExerCube and cfHIIT sessions. Absolute ($z = -2.878$, $p = 0.003$, $r = 0.46$) and relative ($z = -2.837$ $p = 0.005$, $r = 0.45$) average HR values were significantly higher for the cfHIIT session than for the ExerCube training. For the maximal HR, no significant differences were found for absolute ($z = -0.262$, $p = 0.806$, $r = 0.04$) and relative ($z = -0.302$, $p = 0.388$, $r = 0.05$) values. In terms of Borg values, the cfHIIT resulted in a significant higher physical Borg rating ($z = -3.020$, $p = 0.001$, $r = 0.48$) than the ExerCube session. No significant difference was measured for the cognitive Borg ($z = -1.603$, $p = 0.113$, $r = 0.25$).

Secondary Outcomes

The questionnaire data showed significant differences for intrinsic motivation ($z = -3.566$, $p < 0.001$, $r = 0.56$), overall flow score ($z = -3.663$, $p < 0.001$, $r = 0.58$), absorption by activity ($z = -3.436$, $p = 0.001$, $r = 0.54$), perceived importance ($z = -2.518$, $p = 0.012$, $r = 0.40$), and physical activity enjoyment ($z = -3.884$, $p < 0.001$, $r = 0.61$), see **Table 3**. For all of these factors, scores were higher for the ExerCube training session. Additionally, a significant correlation ($r_s = 0.365$, $p = 0.021$) was found between average HR and physical Borg values across all training session data (ExerCube and cfHIIT). No significant correlations were found for $Borg_{physical}$-HR_{max} ($r_s = 0.276$, $p = 0.084$), $Borg_{cognitive}$-HR_{avg} (r_s =-0.224, $p = 0.164$), or Borgcognitive-HRmax (rs = −0.133, $p = 0.412$).

DISCUSSION

Following, we discuss the meaning of our findings in the context of future design and research of effective and attractive fHIIT exergames.

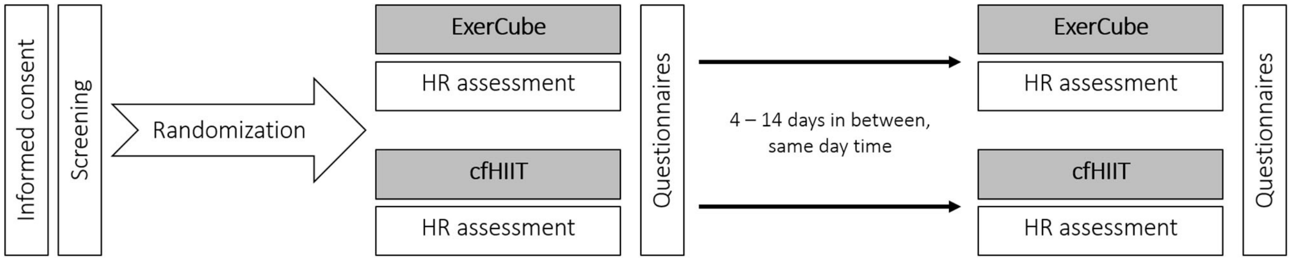


FIGURE 3 | Study procedure for the study comparing fHIIT in an exergame to a conventional small-group class.

TABLE 2 | Comparison of heart rate and Borg values.

	ExerCube	cfHIIT	z	p	r
Average HR [bpm]	155.0 [141.5; 161.3]	159.5 [150.3; 167.0]	−2.878	0.003*	0.46
Average HR as percentage of calculated HRmax	78.7 [72.6; 82.2]	81.1 [77.9; 85.8]	−2.837	0.005*	0.45
Maximal HR [bpm]	182.5 [172.0; 191.0]	180.5 [176.0; 190.8]	−0.262	0.806	0.04
Maximal HR as percentage of calculated HRmax	93.0 [88.7; 97.4]	91.6 [93.6; 101.4]	−0.302	0.388	0.05
$Borg_{physical}$	7.0 [6.0; 8.0]	9.0 [8.0; 9.0]	−3.020	0.001*	0.48
$Borg_{cognitive}$	6.5 [5.0; 8.0]	5.0 [4.0; 6.0]	−1.603	0.113	0.25

*N = 20. Data are shown as median [interquartile range]. Comparisons were calculated using Wilcoxon signed-rank test. *p < 0.05. p-values are exact values and two-tailed. Effect size r, r = 0.10–0.29 indicates a small effect, r = 0.30–0.49 indicates a medium effect and r ≥ 0.50 indicates a large effect. Maximal heart rate was calculate: (HR_{max} = 211 − 0.64 × age). cfHIIT, conventional functional high intensity interval training; HR, heart rate.*

Training Intensity: Effectiveness

The primary interest of the presented work was to objectively and subjectively investigate the intensity of an exergame-based fHIIT with the ExerCube, and thus to explore the feasibility of a specifically designed exergame as a suitable training tool for effective fHIIT. Besides this general proof of feasibility (i.e., reaching the 70–90% range of CHR_{max}), we found implications which seem to be important for future research and development work.

HR-Based Physiological Adaption

For HR_{avg}, the cfHIIT condition showed significant higher HR values compared to the ExerCube condition. This could have been caused by the ExerCube's explorative adaptation algorithm. The physical game difficulty adaptations were triggered by the system, implementing an objective orientation toward 80% CHR_{max} (range: 70–90%) that overlaps with the high intensity zone [80–90% of CHR_{max}, anaerobic zone (78)]. In contrast, adaptations for cfHIIT were triggered by subjective regulations as the participants were allowed to decide on the exercise level themselves, with instruction by the coach.

For safety reasons, the ExerCube's game difficulty adaptation avoided too high HR values with regards to CHR_{max} over a longer period of time, as the exergame is meant to be used in a standalone version without supervision of a coach for the full session. However, fHIIT classes in gyms are always accompanied by a certified coach, and can thus aim for a more persistent high HR_{avg}.

It should also be noted that the CHR_{max} for the ExerCube's physical-difficulty adaption was determined via calculation based on a well-validated generalizable formula (Nes et al., 2013). However, actual HR_{max} can differ; it is a very individual parameter depending on a variety of aspects in addition to age [e.g., gender, fitness level (Nes et al., 2013), and genetics (Wang et al., 2009)]. This could be a reason why the ExerCube's provided training levels did not fully meet respective individual capacity, and thus its HR_{avg} remained slightly below 80% of HR_{max}. Yet the relative HR_{avg} values (% of CHR_{max}) of the ExerCube training did reach values in the fHIIT zone [80–90% of HR_{max} (Edwards, 1994)] for parts of the game duration—and is well-situated in the moderate-intensity zone [70–80% of HR_{max} (Edwards, 1994)]. Furthermore, while the design for safety has to be considered, the measured HR_{max} values show that the ExerCube has the capacity to reach high exercise intensities and to trigger high HR values in young healthy adults.

Future fHIIT exergames should therefore allow for a more individual game difficulty adaption by allowing users to manually insert their pre-assessed individual CHR_{max} or more specific HR prediction models (Ludwig et al., 2018), to then serve as the basis for the algorithm (Hoffmann et al., 2016). In the interest of safety, the implemented explorative algorithm used in the present study could also be refined further to check HR—and if required—adapt more frequently (e.g., every 10–20 s from the very beginning of the game).

Adaptive Training Protocols

Another reason for the significant difference in HR_{avg} could have been the small deviations of the training protocols of both study conditions (i.e., different intervals and sequences in the training structure of the cfHIIT). However, we aimed at designing

TABLE 3 | Comparison of questionnaires.

		ExerCube	cfHIIT	z	p	r
SIMS	Intrinsic motivation	6.5 [5.8; 6.8]	5.1 [4.5; 5.5]	−3.566	<0.001*	0.56
	Identified regulation	6.3 [5.5; 6.7]	6.0 [5.6; 6.7]	−0.029	>0.999	0.01
	External regulation	1.3 [1.0; 2.4]	1.6 [1.3; 2.7]	−0.940	0.367	0.15
	Amotivation	1.0 [1.0; 1.6]	1.3 [1.0; 1.9]	−0.939	0.388	0.15
FSS	Overall	6.0 [5.6; 6.4]	5.4 [4.9; 5.8]	−3.663	<0.001*	0.58
	Fluency of performance	6.3 [5.5; 6.5]	5.7 [5.2; 6.4]	−1.708	0.088	0.27
	Absorption by activity	6.0 [5.5; 6.5]	4.9 [4.5; 5.8]	−3.436	0.001*	0.54
	Perceived importance	1.7 [1.0; 2.2]	1.0 [1.0; 1.8]	−2.519	0.012*	0.40
PACES		6.3 [6.0; 6.6]	5.0 [4.7; 5.5]	−3.884	<0.001*	0.61

*N = 20. Data are shown as median [interquartile range]. Comparisons were calculated using Wilcoxon signed-rank test. *$p < 0.05$. p-values are exact values and two-tailed. Effect size r, r = 0.1 – 0.29 indicates a small effect, r = 0.3 – 0.49 indicates a medium effect and r ≥ 0.5 indicates a large effect. FSS, Flow Short Scale; cfHIIT, conventional functional high intensity interval training; SIMS, Situation Motivation Scale; PACES, Physical activity enjoyment scale.*

a realistic cfHIIT that is comparable to those in fact practiced in the current fitness sector.

It would be interesting to explore different variations of fHIIT protocols in an exergame and link the game difficulty adjustment more closely to the exercises provided in the respective game level. For example, instead of slowing down the game speed when measured HR is too high, provide less exhausting exercises (e.g., holding tasks) and a higher cognitive load.

Effects of Physical-Cognitive Challenge

Another aspect of note is the ExerCube's higher multi-sensory stimulation compared to the cfHIIT, which could have also influenced the HR_{avg}. While the cfHIIT was a single-task training, which required functional movements of participants' own body only, the dual-task training in the ExerCube required participants to concurrently process and react to multi-sensory stimuli (audio-visual, spatial, and game mechanical) while still performing a fHIIT to control the game. This approach constituted more comprehensive executive and attentional functions [(pre-)frontal lobe functions (Funahashi and Andreau, 2013)] than the cfHIIT, and this in turn likely activated more cognitive resources. This was also reflected in the results for perceived exertion of the cognitive domain during our study revealing higher values for the ExerCube compared to the cfHIIT. A subsequent side effect on physical performance can be explained with findings from motor-cognitive research: individuals tend to slow down physical movement when asked to perform a relatively challenging secondary dual-task simultaneously (Yogev-Seligmann et al., 2010). Interestingly, we also found that (while not a significant correlation) participants showed higher HR values for lower cognitive exertion values; in contrast, lower HR values were assessed for higher cognitive load values. This tendency could be caused by the destabilizing effect of dual-tasks that involve competing demands for cognitive and physical resources; this effect is termed "dual-task cost," wherein motor-cognitive interferences can cause deterioration of one or both tasks (Al-Yahya et al., 2011). Thus, we speculate that the multi-sensory stimulation in the ExerCube condition required additional cognitive resources, which in turn limited the physical resources for performance and, therefore, the possibility to reach higher training intensity while also providing additional cognitive training benefits (Benzing et al., 2016; Herold et al., 2018; Stojan and Voelcker-Rehage, 2019).

Future exergame research should further explore and keep in mind these interdependencies. For instance, the ExerCube or a similar exergame could be adapted to investigate effects of varying cognitive loads with a constant physical load on training intensity. Future exergame design should extend existing approaches of physical-cognitive game difficulty adaptation and develop environments that allow for more individualized cognitive-physical and physical-cognitive game challenges. This will allow exergame designers to provide a more individualized dual-domain training (Huang et al., 2014; Hardy et al., 2015; Benzing and Schmidt, 2018; Kappen et al., 2019; Stojan and Voelcker-Rehage, 2019), which could then focus more strongly on either cognitive or on physical challenges depending on the player's needs and skills.

Training Experience: Attractiveness

Additionally of interest to our work was the comparison of the subjective training experience of exergame-based fHIIT with the ExerCube to a cfHIIT. Besides the general proof of feasibility of the ExerCube to be an attractive fHIIT exergame, we again found implications for future research and design toward more appealing exergames.

Shifting Attentional Focus

Regarding training motivation, enjoyment, and flow experience, our study showed significantly higher values in favor of the ExerCube condition. This might have various reasons. In the ExerCube condition, participants' focus seemed to be primarily tied to the game environment and not to their bodily exertion (which indeed was less than in the cfHIIT condition). One indication for this attentional focus shift is the previously discussed higher rated cognitive challenge for the ExerCube condition. Furthermore, flow was significantly higher rated for the ExerCube condition (assessed by FSS) and especially the items "absorption by activity" and "perceived importance." These results match findings of the ExerCube study by Martin-Niedecken et al. (2019), wherein participants reported that they were totally immersed by the game and had to focus on its mechanics to succeed (i.e., a flow experience). In contrast, participants were much more focused on their body with the study's personal coach condition, as they had to concentrate to keep up and perform the exercises correctly. This included more social pressure, i.e., wanting to perform well in front of the coach. These results in combination with those in this study, point toward a trade-off between the two training options. Exergames can provide a degree of playfulness and strong cognitive focus that frees players of the perceived physical and additional social challenges elicited by the presence of coaches. However, coaches provide a degree of guidance and "workout spirit" that leads to greater accuracy in terms of movement; effects that exergames should strive for in their design.

Future exergame design and research should explore in between variations of exergames and trainers to combine the benefits of both approaches (Turmo Vidal et al., 2018).

User-Centered Design

In the exergame condition, game difficulty and complexity were automatically balanced based on each participant's fitness and gaming skills. Thus, in theory, they were never physically over- or under-challenged, nor stressed or bored. Being in this "dual flow" zone is generally considered an optimal workout mode in terms of motivation, enjoyment, and performance (Jackson and Csikszentmihalyi, 1999; Sinclair et al., 2009; Martin-Niedecken and Götz, 2017; Martin-Niedecken et al., 2019). These optimal user-centered demands were reflected by the significant higher rated PACES and the item "intrinsic motivation" (SIMS) for the ExerCube condition, and were likely due to the multi-sensory implementation that was developed and refined by a collaboration of game designers, sport scientists, and the target population in iterations and studies. Co-design allows for including wishes and needs specific to a target group and has been shown to positively affect people's identification with and enjoyment of a product (Birk et al., 2016; Martin-Niedecken and Götz, 2017; Martin-Niedecken et al., 2019). Enjoyment of an activity has a positive impact on physical activity participation and adherence and therefore plays an important role in maintaining an activity-based health care intervention in the long term (Salmon et al., 2003; Hagberg et al., 2009; Rhodes et al., 2009). The results here substantiate the potential of enjoyable exergames for the promotion of physical activity through careful and iterative design and might therefore be a particularly suitable tool for individuals who have trouble motivating themselves to undertake conventional training methods (Wüest et al., 2014; Hoffmann et al., 2016; Moholdt et al., 2017).

Future exergame developments should, therefore, focus on more target population-centric co-designs (i.e., including potential players, but also trainers or therapists in the design process) to ensure that the result meets the players' expectations as well as specific needs and requirements.

Social Exergaming Effects

One difference between our two training stimuli was the single-player mode in the ExerCube session as opposed to the cfHIIT being conducted in a group of 3–4 people, as is common on today's fitness market. We had assumed that this would be a point in favor of the cfHIIT, as social experiences can increase motivation and enjoyment (Campbell et al., 2008; Mueller et al., 2011; Mandryk et al., 2014). However, the questionnaires showed no significant values in favor of cfHIIT. With regards to motivation, flow experience, and enjoyment of physical activity, the ExerCube yielded significantly higher results.

Based on this, it could be assumed that the social factors involved in cfHIIT are not as influential as we had expected. However, we know from related work that in games and exergames the presence of a physical (Emmerich and Masuch, 2018) or virtual (Emmerich and Masuch, 2018; Farrow et al., 2019) co-player or component often enhances player motivation as long as players feel a need to belong (Kaos et al., 2019). This could have increased the experiential quality in the ExerCube. It should also be noted that the social group experience in the cfHIIT also has potential downsides; the social pressure to perform well in front of the coach and other class attendees—similarly observed by Martin-Niedecken et al. (2019) with the personal coach condition—can be overwhelming. Social facilitation is, thus, generally considered positive in exergames, but can instead negatively affect game experiences depending on individual characteristics (e.g., how comfortable is the player at being observed and urged on while working out).

Social facilitation effects in fHIIT exergames are an important aspect for future research. It is in theory possible to play the original Sphery Racer game with the ExerCube in co-located cooperative or competitive mode. This will have to be explored in future work.

Limitations

One limitation of the study consists of the differences between our two training stimuli. While we endeavored to design the two conditions to be as comparable as possible, we also wanted to keep the cfHIIT version as realistic as possible. Thus, the racing sections of the ExerCube are equivalent to with the intervals of the cfHIIT and the lower-intensity pitstops are the equivalent for the resting phases of the cfHIIT. However, there are differences in movement sequence (e.g., repetitive movements per block in the cfHIIT vs. varied movements in the ExerCube) and compositions (e.g., different exercises per block in the cfHIIT vs. accumulated exercises over time in the ExerCube"). This should be considered in future work with the ExerCube.

This also includes not artificially removing potentially beneficial social factors from the cfHIIT condition by exploring this in individual sessions instead of groups. However, as the player experience factors were largely higher for the ExerCube condition, a lack of social factors in exergames may not be as much of an issue as expected. However, it should also be noted that participants experienced the ExerCube for the first time in this study, whereas as some of them had prior experience with cfHIIT group classes. The questionnaire results could thus have been influenced by a novelty effect. Future work has to explore whether this effect remains long-term when players become used to the ExerCube, or when the cfHIIT condition is conducted with pre-existing social groups with prior social bonds (our participants did not know each other or exercise together prior to the study).

Another difference lay in the audio-visual scenarios of the stimuli. The ExerCube provides music that adapts to in-game events, with the addition of sound effects for feedback [shown to be important for exergames in previous work (Martin-Niedecken et al., 2019)]. Our cfHIIT stimuli did have comparable music, however, it was not adaptive beyond following and matching the exercise and rest periods. There are some indications that adaptive music can benefit game experiences (Wharton and Collins, 2011; Rogers and Weber, 2019)—and music certainly positively influences exercise (Karageorghis and Priest, 2012)—nevertheless, this has largely not been explored in the exergame context.

When compared to the exergame condition, another difference is that the cfHIIT featured a degree of subjective self-regulation (participants deciding which exercise version they picked, and how fast and intensely to perform them—albeit with guidance from the coach), while the ExerCube featured more objective adjustments (automatically based on the algorithm). We emphasize that we did offer all participants a mobile device positioned on the floor nearby which showed their current HR in real time. As such, they (including the coach) were in theory able to keep track of their individual training, although we cannot report to what degree they used this option.

Finally, the maximal CHR_{max} in the ExerCube was calculated by a formula that determines relative HR values; these kinds of formulas are based on data from the general population. However, this study was conducted with young healthy adults, whose individual HR_{max} could potentially be higher than what is predicted by the formula. In future work, we will explore whether the determination of individual HR_{max} can provide a more customized, higher training intensity without neglecting safety concerns.

CONCLUSION

The aim of this study was to investigate whether an exergame specifically designed for fHIIT can reach a training intensity comparable with cfHIIT classes and the levels of physical load required for physiological HIIT benefits. Regarding the exergame's training intensity, i.e., its effectiveness, our results reveal that the ExerCube reached high range of physiological training intensity, although the specific adaptation algorithm may need to be adjusted to reach it on average throughout the session. While the cfHIIT yielded higher training intensity (higher average HR) than the ExerCube, participants experienced significantly higher flow, training enjoyment, and motivation in the ExerCube, as well as less perceived physical exertion. Therefore, our results indicate that specifically designed exergames such as the ExerCube are a motivating and enjoyable training approach with the capacity to reach high training intensities. We concluded that exergames or fitness games—if designed properly with regards to fitness protocol (effectiveness) and game design (attractiveness)—have the potential to increase physical activity and training effects to HIIT levels and therefore may be able to facilitate health benefits in young adults. Our results can inform future R&D work which is needed to examine further important aspects in exergames, such as (1) individual and sport-specific determination of physical and cognitive parameters used for pre-game settings and in-game adaptions, (2) refined balancing of cognitive and physical load, and (3) long term effects and training adherence.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

Ethics committee of the ETH Zurich, Switzerland (EK 2015-N-10). Before any measurements were performed, all eligible participants had to sign written informed consent according to the Declaration of Helsinki. The study involved no vulnerable populations. Written informed consent was obtained from the individuals for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

AM-N, AS, and KR drafted the manuscript and provided substantial contributions to the conception and design of the manuscript. AM-N, AS, and AM created the study design and carefully selected the assessment methods. AM-N co-designed the ExerCube stimuli, whilst AM and AS co-created the conventional fHIIT protocol. AM conducted the study (supervised by AS, AM-N, and EB). AS led data analysis and interpretation. AM-N and KR also contributed to the latter. All authors critically reviewed and approved the final manuscript.

ACKNOWLEDGMENTS

The authors kindly thank Sphery Ltd. for providing the infrastructure and technical support throughout the study duration, which ensured a smooth implementation of the study.

REFERENCES

Al-Yahya, E., Dawes, H., Smith, L., Dennis, A., Howells, K., and Cockburn, J. (2011). Cognitive motor interference while walking: a systematic review and meta-analysis. *Neurosci. Biobehav. Rev.* 35, 715–728. doi: 10.1016/j.neubiorev.2010.08.008

Babraj, J. A., Vollaard, N. B., Keast, C., Guppy, F. M., Cottrell, G., and Timmons, J. A. (2009). Extremely short duration high intensity interval training substantially improves insulin action in young healthy males. *BMC endocr. disord.* 9:3. doi: 10.1186/1472-6823-9-3

Barathi, S. C., Finnegan, D. J., Farrow, M., Whaley, A., Heath, P., Buckley, J., et al. (2018). "Interactive feedforward for improving performance and maintaining intrinsic motivation in VR exergaming," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal, QC), 1–14. doi: 10.1145/3173574.3173982

Benzing, V., Heinks, T., Eggenberger, N., and Schmidt, M. (2016). Acute cognitively engaging exergame-based physical activity enhances executive functions in adolescents. *PloS one* 11:e0167501. doi: 10.1371/journal.pone.0167501

Benzing, V., and Schmidt, M. (2018). Exergaming for children and adolescents: strengths, weaknesses, opportunities and threats. *J. clin. med.* 7:422. doi: 10.3390/jcm7110422

Best, J. R. (2015). Exergaming in youth. *Z. Psychol.* 221, 72–78. doi: 10.1027/2151-2604/a000137

Bianchi-Berthouze, N., Kim, W. W., and Patel, D. (2007). "Does body movement engage you more in digital game play? and why?" in *International conference on affective computing and intelligent interaction,* eds A. Paiva, R. Prada, and R. W. Picards (Berlin: Springer), 102–113. doi: 10.1007/978-3-540-74889-2_10

Biddle, S. J., and Asare, M. (2011). Physical activity and mental health in children and adolescents: a review of reviews. *Br. j. sports med.* 45, 886–895. doi: 10.1136/bjsports-2011-090185

Birk, M. V., Atkins, C., Bowey, J. T., and Mandryk, R. L. (2016). "Fostering intrinsic motivation through avatar identification in digital games," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Diego, CA), 2982–2995. doi: 10.1145/2858036.2858062

Borg, G. A. (1982). Psychophysical bases of perceived exertion. *Med. Sci. Sports Exerc*. 14, 377–381. doi: 10.1249/00005768-198205000-00012

Buckley, S., Knapp, K., Lackie, A., Lewry, C., Horvey, K., Benko, C., et al. (2015). Multimodal high-intensity interval training increases muscle function and metabolic performance in females. *Appl. Physiol. Nutr. Metab.* 40, 1157–1162. doi: 10.1139/apnm-2015-0238

Burgomaster, K. A., Howarth, K. R., Phillips, S. M., Rakobowchuk, M., MacDonald, M. J., McGee, S. L., et al. (2008). Similar metabolic adaptations during exercise after low volume sprint interval and traditional endurance training in humans. *J. physiol.* 586, 151–160. doi: 10.1113/jphysiol.2007.142109

Byrne, A. M., and Kim, M. (2019). The Exergame as a tool for mental health treatment. *J. Creat. Mental Health* 14, 465–477. doi: 10.1080/15401383.2019.1627263

Campbell, T., Ngo, B., and Fogarty, J. (2008). "Game design principles in everyday fitness applications," in *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work* (San Diego, CA), 249–252. doi: 10.1145/1460563.1460603

Caria, M. A., Tangianu, F., Concu, A., Crisafulli, A., and Mameli, O. (2007). Quantification of spinning® bike performance during a standard 50-minute class. *J. sports sci.* 25, 421–429. doi: 10.1080/02640410600718533

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. New York, NY: Routledge. doi: 10.4324/9780203771587

de Bruin, E. D., Rebsamen, S., Knols, R., and Baschung Pfister, P. (2019). Exergame-driven high-intensity interval training in untrained community dwelling older adults: a formative one group quasi-experimental feasibility trial. *Front. physiol.* 10:1019. doi: 10.3389/fphys.2019.01019

Edwards, S. (1994). *The Heart Rate Monitor Book.* Sacramento, CA: Fleet Feet Press. doi: 10.1249/00005768-199405000-00020

Emmerich, K., and Masuch, M. (2018). "Watch me play: does social facilitation apply to digital games?," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal, QC), 1–12. doi: 10.1145/3173574.3173674

Farrow, M., Lutteroth, C., Rouse, P. C., and Bilzon, J. L. (2019). Virtual-reality exergaming improves performance during high-intensity interval training. *Eur. j. sport sci.* 19, 719–727. doi: 10.1080/17461391.2018.1542459

Feito, Y., Heinrich, K. M., Butcher, S. J., and and, W.,Poston, S. C. (2018). High-intensity functional training (HIFT): definition and research implications for improved fitness. *Sports* 6:76. doi: 10.3390/sports6030076

Folsom, A. R., Kushi, L. H., and Hong, C. P. (2000). Physical activity and incident diabetes mellitus in postmenopausal women. *Am. J. Public Health* 90:134. doi: 10.2105/AJPH.90.1.134

Foster, C., Farland, C. V., Guidotti, F., Harbin, M., Roberts, B., Schuette, J., et al. (2015). The effects of high intensity interval training vs steady state training on aerobic and anaerobic capacity. *J. sports sci. med.* 14, 747–755. doi: 10.1249/01.mss.0000476771.63318.52

Funahashi, S., and Andreau, J. M. (2013). Prefrontal cortex and neural mechanisms of executive function. *J. Physiol. Paris* 107, 471–482. doi: 10.1016/j.jphysparis.2013.05.001

Gibala, M. J., Little, J. P., MacDonald, M. J., and Hawley, J. A. (2012). Physiological adaptations to low-volume, high-intensity interval training in health and disease. *J. physiol.* 590, 1077–1084. doi: 10.1113/jphysiol.2011.224725t

Guay, F., Vallerand, R. J., and Blanchard, C. (2000). On the assessment of situational intrinsic and extrinsic motivation: the Situational Motivation Scale (SIMS). *Motiv. emot.* 24, 175–213. doi: 10.1023/A:1005614228250

Hagberg, L., Lindahl, B., Nyberg, L., and Hellénius, M. L. (2009). Importance of enjoyment when promoting physical exercise. *Scand. j. med. sci. sports* 19, 740–747. doi: 10.1111/j.1600-0838.2008.00844.x

Haller, J. C., Jang, Y. H., Haller, J., Shaw, L., and Wünsche, B. C. (2019). "HIIT the road: using virtual spectator feedback in HIIT-based exergaming," in *Proceedings of the Australasian Computer Science Week Multiconference* (Sydney, NSW), 1–9. doi: 10.1145/3290688.3290752

Hardy, S., Dutz, T., Wiemeyer, J., Göbel, S., and Steinmetz, R. (2015). Framework for personalized and adaptive game-based training programs in health sport. *Multimedia Tools Appl.* 74, 5289–5311. doi: 10.1007/s11042-014-2009-z

Herold, F., Hamacher, D., Schega, L., and Müller, N. G. (2018). Thinking while moving or moving while thinking–concepts of motor-cognitive training for cognitive performance enhancement. *Front. aging neurosci.* 10:228. doi: 10.3389/fnagi.2018.00228

Hoffmann, K., Wiemeyer, J., and Hardy, S. (2016). "Prediction and control of the individual heart rate response in exergames," in *Proceedings of the 10th International Symposium on Computer Science in Sports (ISCSS)* (Loughborough), 171–178. doi: 10.1007/978-3-319-24560-7_22

Huang, T., Larsen, K., M., Ried-Larsen, Møller, N., and Andersen, L. B. (2014). The effects of physical activity and exercise on brain-derived neurotrophic factor in healthy humans: a review. *Scand. j. med. sci. sports* 24, 1–10. doi: 10.1111/sms.12069

Isbister, K. (2016). *How games move us: Emotion by design.* Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9267.001.0001

Jackson, S. A., and Csikszentmihalyi, M. (1999). *Flow in sports Human Kinetics*. Champaign, IL: Human Kinetics.

Joronen, K., Aikasalo, A., and Suvitie, A. (2017). Nonphysical effects of exergames on child and adolescent well-being: a comprehensive systematic review. *Scand. j. caring sci.* 31, 449–461. doi: 10.1111/scs.12393

Kajastila, R., and Hämäläinen, P. (2015). Motion games in real sports environments. *interactions* 22, 44–47. doi: 10.1145/2731182

Kaos, M. D., Rhodes, R. E., Hämäläinen, P., and Graham, T. N. (2019). "Social play in an exergame: how the need to belong predicts adherence," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow), 1–13. doi: 10.1145/3290605.3300660

Kappen, D. L., Mirza-Babaei, P., and Nacke, L. E. (2019). Older adults' physical activity and exergames: a systematic review. *Int. J. Hum. Comput. Interact.* 35, 140–167. doi: 10.1080/10447318.2018.1441253

Karageorghis, C. I., and Priest, D. L. (2012). Music in the exercise domain: a review and synthesis (part I). *Int. rev. sport exerc. psychol.* 5, 44–66. doi: 10.1080/1750984X.2011.631026

Keesing, A., Ooi, M., Wu, O., Ye, X., Shaw, L., and Wünsche, B. C. (2019). "Hiit with hits: using music and gameplay to induce hiit in exergames," in *Proceedings of the Australasian Computer Science Week Multiconference* (Sydney), 1–10. doi: 10.1145/3290688.3290740

Kemi, O. J., and Wisløff, U. (2010). High-intensity aerobic exercise training improves the heart in health and disease. *J. cardiopulm. rehabil. prev.* 30, 2–11. doi: 10.1097/HCR.0b013e3181c56b89

Kendzierski, D., and DeCarlo, K. J. (1991). Physical activity enjoyment scale: two validation studies. *J. sport exerc. psychol.* 13, 50–64. doi: 10.1123/jsep.13.1.50

Kilpatrick, M. W., Jung, M. E., and Little, J. P. (2014). High-intensity interval training: a review of physiological and psychological responses. *ACSM's Health Fit. J.* 18, 11–16. doi: 10.1249/FIT.0000000000000067

Kim, Y. S., Prestopnik, N., and Biocca, F. A. (2014). Body in the interactive game: how interface embodiment affects physical activity and health behavior change. *Comput. Hum. Behav.* 36, 376–384. doi: 10.1016/j.chb.2014.03.067

Kliszczewicz, B., McKenzie, M., and Nickerson, B. (2019). Physiological adaptation following four-weeks of high-intensity functional training. *Vojnosanitetski pregled* 76, 272–277. doi: 10.2298/VSP170228095K

Lee, S., Kim, W., Park, T., and Peng, W. (2017). The psychological effects of playing exergames: a systematic review. *Cyberpsychol. Behav. Soc. Netw.* 20, 513–532. doi: 10.1089/cyber.2017.0183

Li, J.,Theng, Y.-L., and Foo, S. (2016). Effect of exergames on depression: a systematic review and meta-analysis. *Cyberpsychol. Behav. Soc. Netw.* 19, 34–42. doi: 10.1089/cyber.2015.0366

Lu, A. S., Kharrazi, H., Gharghabi, F., and Thompson, D. (2013). A systematic review of health videogames on childhood obesity prevention and intervention. *Games Health J.* 2, 131–141. doi: 10.1089/g4h.2013.0025

Ludwig, M., Hoffmann, K., Endler, S., Asteroth, A., and Wiemeyer, J. (2018). Measurement, prediction, and control of individual heart rate responses to exercise—basics and options for wearable devices. *Front. physiol.* 9:778. doi: 10.3389/fphys.2018.00778

Lyons, E. J. (2015). Cultivating engagement and enjoyment in exergames using feedback, challenge, and rewards. *Games health j.* 4, 12–18. doi: 10.1089/g4h.2014.0072

MacInnis, M. J., and Gibala, M. J. (2017). Physiological adaptations to interval training and the role of exercise intensity. *J. physiol.* 595, 2915–2930. doi: 10.1113/JP273196

Macvean, A., and Robertson, J. (2013). "Understanding exergame users' physical activity, motivation and behavior over time," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris), 1251–1260. doi: 10.1145/2470654.2466163

Mandryk, R. L., Gerling, K. M., and Stanley, K. G. (2014). "Designing games to discourage sedentary behaviour," in *Playful User Interfaces* (Singapore: Springer), 253–274. doi: 10.1007/978-981-4560-96-2_12

Márquez Segura, E., Waern, A., Moen, J., and Johansson, C. (2013). "The design space of body games: technological, physical, and social design," in *Proceedings of the SIGCHI conference on Human Factors in computing systems* (Paris), 3365–3374. doi: 10.1145/2470654.2466461

Marshall, J., Benford, S., and Pijnappel, S. (2016). Expanding exertion gaming. *Int. J. Hum. Comp. Stud.* 90, 1–13. doi: 10.1016/j.ijhcs.2016.02.003

Martin-Niedecken, A. L., and Götz, U. (2017). "Go with the dual flow: evaluating the psychophysiological adaptive fitness game environment "Plunder Planet," in *Joint International Conference on serious games* (Valencia), 32–43. doi: 10.1007/978-3-319-70111-0_4

Martin-Niedecken, A. L., and Mekler, E. D. (2018). "The ExerCube: participatory design of an immersive fitness game environment," in *Joint International Conference on Serious Games* (Darmstadt), 263–275. doi: 10.1007/978-3-030-02762-9_28

Martin-Niedecken, A. L., Rogers, K., Turmo Vidal, L., Mekler, E. D., and Márquez Segura, E. (2019). "ExerCube vs. personal trainer: evaluating a holistic, immersive, and adaptive fitness game setup," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow), 1–15. doi: 10.1145/3290605.3300318

McRae, G., Payne, A., Zelt, J. G., Scribbans, T. D., Jung, M. E., Little, J. P., et al. (2012). Extremely low volume, whole-body aerobic–resistance training improves aerobic fitness and muscular endurance in females. *Appl. Physiol. Nutr. Metab.* 37, 1124–1131. doi: 10.1139/h2012-093

Menz, V., Marterer, N., Amin, S. B., Faulhaber, M., Hansen, A. B., and Lawley, J. S. (2019). Functional vs. running low-volume high-intensity interval training: effects on VO2max and muscular endurance. *J. sports sci. med.* 18, 497–504.

Milanović, Z., Sporiš, G., and Weston, M. (2015). Effectiveness of high-intensity interval training (HIT) and continuous endurance training for VO2max improvements: a systematic review and meta-analysis of controlled trials. *Sports med.* 45, 1469–1481. doi: 10.1007/s40279-015-0365-0

Moholdt, T., Weie, S., Chorianopoulos, K., Wang, A. I., and Hagen, K. (2017). Exergaming can be an innovative way of enjoyable high-intensity interval training. *BMJ open sport exerc. med.* 3:e000258. doi: 10.1136/bmjsem-2017-000258

Motl, R. W., Dishman, R. K., Saunders, R., Dowda, M., Felton, G., and Pate, R. R. (2001). Measuring enjoyment of physical activity in adolescent girls. *Am. j. prev. med.* 21, 110–117. doi: 10.1016/S0749-3797(01)00326-9

Mueller, F., and Isbister, K. (2014). "Movement-based game guidelines," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, ON), 2191–2200. doi: 10.1145/2556288.2557163

Mueller, F. F., Edge, D., Vetere, F., Gibbs, M. R., Agamanolis, S., Bongers, B., et al. (2011). "Designing sports: a framework for exertion games," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC), 2651–2660. doi: 10.1145/1978942.1979330

Nes, B., Janszky, I., Wisløff, U., Støylen, A., and Karlsen, T. (2013). Age-predicted maximal heart rate in healthy subjects: the HUNT fitness study. *Scand. j. med. sci. sports* 23, 697–704. doi: 10.1111/j.1600-0838.2012.01445.x

Oh, Y., and Yang, S. (2010). "Defining exergames and exergaming," in *Proceedings of Meaningful Play*, 1–17.

Pasch, M., Bianchi-Berthouze, N., van Dijk, B., and Nijholt, A. (2009). Movement-based sports video games: investigating motivation and gaming experience. *Entertain. Comput.* 1, 49–61. doi: 10.1016/j.entcom.2009.09.004

Peng, W. J., Lin, H., and Crouse, J. (2011). Is playing exergames really exercising? A meta-analysis of energy expenditure in active video games. *Cyberpsychol. Behav. Soc. Netw.* 14, 681–688. doi: 10.1089/cyber.2010.0578

Rheinberg, F., Vollmeyer, R., and Engeser, S. (2003). "Measuring flow experiences," in *Diagnostik von Motivation und Selbstkonzept, Test und Trends*, eds J. Stiensmeier-Pelster and F. Rheinberg (Göttingen: Hogrefe), 261–279.

Rhodes, R. E., Warburton, D. E., and Bredin, S. S. (2009). Predicting the effect of interactive video bikes on exercise adherence: an efficacy trial. *Psychol. health med.* 14, 631–640. doi: 10.1080/13548500903281088

Rogers, K., and Weber, M. (2019). "Audio habits and motivations in video game players," in *Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound on ZZZ* (Nottingham), 45–52. doi: 10.1145/3356590.3356599

Roy, M., Williams, S. M., Brown, R. C., Meredith-Jones, K. A., Osborne, H., Jospe, M., et al. (2018). High-intensity interval training in the real world: outcomes from a 12-month intervention in overweight adults. *Med. Sci. Sports Exerc.* 50, 1818–1826. doi: 10.1249/MSS.0000000000001642

Salmon, J., Owen, N., Crawford, D., Bauman, A., and Sallis, J. F. (2003). Physical activity and sedentary behavior: a population-based study of barriers, enjoyment, and preference. *Health psychol.* 22:178. doi: 10.1037/0278-6133.22.2.178

Schaun, G. Z., and Del Vecchio, F. B. (2018). High-intensity interval exercises' acute impact on heart rate variability: comparison between whole-body and cycle ergometer protocols. *J. Strength Cond. Res.* 32, 223–229. doi: 10.1519/JSC.0000000000002180

Shafer, D. M., Carbonara, C. P., and Popova, L. (2014). Controller required? The impact of natural mapping on interactivity, realism, presence, and enjoyment in motion-based video games. *Presence Teleoper. Virtual Environ.* 23, 267–286. doi: 10.1162/PRES_a_00193

Shaw, L. A., Wuensche, B. C., Lutteroth, C., Buckley, J., and Corballis, P. (2017). "Evaluating sensory feedback for immersion in exergames," in *Proceedings of the Australasian Computer Science Week Multiconference* (Geelong), 1–6. doi: 10.1145/3014812.3014823

Sinclair, J., Hingston, P., and Masek, M. (2007). "Considerations for the design of exergames," in *Proceedings of the 5th international conference on Computer graphics and interactive techniques in Australia and Southeast Asia* (Perth, WA), 289–295. doi: 10.1145/1321261.1321313

Sinclair, J., Hingston, P., and Masek, M. (2009). "Exergame development using the dual flow model," in *Proceedings of the Sixth Australasian Conference on Interactive Entertainment* (Sydney, NSW), 1–7. doi: 10.1145/1746050.1746061

Steinbeck, K. S. (2001). The importance of physical activity in the prevention of overweight and obesity in childhood: a review and an opinion. *Obes. rev.* 2, 117–130. doi: 10.1046/j.1467-789x.2001.00033.x

Stojan, R., and Voelcker-Rehage, C. (2019). A systematic review on the cognitive benefits and neurophysiological correlates of exergaming in healthy older adults. *J. clin. med.* 8:734. doi: 10.3390/jcm8050734

Tabata, I., Nishimura, K., Kouzaki, M., Hirai, Y., Ogita, F., Miyachi, M., et al. (1996). Effects of moderate-intensity endurance and high-intensity intermittent training on anaerobic capacity and VO_2max. *Med. Sci. Sports Exerc.* 28, 1327–1330. doi: 10.1097/00005768-199610000-00018

Teixeira, P. J., Carraça, E. V., Markland, D., Silva, M. N., and Ryan, R. M. (2012). Exercise, physical activity, and self-determination theory: a systematic review. *Int. j. behav. nutr. phys. act.* 9:78. doi: 10.1186/1479-5868-9-78

Thompson, W. (2018). Worldwide survey of fitness trends for 2019. *ACSM's Health Fit. J.* 22, 10–17. doi: 10.1249/FIT.0000000000000438

Thompson, W. R., Gordon, N. F., and Pescatello, L. S. (2010). *ACSM's guidelines for exercise testing and prescription.* Philadelphia, PA: Lippincott Williams and Wilkins.

Tondello, G. F., Mora, A., Marczewski, A., and Nacke, L. E. (2019). Empirical validation of the gamification user types hexad scale in English and Spanish. *Int. J. Hum. Comput. Stud.* 127, 95–111. doi: 10.1016/j.ijhcs.2018.10.002

Trost, S. G., Owen, N., Bauman, A. E., Sallis, J. F., and Brown, W. (2002). Correlates of adults' participation in physical activity: review and update. *Med. sci. sports exerc.* 34, 1996–2001. doi: 10.1097/00005768-200212000-00020

Turmo Vidal, L., Márquez Segura, E., and Waern, A., (2018). "Movement correction in instructed fitness training: design recommendations and opportunities," in *Proceedings of the 2018 Designing Interactive Systems Conference* (Hong Kong), 1041–1054. doi: 10.1145/3196709.3196789

Valenzuela, T., Okubo, Y., Woodbury, A., Lord, S. R., and Delbaere, K. (2018). Adherence to technology-based exercise programs in older adults: a systematic review. *J. Geriatr. Phys. Ther.* 41, 49–61. doi: 10.1519/JPT.0000000000000095

Wang, X., Ding, X., Su, S., Li, Z., Riese, H., Thayer, J. F., et al. (2009). Genetic influences on heart rate variability at rest and during stress. *Psychophysiology* 46, 458–465. doi: 10.1111/j.1469-8986.2009.00793.x

Weiss, T., Kreitinger, J., Wilde, H., Wiora, C., Steege, M., Dalleck, L., et al. (2010). Effect of functional resistance training on muscular fitness outcomes in young adults. *J. Exerc. Sci. Fit.* 8, 113–122. doi: 10.1016/S1728-869X(10)60017-2

Weston, K. S., Wisløff, U., and Coombes, J. S. (2014). High-intensity interval training in patients with lifestyle-induced cardiometabolic disease: a systematic review and meta-analysis. *Br. J. Sports Med.* 48, 1227–1234. doi: 10.1136/bjsports-2013-092576

Wharton, A., and Collins, K. (2011). Subjective measures of the influence of music customization on the video game play experience: a pilot study. *Game Stud. int. j. comput. game res.* 11.

Wilke, J., Kaiser, S., Niederer, D., Kalo, K., Engeroff, T., Morath, C., et al. (2019). Effects of high-intensity functional circuit training on motor function and sport motivation in healthy, inactive adults. *Scand. j. med. sci. sports* 29, 144–153. doi: 10.1111/sms.13313

World Health Organization (2010). *Global recommendations on physical activity for health*. World Health Organization.

Wüest, S., van de Langenberg, R., and de Bruin, E. D. (2014). Design considerations for a theory-driven exergame-based rehabilitation program to improve walking of persons with stroke. *Eur. Rev. Aging Phys. Act.* 11, 119–129. doi: 10.1007/s11556-013-0136-6

Xiong, S., Zhang, P., and Gao, Z. (2019). Effects of Exergaming on Preschoolers' Executive Functions and Perceived Competence: a Pilot Randomized Trial. *J. clin. med.* 8:469. doi: 10.3390/jcm8040469

Yogev-Seligmann, G., Rotem-Galili, Y., Mirelman, A., Dickstein, R., Giladi, N., and Hausdorff, J. M. (2010). How does explicit prioritization alter walking during dual-task performance? Effects of age and sex on gait speed and variability. *Phys. ther.* 90, 177–186. doi: 10.2522/ptj.20090043

Conflict of Interest: AM-N is the Co-Founder and CEO of the startup company Sphery Ltd. that developed the ExerCube. AS has been working at Sphery since November 2019, but was not yet employed at the time of the study conduction. No revenue was paid (or promised to be paid) directly to AM-N, to Sphery or the research institutions.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 15 January 2021
doi: 10.3389/fpsyg.2020.01902

Check for updates

Me, Myself, and Not-I: Self-Discrepancy Type Predicts Avatar Creation Style

*Mitchell G. H. Loewen[1,2], Christopher T. Burris[3] and Lennart E. Nacke[2,4]**

[1] Department of Psychology, University of Waterloo, Waterloo, ON, Canada, [2] HCI Games Group, The Games Institute, University of Waterloo, Waterloo, ON, Canada, [3] Department of Psychology, St. Jerome's University, Waterloo, ON, Canada, [4] Department of Communication Arts & Stratford School of Interaction Design and Business, University of Waterloo, Waterloo, ON, Canada

In video games, identification with avatars—virtual entities or characters driven by human behavior—has been shown to serve many interpersonal and intraindividual functions (like social connection, self-expression, or identity exploration) but our understanding of the psychological variables that influence players' avatar choices remains incomplete. The study presented in this paper tested whether players' preferred style of avatar creation is linked to the magnitude of self-perceived discrepancies between who they are, who they aspire to be, and who they think they should be. One-hundred-and-twenty-five undergraduate gamers indicated their preferred avatar creation style and completed a values measure from three different perspectives: their actual, ideal, and ought selves. The average actual/ideal values discrepancy was greater among those who preferred idealized avatars vs. those who preferred realistic avatars. The average actual/ought values discrepancy was greater among those who preferred completely different avatars (i.e., fantasy/role-players) vs. those who preferred realistic avatars. These results, therefore, offer additional evidence that self-discrepancy theory is a useful framework for understanding avatar preferences.

Keywords: avatars, self-discrepancy theory, values, video games, self-perception

OPEN ACCESS

Edited by:
Marc Herrlich,
University of Kaiserslautern, Germany

Reviewed by:
Daniel Roth,
Technical University of Munich, Germany
Joan Llobera,
Universitat de Barcelona, Spain

Correspondence:
Lennart E. Nacke
lennart.nacke@acm.org

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Psychology

Received: *31 January 2020*
Accepted: *10 July 2020*
Published: *15 January 2021*

Citation:
Loewen MGH, Burris CT and Nacke LE (2021) Me, Myself, and Not-I: Self-Discrepancy Type Predicts Avatar Creation Style. Front. Psychol. 11:1902. doi: 10.3389/fpsyg.2020.01902

1. INTRODUCTION

Over the past few decades, video games have become an integral component of popular culture and currently generate more revenue than the Hollywood movie industry (Nath, 2016). Unlike movies, many video games allow players to interact with and experience a digital environment via avatars. Although the use of the term "avatar" dates back to early multi-user dungeon text-based games, its application in virtual worlds and consequently video games designates control by a human actor (see Bailenson and Blascovich, 2004) rather than the artificial intelligence characteristic of game agents, often referring to digital non-player characters or entities, whose behavior is controlled through algorithms (Roth et al., 2017; Waltemate et al., 2018).

The first-person perspective provided by avatars facilitates a level of identification (Cohen, 2001) with the characters in the game that surpasses alternative visual media, such as film and television (Klimmt et al., 2009). Adopting a first-person perspective hinges on a sense of virtual embodiment, which is facilitated by an avatar's behavioral and photographic realism (Bailenson and Blascovich, 2004). The implications of such virtual embodiment or self-presence (Biocca, 1997; Jin and Park, 2009; Slater et al., 2009) are often considerable. For example, people will often conform to their digital self-representation even when that virtual body is unlike their own (cf. the so-called Proteus effect: Yee and Bailenson, 2007).

Consequently, avatars have been shown to serve a variety of psychological functions: interpersonal ones, such as social connection (e.g., Lomanowska and Guitton, 2014; Song and Fox, 2016) as well as intraindividual ones, such as self-expression (e.g., Sung et al., 2011) and identity exploration (e.g., Bessière et al., 2007; Hefner et al., 2007; Klimmt et al., 2010). The potential for positive applications of avatars in addressing domains, such as health and well-being (e.g., Fox and Bailenson, 2009; Jin, 2011; Behm-Morawitz, 2013) and the reduction of implicit racial bias (e.g., Banakou et al., 2016) have also been explored.

With the above as a backdrop, researchers have begun to identify factors that guide an individual's choices concerning how an avatar is chosen and/or customized, with a recurrent focus on the extent to which avatars resemble their users in physical and/or psychological characteristics. Mancini and Sibilla (2017, p. 275) recently stated that "there is at present no consensus on this issue, some studies have reported that players use their offline self as a starting point for the construction of their characters, and some others reported that players sometimes build characters which are totally disconnected from their offline self."

In addition, the concept of self-presence helps frame the effect a player's virtual self can have on the "perception of one's body (i.e., body schema or body image), physiological states, emotional states, perceived traits, and identity" (Biocca, 1997). This led to research suggesting that an avatar could have a positive influence on well-being and health appearance and behaviors (Fox and Bailenson, 2009; Jin, 2011; Behm-Morawitz, 2013). Along those lines, studies investigated the concept of parasocial interaction with an avatar, where self-presence was understood as "the extent of game players' interpersonal involvement with their avatar and the extent to which game players perceive themselves as interacting with the avatar" (Jin and Park, 2009).

Paralleling Mancini and Sibilla (2017) and other researchers (e.g., Dunn and Guadagno, 2012; Villani et al., 2016), we suggest that self-discrepancy theory (SDT) (Higgins, 1987) provides a useful framework for making sense of avatar choices. According to SDT, the self can be understood in terms of three domains:

1. the *actual self* (the attributes that someone—self or other—perceives that the target person actually possesses);
2. the *ideal self* (the attributes that someone wants the target person to possess), and
3. the *ought self* (the attributes that someone believes that the target person should possess).

Discrepancies between the actual self and either self-guide (i.e., the ideal self or the ought self) have been posited to evoke distinct classes of emotions as well as motivation to resolve the perceived discrepancies.

As Mancini and Sibilla (2017) themselves noted, extant avatar research inspired by SDT has focused on the actual and ideal selves; the ought self has been overlooked. We posit that discrepancies centered on the ought self may help explain why some individuals prefer avatars that are "totally disconnected from their offline self" (p. 275), however. Testing this possibility was a primary goal of the present research.

Our methodology incorporated an operationalization of Neustaedter and Fedorovskaya (2009)'s avatar preference typology. Within their framework, so-called "Realistics" aim for continuity between their digital and real-life selves by attempting to make the former similar to the latter in terms of appearance. "Ideals" are more selective, as their constructed avatars reflect the "best" parts of themselves, and/or traits and characteristics to which they aspire. In contrast, "Fantasies" and "Roleplayers" maintain a clear distinction between their real and virtual selves, which often differ markedly. Indeed, an adopted avatar allows the player an opportunity to explore virtual worlds through the eyes of a persona quite unlike themselves: Fantasies do so via one's virtual self, whereas Roleplayers do so via multiple virtual selves. Overall, then, the differences between chosen avatars and their player-creators range from minimal (Realistic) to moderate (Ideal) to substantial (Fantasy/Roleplayer).

In SDT terms, we would expect Ideals to be more likely to perceive a discrepancy between their actual and ideal selves compared to Realistics: Ideals' avatars reflect the "best" parts of players and/or traits and characteristics to which they aspire, whereas the avatars of Realistics are arguably less aspirational, instead closely resembling the players themselves. Like those of Ideals, Fantasies/Roleplayers' virtual selves differ from their real-world personas, but the differences are so great that they are not likely to be the result of simple aspiration. Consequently, we would have no clear conceptual basis for predicting that the average actual-ideal self-discrepancy of Fantasies/Roleplayers would differ from those of Realistics.

We would, however, expect Fantasies/Roleplayers to be more likely to perceive actual-ought self-discrepancies relative to Realistics. Actual-ought self-discrepancies have been linked conceptually to resentment and fear of negative social evaluation (Higgins, 1987). Consequently, identifying with an avatar that is wholly different from the real-world self-amidst the protective anonymity of virtual environments may allow Fantasies/Roleplayers dealing with actual-ought self-discrepancies to explore and express personal attributes perceived to be too taboo or risky to own or express in the real world (Crenshaw and Nardi, 2014; Mancini and Sibilla, 2017). In contrast, we would have no conceptual basis for expecting Ideals and Realistics to differ with respect to actual-ought self-discrepancies.

To test these hypotheses, participants who had created or customized at least one avatar as part of an online gaming experience completed a brief, cross-culturally validated values measure three times—that is, from the perspective of their actual, ideal, and ought selves (meaning: completion of a short personality inventory from the perspective of the actual, ideal, and avatar selves in Mancini and Sibilla, 2017). They also indicated their preferred style of avatar creation based on descriptions adapted from Neustaedter and Fedorovskaya (2009). We subsequently computed actual-ideal and actual-ought values discrepancy scores and compared the resulting means among three groups: Realistics, Ideals, and Fantasies/Roleplayers (henceforth referred to as "Differents")[1].

2. METHOD

In this section, all dependent measures, conditions, and data exclusions are reported. The sample size was maximized in the context of practical and temporal constraints—specifically, the number of study-specific volunteer slots allotted by the research pool coordinator coupled with the first author's fixed timeframe for completing his thesis on which this report is based. The study received ethics approval from the ethics board at the authors' home institution.

2.1. Participants

One hundred fifty-two undergraduates registered with the psychology research pool at the University of Waterloo agreed to participate in an online study described as investigating "the relationship between video game players and their in-game avatars to better understand how and why players create the avatars that they do" in exchange for extra course credit. Would-be participants were asked to sign up for the study only if they had previously played a Massively Multiplayer Online Game (MMOG) that involved avatar creation. Median age was 20, with 97% of the sample between ages 17 and 24.

Twenty-seven participants were excluded from the final data set. Specifically, based on their responses to the screening questions, three had not played a game that involved avatar customization and seven did not provide a name and/or description of the game they had played. In addition, 13 participants reported nearly identical responses (e.g., "6") to every question across all selves and avatars (suggesting inattention to item content), one did not respond to the avatar-creation-style item, one did not complete the ought-self measure, one completed the study twice (and so the second set of responses was removed), and one did not complete any of the key measures. Thus, the final sample consisted of 125 participants (51 female, 72 male, 2 other/no response; 38% Euro-Canadian, 38% East Asian, 24% other).

2.2. Materials

2.2.1. Screening Questions

To ensure that participants met inclusion criteria (see above), they were asked to specify which MMOG they had played the most, to provide a brief description of the game, and to describe if the game allowed for avatar customization (the three MMOGs most frequently listed by final-sample participants were *MapleStory* [$n = 19$], *World of Warcraft* [$n = 19$], and *RuneScape* [$n = 14$]; no other listed game exceeded $n = 5$). They were also asked how long (in months) they had played the specified game ($M = 18.24$; $SD = 21.28$).

2.2.2. Self-Discrepancy Measure

To assess the magnitude of participants' actual/ideal and actual/ought self-discrepancies across a broad, significant personal domain, participants were asked to complete the short version of Schwartz's Value Survey (SSVS) as presented in Lindeman and Verkasalo (2005) (see also **Appendix**) under three different instructional sets using a −2 to 14 response range with the following anchors: −2 = opposed to my values; 0 = not important; 6 = important; 12 = very important; 14 = of supreme importance. Thus, participants rated the importance of values, such as "power" and "self-direction" from the perspective of: (1) the actual self (i.e., "how you truly see yourself"); (2) the ideal self (i.e., "how you would like to be"); and (3) the ought self (i.e., "how you think others think you should be").

Mean actual-ideal and actual-ought discrepancy scores for each individual were computed by averaging the absolute values of the discrepancy scores for each of the 10 relevant values pairs (e.g., *actual Hedonism—ideal Hedonism*, or *actual Benevolence—ought Benevolence*). Given this computational strategy and the fact that the SSVS uses single items to assess each of the 10 values represented within Schwartz's circumplex model, an internal consistency coefficient could not be computed (but see Lindeman and Verkasalo, 2005 for psychometric information concerning the SSVS in its original form).

2.2.3. Avatar Creation Style

Participants selected their preferred avatar creation style from three descriptions based on Neustaedter and Fedorovskaya (2009)—that is, *Realistic*, *Ideal*, and *Different*, respectively:
When you create avatars in games, which of the following statements best describes you (choose only one)?

1. When I create avatars in games, I tend to create them as realistic and similar to myself as possible.
2. When I create avatars in games, I tend to create them as an idealized version of myself.
3. When I create avatars in games, I tend to create them as someone distinctly different from myself.

2.2.4. Procedure

Participants completed the study online in a time and location of their choosing. They first completed the screening questions followed by the three versions of the SSVS (in fixed actual/ideal/ought order) and then indicated their preferred avatar creation style. They subsequently provided basic demographic information, reported how many hours per week they spent playing video games ($M = 9.62$; $SD = 10.35$)[2], and received online debriefing[3].

[1] Neustaedter and Fedorovskaya (2009) suggested that Roleplayers and Fantasies might constitute conceptually distinct subtypes of "Different" avatar creators. Of the 30 Differents in our sample, only six indicated that they created multiple avatars (which is the primary distinguishing feature of Roleplayers). Consequently, we did not attempt analyses intended to differentiate the two Different subtypes in the present study.

[2] One-way analyses of variance (ANOVAs) revealed no significant relationship between avatar creation style and either months of self-reported experience with the specified MMOG or self-reported hours of video game play per week (both $Fs < 1$), so neither time variable will be discussed further.

[3] As part of a secondary research question, participants also completed a 10-item measure, based on Livingston et al. (2014), of the economic (utilitarian) values that players ascribe to their avatars. The measure was completed once for each avatar participants indicated having created or customized. Results associated with this measure will not be discussed here.

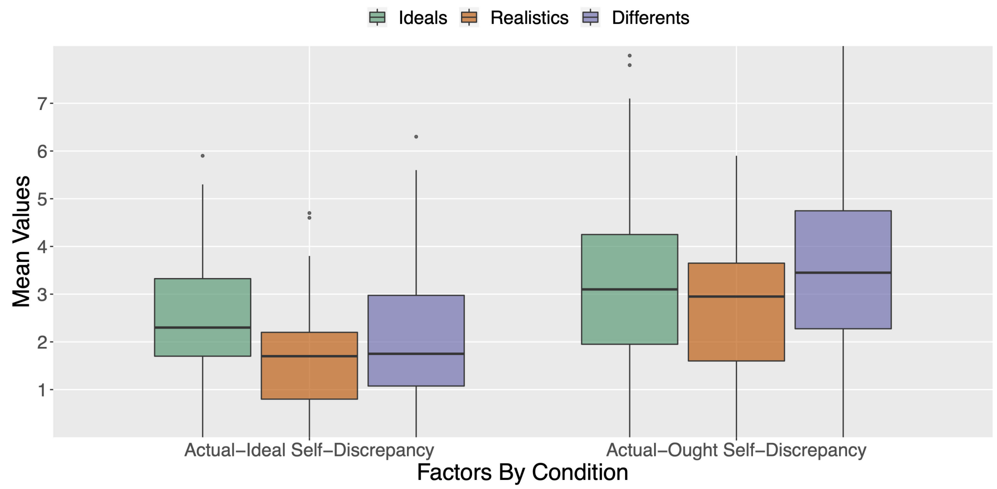


FIGURE 1 | Comparison of Realistics, Ideals, and Differents for their actual-ideal and actual-ought self-discrepancy scores.

3. RESULTS

The avatar creation style breakdown in the present sample was 27 Realistics, 68 Ideals, and 30 Differents. Preliminary two-way analyses of variance (ANOVAs) revealed that neither the gender main effect nor the gender × avatar creation style interaction was significant for either the actual-ideal or the actual-ought self-discrepancy scores (all $p_s > 0.10$), so gender will not be discussed further[4]. Thus, for hypothesis testing purposes, we conducted separate one-way analyses of variance (ANOVAs) to test for possible links between avatar creation style and actual-ideal and actual-ought self-discrepancies. Given our focus on composite scales rather than single Likert-type items, and given that homogeneity of variance tests yielded non-significant results for both the actual-ideal ($p = 0.371$) and actual-ought discrepancy scores ($p = 0.441$), use of the F-statistic is defensible (see also Carifio and Perla, 2008).

There was a significant main effect of avatar creation style on actual-ideal self-discrepancy scores, $F_{(2,122)} = 3.36, p = 0.038, \eta_p^2 = 0.052$. *Post-hoc* pairwise comparisons (least significant difference procedure) revealed the expected pattern (see **Figure 1**): Ideals reported significantly ($p = 0.011$) higher actual-ideal self-discrepancy scores ($M = 2.57; SD = 1.81$) compared to Realistics ($M = 1.81; SD = 1.02$). In contrast, neither Ideals nor Realistics differed significantly from Differents ($M = 2.42; SD = 1.37; p = 0.590$ and $p = 0.080$, respectively).

There was also a significant main effect of avatar creation style on actual-ought self-discrepancy scores, $F_{(2,122)} = 3.15, p = 0.046, \eta_p^2 = 0.049$. *Post-hoc* pairwise comparisons revealed the expected pattern: Differents reported significantly ($p = 0.047$) higher actual-ought self-discrepancy scores ($M = 4.30; SD = 2.18$) compared to Realistics ($M = 3.31; SD = 1.85$). Actual-ought self-discrepancy scores were also significantly ($p = 0.018$) higher for Differents than for Ideals ($M = 3.32; SD = 1.72$). In contrast, actual-ought self-discrepancy scores for Realistics and Ideals did not differ significantly ($p = 0.972$).

[4]For more poignant discussions of constraints on gender expression in MMOGs, see McArthur et al. (2015), McArthur (2017), and Pace et al. (2009).

4. DISCUSSION

Guided by self-discrepancy theory (SDT), we conducted the present research to gain a better understanding of how and why video game players select or create the avatars that they do. We reasoned that players' preferred style of avatar creation could be linked to the magnitude of self-perceived discrepancies between who they are, who they aspire to be, and who they think they should be. To test this idea, MMOG players indicate their preferred avatar creation style and completed a values measure from three different perspectives. Methodologically speaking, our approach differed from previous discrepancy-based avatar research in at least three ways. First, our participants completed all three discrepancy-related measures from a self-perspective (rather than one or more from the perspective of an avatar). Second, an ought self-measure was included among these three. Third, our computational approach focused solely on the magnitude of self-discrepancies, not their direction (vs., e.g., Mancini and Sibilla, 2017).

As hypothesized, the perceived values discrepancy between actual and ideal self averaged higher among those who preferred idealized avatars compared to those who preferred realistic avatars; those who preferred completely different avatars (i.e., fantasy/role-players) averaged in between. Also as hypothesized, the perceived values discrepancy between actual and ought self averaged higher among those who preferred different avatars compared to those who preferred realistic avatars; the actual-ought discrepancy among fantasy/role-players (i.e., different avatars) also averaged higher compared to those who preferred idealized avatars.

These results contribute to the existing empirical literature concerning the extent to which avatars serve a compensatory function among those who perceive gaps between who they are and who they want to be or (think they) should be. That is, whereas idealized avatars embody aspirations, wholly different avatars seem to reflect a casting off of perceived demands within the relative safety of a virtual game world's "Magic Circle" (where players will conform to how they represent themselves digitally, manifesting in deviant or aspirational behavior in line with the Proteus effect as discussed by Yee and Bailenson, 2007). In contrast, true-to-self (realistic) avatars tend to be preferred by those who perceive their various selves to be in comparative alignment.

Understanding avatar creation style from an SDT perspective that includes the ought self as well as the ideal self opens up intriguing avenues for subsequent research. Indeed, in its original formulation, SDT was intended to help understand emotions, with actual-ideal discrepancy mapping onto depressive affect and actual-ought discrepancy mapping onto anxiety (Higgins, 1987). It might therefore be worthwhile to explore the potential for avatar creation style to serve not only as a proxy indicator of psychological well-being, but also as a clue concerning the domain(s) in which adjustment difficulties may lie. For example, a strong preference for "different" avatars might suggest that an individual is struggling with one or more identity elements in their real life (e.g., sexual orientation, religious disillusionment) that may be subject to censure in their social environment. Extending this reasoning, shifts in preferred avatar creation style over time could be of diagnostic value.

The psychosocial consequences of discrepancy-congruent or discrepancy-incongruent gameplay should also be explored. For example, do individuals with a substantial actual-ideal discrepancy feel better after playing as an idealized avatar? Would individuals who lack substantive self-discrepancies feel disoriented after playing as a fantasy/role-play avatar? How enduring are such effects?

The adoption of specific avatar creation styles could have therapeutic value. Thus, in line with research suggesting positive effects of feeling self-present in a game world (Fox and Bailenson, 2009; Jin, 2011; Behm-Morawitz, 2013), game developers could actively promote diverse avatar creation styles based on the needs of their community. Given the volatility of online communities, this could be a helpful tool for game community managers seeking to improve the collective well-being and/or possibly reduce the toxicity of online gaming communities. In this sense, knowledge of avatar creation styles provides a missing link between the self-expressive world of virtual characters and the real-life interaction displayed in out-of-game community behaviors.

With respect to limitations, the present study sampled only Canadian psychology undergraduates, but gender and ethnic diversity was considerable (40% women, 62% non-Euro-Canadian). The avatar creation style instrument created for this study relied on a single-item, forced-choice format, which may have sacrificed some information that multi-item, continuous measures could have provided. For example, a basic distinction between an avatar's appearance and the avatar's in-game behavior could prove important. Moreover, although avatar style preferences are at least somewhat stable within individual players (see Mancini and Sibilla, 2017), contextual factors can also shape specific choices (e.g., Triberti et al., 2017). Discrepancy scores in the present study were generated based on a novel administration of a values scale, although the values dimensions assessed have been among the most comprehensive and cross-culturally validated constructs in psychology (see Lindeman and Verkasalo, 2005).

Notwithstanding the present study's limitations, our results demonstrate SDT's usefulness with respect to understanding the link between players and their avatars. Indeed, our results suggest that game designers would do well to ensure that players have the tools to fashion avatars that feel "right," for avatar creation appears to be driven—at least in part—by the oughts and ideals that the players carry within them.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Psychology Delegated Ethics Review Committee (DERC), ORE #: 21832. The patients/participants provided their electronic informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

ML designed the studies, performed the statistical analysis together with CB, interpreted the study results together with CB and LN, and wrote the first full draft of the manuscript together with CB. LN co-supervised ML, edited the drafts, and provided the feedback on the project and on the manuscript as well as rewriting portions of it from the last draft to this submission. All authors contributed to the article and approved the submitted version.

FUNDING

This work was made possible by Lennart Nacke's NSERC Discovery Grant 2018-06576, the Canada Foundation for Innovation Infrastructure Fund 35819 SURGE—The Stratford User Research and Gameful Experiences Lab, the NSERC CREATE SWaGUR grant, and the Social Sciences and Humanities Research Council Canada Grant 895-2011-1014 (IMMERSe).

ACKNOWLEDGMENTS

We thank all participants in the study and The Games Institute at the University of Waterloo for their support. We also thank Katja Rogers for lending us support with her R and boxplot creation skills.

REFERENCES

Bailenson, J., and Blascovich, J. (2004). "Avatars," in *Encyclopedia of Human-Computer Interaction*, ed Bainbridge, W. S. (Great Barrington, MA: Berkshire Publishing Group), 64–68.

Banakou, D., Hanumanthu, P. D., and Slater, M. (2016). Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias. *Front. Hum. Neurosci*. 10:601. doi: 10.3389/fnhum.2016.00601

Behm-Morawitz, E. (2013). Mirrored selves: the influence of self-presence in a virtual world on health, appearance, and well-being. *Comput. Hum. Behav*. 29, 119–128. doi: 10.1016/j.chb.2012.07.023

Bessière, K., Seay, A. F., and Kiesler, S. (2007). The ideal elf: identity exploration in world of warcraft. *Cyberpsychol. Behav*. 10, 530–535. doi: 10.1089/cpb.2007.9994

Biocca, F. (1997). The Cyborg's dilemma: progressive embodiment in virtual environments. *J. Comput. Mediat. Commun*. 3. doi: 10.1111/j.1083-6101.1997.tb00070.x

Carifio, J., and Perla, R. (2008). Resolving the 50-year debate around using and misusing likert scales. *Med. Educ*. 42, 1150–1152. doi: 10.1111/j.1365-2923.2008.03172.x

Cohen, J. (2001). Defining identification: a theoretical look at the identification of audiences with media characters. *Mass Commun. Soc*. 4, 245–264. doi: 10.1207/S15327825MCS0403_01

Crenshaw, N., and Nardi, B. (2014). "What's in a name? Naming practices in online video games," in *Proceedings of the First ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play, CHI PLAY '14* (New York, NY: Association for Computing Machinery), 67–76. doi: 10.1145/2658537.2658685

Dunn, R. A., and Guadagno, R. E. (2012). My avatar and me–gender and personality predictors of avatar-self discrepancy. *Comput. Hum. Behav*. 28, 97–106. doi: 10.1016/j.chb.2011.08.015

Fox, J., and Bailenson, J. N. (2009). Virtual self-modeling: the effects of vicarious reinforcement and identification on exercise behaviors. *Media Psychol*. 12, 1–25. doi: 10.1080/15213260802669474

Hefner, D., Klimmt, C., and Vorderer, P. (2007). "Identification with the player character as determinant of video game enjoyment," in *International Conference on Entertainment Computing* (Springer), 39–48. doi: 10.1007/978-3-540-74873-1_6

Higgins, E. T. (1987). Self-discrepancy: a theory relating self and affect. *Psychol. Rev*. 94:319. doi: 10.1037/0033-295X.94.3.319

Jin, S.-A. A. (2011). "It feels right. therefore, i feel present and enjoy": the effects of regulatory fit and the mediating roles of social presence and self-presence in avatar-based 3D virtual environments. *Presence Teleoper. Virtual Environ*. 20, 105–116. doi: 10.1162/pres_a_00038

Jin, S.-A. A., and Park, N. (2009). Parasocial interaction with my avatar: Effects of interdependent self-construal and the mediating role of self-presence in an avatar-based console game, WII. *Cyberpsychol. Behav*. 12, 723–727. doi: 10.1089/cpb.2008.0289

Klimmt, C., Hefner, D., and Vorderer, P. (2009). The video game experience as "true" identification: a theory of enjoyable alterations of players' self-perception. *Commun. Theory* 19, 351–373. doi: 10.1111/j.1468-2885.2009.01347.x

Klimmt, C., Hefner, D., Vorderer, P., Roth, C., and Blake, C. (2010). Identification with video game characters as automatic shift of self-perceptions. *Media Psychol*. 13, 323–338. doi: 10.1080/15213269.2010.524911

Lindeman, M., and Verkasalo, M. (2005). Measuring values with the short Schwartz's value survey. *J. Pers. Assess*. 85, 170–178. doi: 10.1207/s15327752jpa8502_09

Livingston, I. J., Gutwin, C., Mandryk, R. L., and Birk, M. (2014). "How players value their characters in world of warcraft," in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work* & *Social Computing, CSCW'14* (New York, NY: Association for Computing Machinery), 1333–1343. doi: 10.1145/2531602.2531661

Lomanowska, A. M., and Guitton, M. J. (2014). My avatar is pregnant! Representation of pregnancy, birth, and maternity in a virtual world. *Comput. Hum. Behav*. 31, 322–331. doi: 10.1016/j.chb.2013.10.058

Mancini, T., and Sibilla, F. (2017). Offline personality and avatar customisation. discrepancy profiles and avatar identification in a sample of mmorpg players. *Comput. Hum. Behav*. 69, 275–283. doi: 10.1016/j.chb.2016.12.031

McArthur, V. (2017). "The UX of avatar customization," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17* (New York, NY: Association for Computing Machinery), 5029–5033. doi: 10.1145/3025453.3026020

McArthur, V., Teather, R. J., and Jenson, J. (2015). "The avatar affordances framework: Mapping affordances and design trends in character creation interfaces," in *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play, CHI PLAY '15* (New York, NY: Association for Computing Machinery), 231–240. doi: 10.1145/2793107.2793121

Nath, T. (2016). *Investing in Video Games. This Industry Pulls in More Revenue Than Movies, Music*. Nasdaq.com. Retrieved from: http://www.nasdaq.com/article/investing-in-video-games-this-industry-pulls-in-more-revenue-than-movies-music-cm634585

Neustaedter, C., and Fedorovskaya, E. (2009). "Presenting identity in a virtual world through avatar appearances," in *Proceedings of Graphics Interface 2009, GI '09* (Toronto, ON: Canadian Information Processing Society), 183–190. Available online at: https://dl.acm.org/doi/proceedings/10.5555/1555880

Pace, T., Houssian, A., and McArthur, V. (2009). Are socially exclusive values embedded in the avatar creation interfaces of MMORPGs? *J. Inform. Commun. Ethics Soc*. 7, 192–210. doi: 10.1108/14779960910955909

Roth, D., Lugrin, J.-L., von Mammen, S., and Latoschik, M. E. (2017). "Controllers & inputs: masters of puppets," in *Avatar, Assembled: The Social and Technical Anatomy of Digital Bodies, Chapter 29*, ed J. Banks (New York, NY: Peter Lang), 281–290.

Slater, M., Perez-Marcos, D., Ehrsson, H., and Sanchez-Vives, M. (2009). Inducing illusory ownership of a virtual body. *Front. Neurosci*. 3:29. doi: 10.3389/neuro.01.029.2009

Song, W., and Fox, J. (2016). Playing for love in a romantic video game: avatar identification, parasocial relationships, and chinese women's romantic beliefs. *Mass Commun. Soc*. 19, 197–215. doi: 10.1080/15205436.2015.1077972

Sung, Y., Moon, J., Kang, M., and Lin, J.-S. (2011). Actual self vs. avatar self: the effect of online social situation on self-expression. *J. Virtual Worlds Res*. 4:4–21. doi: 10.4101/jvwr.v4i1.1927

Triberti, S., Durosini, I., Aschieri, F., Villani, D., and Riva, G. (2017). Changing avatars, changing selves? The influence of social and contextual expectations on digital rendition of identity. *Cyberpsychol. Behav. Soc. Netw*. 20, 501–507. doi: 10.1089/cyber.2016.0424

Villani, D., Gatti, E., Triberti, S., Confalonieri, E., and Riva, G. (2016). Exploration of virtual body-representation in adolescence: the role of age and sex in avatar customization. *SpringerPlus* 5:740. doi: 10.1186/s40064-016-2520-y

Waltemate, T., Gall, D., Roth, D., Botsch, M., and Latoschik, M. E. (2018). The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE Trans. Vis. Comput. Graph*. 24, 1643–1652. doi: 10.1109/TVCG.2018.2794629

Yee, N., and Bailenson, J. (2007). The Proteus effect: the effect of transformed self-representation on behavior. *Hum. Commun. Res*. 33, 271–290. doi: 10.1111/j.1468-2958.2007.00299.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX

Short Schwartz's Value Survey

In the following questionnaire, you will be asked about the importance of a range of values from the perspective of your actual, ideal, and ought selves.

Your **"actual self"** refers to **how you truly see yourself**. Your **"ideal self"** refers to **how you would like to be**. Your **"ought self"** refers to **how you think others think you should be**.

Each of the next three pages of the questionnaire will focus on one of these selves.

1. Please rate the importance of the following values as guiding principles from the perspective of how you truly see yourself (that is, your "actual self").
2. Now rate the importance of the same values from the perspective of how you would like to be (that is, your "ideal self").
3. Finally, please rate these values once more from the perspective of how others think you should be (that is, your "ought self").

(−2 = Opposed to my values, 0 = Not Important, 6 = Important, 12 = Very Important, 14 = Of Supreme Importance) POWER (social power, authority, wealth)

ACHIEVEMENT (success, capability, ambition, influence on people and events)

HEDONISM (gratification of desires, enjoyment in life, self-indulgence)

STIMULATION (daring, a varied and challenging life, an exciting life)

SELF-DIRECTION (creativity, freedom, curiosity, independence, choosing one's own goals)

UNIVERSALISM (broad-mindedness, beauty of nature and arts, social justice, a world at peace, equality, wisdom, unity with nature, environmental protection)

BENEVOLENCE (helpfulness, honesty, forgiveness, loyalty, responsibility)

TRADITION (respect for tradition, humbleness, accepting one's portion in life, devotion, modesty)

CONFORMITY (obedience, honoring parents and elders, self-discipline, politeness)

SECURITY (national security, family security, social order, cleanliness, reciprocation of favors).

frontiers
in Computer Science

ORIGINAL RESEARCH
published: 16 April 2021
doi: 10.3389/fcomp.2021.531713

How the Visual Design of Video Game Antagonists Affects Perception of Morality

Reyhan Pradantyo[1], *Max V. Birk*[2] *and Scott Bateman*[1*]

[1] Human-Computer Interaction Lab, Faculty of Computer Science, University of New Brunswick, Fredericton, NB, Canada,
[2] Department of Industrial Design, Eindhoven University of Technology, Eindhoven, Netherlands

The visual design of antagonists—typically thought of as "bad guys"—is crucial for game design. Antagonists are key to providing the backdrop to a game's setting and motivating a player's actions. The visual representation of antagonists is important because it affects player expectations about the character's personality and potential actions. Particularly important is how players perceive an antagonist's morality. For example, an antagonist appearing disloyal might foreshadow betrayal; a character who looks cruel suggests that tough fights are ahead; or, a player might be surprised when a friendly looking character attacks them. Today, the art of designing character morality is informed by archetypal elements, existing characters, and the artist's own background. However, little work has provided insight into how an antagonist's appearance can lead players to make moral judgments. Using Mechanical Turk, we collected participant ratings on a stimulus image set of 105 antagonists from popular video games. The results of our work provide insights into how the visual attributes of antagonists can influence judgments of character morality. Our findings provide a valuable new lens for understanding and deepening an important aspect of game design. Our results can be used to help ensure that a particular character design has the best chance to be universally seen as "evil," or to help create more complex and conflicted emotional experiences through carefully designed characters that do not appear to be bad. Our research extends current research practices that seek to build an understanding of game design and provides exciting new directions for exploring how design and aesthetic practices can be better studied and supported.

Keywords: video games, morality, visual design, empirical methods, antagonists, bad guys, character design, visual attributes

OPEN ACCESS

Edited by:
Z. O. Toups,
New Mexico State University,
United States

Reviewed by:
Jaime Banks,
Texas Tech University, United States
Samuel Marcos-Pablos,
University of Salamanca, Spain

Correspondence:
Scott Bateman
scottb@unb.ca

Specialty section:
This article was submitted to
Human-Media Interaction,
a section of the journal
Frontiers in Computer Science

Received: *31 January 2020*
Accepted: *25 February 2021*
Published: *16 April 2021*

Citation:
Pradantyo R, Birk MV and Bateman S (2021) How the Visual Design of Video Game Antagonists Affects Perception of Morality.
Front. Comput. Sci. 3:531713.
doi: 10.3389/fcomp.2021.531713

INTRODUCTION

Antagonists—who are often thought of as "bad guys"—are a critical part of game design. Antagonists often drive the story of a game, by acting as a catalyst for conflict, thereby influencing player choices and providing important challenges (e.g., a "boss fight"; Vorderer et al., 2003; Schell, 2008; Przybylski et al., 2010). Villains that inspire and challenge the player will keep them from losing interest in the goal (Manninen and Kujanpää, 2007). While character design often involves creating a backstory and defining behaviors and abilities, one key way that characters are initially experienced by players is through their visual attributes—what they look like (Bar et al., 2006).

We draw from-visual stereotypes to predict character attributes and behaviors. Our presumptions are based on common references and traits. First impressions of a new character have been shown to be persistent even if impressions are contradicted or more nuanced information are revealed (Haake and Gulz, 2008). Visual attributes help game designers to communicate elements of a game's theme, story, and challenge, and steer player behavior (Baranowski et al., 2008; Schell, 2008; Przybylski et al., 2010; Bakkes et al., 2012; Mohd Tuah et al., 2017). For example, the slanting eyebrows of the Goombas in *Super Mario Bros*. Nintendo 1983 help convey to the player that they are not happy, and that the player should get out of the way. Important to understand how people perceive the intent and actions of characters is *morality*, which is the "… differentiation of intentions, decisions and actions between those that are distinguished as proper and those that are improper" (Long et al., 1987). People use characters' appearances to help make judgments about their morality, and morality perception greatly affect people's enjoyment of games and other media (Eden et al., 2015). Thus, game designers and artists try to match designs with how they want a player to interpret their characters, whether it is in a congruent way (e.g., a bad guy who looks evil) or an incongruent way (e.g., a bad guy who looks innocent, non-threatening and friendly). Providing information about the perception of higher level personality traits with the visual attributes of antagonists could be extremely useful for game designers (McLaughlin, 2012).

Because the visual design of characters is a critical part of game design, we provide a first study examining how different visual attributes lead to different moral interpretations of antagonists. To do this, we first created a stimulus set of 105 images of antagonists that span a wide range of successful games from the last 5 years. Next, we conducted a survey on Mechanical Turk ($N = 283$), in two parts. Part 1 solicited rating of antagonist images using the CMFQ-S (Character Moral Foundations Questionnaire–Short, a short, validated scale previously used in the interpretation of character morality). In part 2, we gathered people's judgments on the saliency of visual attributes that featured "prominently" in the design of antagonists. By assembling our two data sets, we are able to provide evidence-based insights into many of the important visual attributes used in the design of video game antagonists and relate them directly to judgments of morality.

The findings of our work provide valuable new insights and deepen our understanding of how character design—an important aspect of game design—is interpreted. In practice, our results can be used to help ensure that a particular character design has the best chance to be universally seen as "evil," or to help create more complex and conflicted emotional experiences through carefully designed characters that do not appear to be bad. More broadly, we believe the methodology that we have identified can be leveraged, extended and strengthened to improve game design research, to better characterize current practices and cultural experience, and to build more precise and engaging entertainment experiences.

LITERATURE REVIEW

Antagonists in Video Games

The tension between the main character (i.e., protagonist) of a story and their opponent (i.e., antagonist) is ubiquitous to fiction and fuel many dramatic situations; e.g., when Darth Vader reveals himself to be Luke Skywalker's father. Looking at the common template of the monomyth or the hero's journey (Lane, 2017), the antagonist is the cause of going on a journey, and provides the reason that challenges need to be faced, while providing temptations on the way (e.g., Vader to Luke: "Join me and together we can rule the galaxy as father and son." *The Empire Strikes Back,* Lucas Films 1982). In video games, antagonists fulfill a similar role—in the *Super Mario Bros.* series (Nintendo 1985) Bowser keeps Princess Peach hostage so that Mario can set out to rescue her; in the *Sonic the Hedgehog* series (Sega 1989) Dr. Robotnik/Eggman aims to achieve world domination which Sonic tries to prevent.

Villains are central to every culture, because they provide a moral compass (Eden et al., 2015)—they show behaviors that are threatening to society, because they cause others physical harm, deny the rights and freedom of others, create chaos, would betray others, or perform actions that are disgusting. As such, villains are on the opposite line of moral behavior, which helps us relate to the hero's efforts and understand their drive.

Visual Attributes of Villains and Archetypes

It has been well-established that there are clear differences between how heroes and villains are visually represented and that this affects people's judgments about these characters (Hoffner and Cantor, 1991; Eden et al., 2015; Grizzard et al., 2018). Narratives often use tropes or clichés that the audiences are familiar with—such as the damsel in distress trope used in Mario—to set expectations and to make clear what actions will need to be taken. Similar to narrative tropes, character designers use visual archetypes (Haake and Gulz, 2008; e.g., the muscle packed action hero; or the magician in long robes) to provide visual affordances for players (e.g., recognizing a character that will likely use brute force vs. magic) to motivate player actions. For example, when facing a life or death decision we would act differently toward an immoral character (e.g., someone who puts themselves at risk to help an injured child in a dire situation vs. someone who always acts in their own best interests); or, someone who betrays their team or family, compared to someone who has displayed moral behavior (e.g., acting with loyalty even while being tempted toward disloyalty).

Classic villains such as the gangster wearing a fedora, a striped suit, and two-tone shoes, or the long-nosed witch with a tall hat, and crooked teeth, are well-known and easily identified. Literature and drama are the source of villain archetypes (Fahraeus and Yakali Çamoglu, 2011), but archetypes are present in games as well, e.g., the mentally unstable villain Joker in *Batman: Arkham Asylum* (Rocksteady 2009), or the superior species like the Sectoids in *XCOM 2: Enemy Unknown* (Fireaxis Games 2016). While the presented archetypes could be applied across genders, age groups, and race, the majority of villains are male (Ivory, 2006), with an observable uptake in female villains (Lindner et al., 2019). In the design of villain character designers show several preferences, such as Classic Villain—TV Tropes (n.d.):

- **Display a common vice:** antagonists often represent a sin or a vice; e.g., wrath, gluttony, pride; e.g., *God of War*'s Baldur represents wrath through his visual display of anger and rage (Sony Interactive Entertainment 2018);

- **Display a common moral flaw:** many villains have at least one moral flaw; e.g., they are deeply disloyal or careless toward the wellbeing of others; e.g., the ruthless Handsome Jack in *Borderlands 2* (2K Games, 2012);
- **Distinct Color:** antagonists are visually distinct through the use of color (Lundwall, 2017); e.g., antagonists are often represented with dark color palettes, while protagonists are bright; e.g., Link's primary color is green, the color of hope, wielding the bright Master Sword, while the evil Ganon is black and red (*The Legend of Zelda: Breath of the Wild*, Nintendo, 2017); and,
- **Distinct Form:** antagonists are visually distinct through the use of shape (Ekström, 2013); e.g., the spiky Sauron (*Middle-earth: Shadow of War*, Warner Bros. Interactive Entertainment 2017); or, size, e.g., the oversized Onyxia in *World of Warcraft* (Blizzard Entertainment, 2004).

While designers often draw from their experience, mood boards, and previous characters with similar traits, media psychology and communication studies provide theoretical frameworks to characterize the effects and implications of designing for moral judgments. So how exactly are moral judgments formed?

Appearance and Moral Foundations

As a member of any society, we learn what is right and what is wrong, and we learn to associate certain forms of appearance with morally questionable behavior (Klapp, 1954). Motorcycle gangs like the Hell's Angles, for example, wear vests with patches—identifying them as gang members—and are associated with violence. Or, the slick look of a wall street banker that suggests a dedication to personal gain, the willingness to put personal gain before others, and to cause chaos and disorder through immoral actions like morally questionable stock trades. Media commonly draws from imagery that is reminiscent of morally corrupt parts of a historical or current society to make it easy for the audience to identify the moral stance of a character (Klapper, 1960).

Several theories from different fields [e.g., psychology (Kohlberg, 1971; Diessner et al., 2008; Doris et al., 2020), philosophy (Haidt and Joseph, 2008), sociology (Boltanski and Thévenot, 2000), law (Raz, 1995), and communication studies (Fiske et al., 2007; Eden et al., 2015)], provide a nuanced perspectives on how morality might be communicated through a person's appearance (Haidt and Joseph, 2008), for example, present the moral foundation theory (MFT). Moral foundation theory offers a pluralistic perspective on moral, suggesting that morality is judged on in five domains: harm/care, fairness/reciprocity, ingroup/loyalty, authority/respect, purity/sanctity.

Building on Haidt and Joseph's theory, Grizzard et al. (2019) evaluated and extended the character morality questionnaire. Their questionnaire asks participants to indicate their agreement to questions such as "This character would physically hurt another person."

Linking appearance with emotional responses, people are capable of making split-second judgments of others (Willis and Todorov, 2006). It is suggested that both men and women are influenced by physiognomy in day-to-day life. In this study, participants rated faces based on attractiveness, likeability, competence, trustworthiness, and aggressiveness with insignificantly no difference between being with or without time constraints. This spontaneous detection skill is suggested to be essential for survival. These papers have also studied how and why people perceive stereotypes of good guys and bad guys (Secord et al., 1953; Bull and Green, 1980; Goldstein et al., 1984; Yarmey, 1993; Flowe, 2012; Croley et al., 2017), and discuss moral perceptions based on physiognomy and other visual attributes. To measure morality, the Character Moral Foundations Questionnaire (CMFQ) (Eden et al., 2015; Grizzard et al., 2019) was often used. In all cases, visual attributes are capable of affecting the peoples' moral judgments of characters.

Studies have shown that the perceived morality among heroes and villains (Eden et al., 2015) in media have strong connections with viewer's enjoyment (Sanders and Tsay-Vogel, 2016; Eden et al., 2017). A well-known theory in understanding the ties between media enjoyment and morality is affective disposition theory (ADT) (Raney, 2006). Affective disposition theory suggests that viewers interpret characters as liked or disliked based on how they judge the character's morality. The outcome of any event affects the viewer's enjoyment, depending on the congruence of the viewer's expectations: highly liked characters who experience positive outcomes and less liked characters who experience negative outcomes increase viewer enjoyment (Raney, 2004). Consequently, enjoyment decreases when unexpected events occur, such as when a liked character experiences a negative outcome or a disliked character experiences a positive outcome.

Moral Judgments

The interaction between media and entertainment use, media experience, and moral judgment, has is at the center of ADT (Zillmann, 1996). Affective disposition theory engages with how viewers perceive and assess a character based on their actions and determine if a character is good or bad. From the viewer perspective defining a character as good or bad creates tension and suspense, because depending on the characters moral leaning the audience tries to predict future action and observes if a character acts according to the ascribed moral category. The perceived disposition affects the audience's enjoyment of a narrative. The game the Last of Us 2, plays with character expectations. The player starts out playing the character Abby Anderson. Playing from Abby's view the player first likes Abby. Deeper into the narrative Abby commits violent actions against characters that were established as "good," which leads to Abby being depicted as "bad" due to her actions. Abby being depicted as a "bad" character creates expectations regarding Abby's future actions and conflicts for the player when they need to take on Abby's role playing now a "bad" character themselves. Applied to games ADT would suggest that the expectations regarding the disposition of the character is foundational for the player's affective response; e.g., the feeling of disgust or despair when a "bad" character falls in carnage or kills a good character or the positive feeling when the hero prevails and experience success (Raney, 2004) offers two complementing amendments to ADT: (1) the formation of an affective disposition sometimes precedes the moral evaluation of a character (for evidence see Grizzard et al., 2018), and (2) the ascribed disposition "good" or "bad" leads to an interpretation of a character's actions in line with expectations. Both concepts are interesting our research, because (1) suggests that the simple interpretation of a character's appearance affects moral decision and (2) that the interpretation potentially influences how further actions of such a character are perceived.

The theory has practical value for our research, because it argues that the dispositions we ascribe to a character may be relevant for our entertainment experience and might lead to emotional experiences. Hence, judging a character as "bad" and differentiate potential future actions based on the character's appearance, e.g., a character judged as impure who commits violent actions, might be exactly the form of tension a game-designer aims for. The tyrant Pagan Min in *Far Cry 4* (Ubisoft), who is introduced to the player by killing one of his commanders using a pen, might create an expectation of unpredictable violence and could facilitate emotional experiences for the player; e.g., fear of Min's unpredictable actions.

To advance media theories and ADT—which focused on short-term affective engagement with media—and provide a more wholistic perspective on morality and media effects on society, Tamborini (2011) suggested the Model of Intuitive Morality and Exemplars (MIME). Model of intuitive morality and exemplars suggests that strong moral beliefs are uphold by media selection, i.e., we like content that fits within our overall moral belief system and is therefore more likely to be selected, and reinforces our moral belief system. Build on Moral Foundation Theory (Haidt and Joseph, 2008), MIME follows a dual-processing logic and suggests that we evaluate events intuitively (process 1) unless they are not within expectations or too complex, then we are deliberately rational to comprehend the given events (process 2). Model of intuitive morality and exemplars also draws from exemplification theory (Zillmann, 1999), which suggests that recent or frequent events or concrete and highly emotional exemplars increase moral judgment. From a game-designers perspective MIME might explain preferences and playstyle of a player by considering previous media preference and moral examples within these games. Game designers can make use of character expectations and effects of creating unexpected scenarios, e.g., a morally corrupt character helping a vulnerable protagonist.

Model of intuitive morality and exemplars has been applied in studies to analyze or discuss the effect of videogames (Tamborini et al., 2011, 2017; Eden et al., 2014) and game characters (Joeckel et al., 2012; Tamborini et al., 2013; Boyan et al., 2015) and provides guidance to understand morality processes during game play and a framework to understand how videogames shape audience's moral intuitions, and subsequently media interpretation and response. Looking at the morality of the characters we present, MIME supports that our interpretation and moral judgment is a result of the media context and its effect on our moral intuition. The gangster world of Martin Scorsese, contributed to our interpretation of Italian man in needle striped suits with a fedora as gangsters, and informed our moral intuition to interpret video game character in games such as *Grant Theft Auto* (Rockstar Games 1997) or *Mafia* (Illusion Softworks 2002) that dress the same way as similarly morally corrupt.

Morality in Video Games

In comparison to other media forms, video games put the player into the driver's seat, resulting in a context where moral actions are not just observed, but actively executed (DeVane and Squire, 2008). Video games enable players to explore their moral values through the protagonist, by making moral decisions of any kind themselves and act in environments where moral values are deviate from the values of modern society.

In video games, morality and its different dimensions set players expectations—for example, the criminal setting of *Grand Theft Auto* (Rockstar Games 1997) puts the player in the role of a criminal in a fictional city. The mechanics and rules of the game reinforce morally questionable behavior such as beating up people, stealing cars, or destroying property. However, even in a criminal world not all moral dimensions are abolished: e.g., loyalty toward gang comrades remains relevant. Morality in such games such has been intensively studied and has triggered heated public and academic debates about transfer effects of violence (Ferguson, 2008).

In games where the player has a choice about the moral compass of their character, we usually find indicators of their standing in society represented by the people they can talk to (*Mass Effect*; BioWare 2007), the availability of dialog (*Detroit Become Human*; Quantic Dream 2018), or visual indicators—the classic game *Ultima Online* (Origin Systems 1997), for example, assigned a special name tag to individuals who attacked other players.

The antagonist and the moral beliefs they project have implication for the presumptions of the player, and subsequently their intuition about in-game situations (Joeckel et al., 2012). For example, when interacting with an antagonist that has not acted fairly, the player would mistrust their offers. The narrative-driven zombie game series, *The Walking Dead* (Telltale Games 2012–2019), frequently presents players with situations where they need to judge the moral compass of the players around them; e.g., when offered food from a group that might or might not have engaged in cannibalism.

From a designer's perspective, the visual attributes (e.g., an eye patch or a scar), that inform players about the morality of a character are important to effectively communicate a character's moral standing. This is particularly difficult, considering that a character's moral is not just judged on a single axis from good to bad, but on visual elements that speak to their fairness, willingness to physically hurt others, their loyalty, how willing they are to follow rules, and their ability to engage in disgusting behavior. But how do we approach the complex effects that small details like the nose on how a character is interpreted?

Quantifying Visual Experiences

Jacobsen (2006) outlines how aesthetics can be captured by applying scientific method: by manipulating size and shape of body parts, e.g., waist-to-hip ratios, or evaluating the effect of abstract patterns in comparison to known stimuli. A model to understand aesthetic experiences has been presented by Leder et al. (2004). The authors suggest that aesthetic experiences are context dependent and show that aesthetic experiences are a complex interaction between cognition, affect, and perceptual processes—e.g., judging visual complexity, or relating an experience to prior memories are different cognitive processes.

Researchers and designers have applied several techniques to understand the interpretation of design. Hassenzahl (2004), for example, has investigated the consistency of beauty judgments, and operationalizes beauty. Reinecke et al. (2013) investigated the appeal of websites, in terms of visual complexity, colorfulness, and appeal. The created model of visual appeal combined with basic demographics explained about 50% of resulting appeal ratings. These are similar to Tuch et al.'s (2012) findings, which show that prototypes and visual complexity affect the aesthetic perception of a website, but that the amount of time that a website is viewed matters.

Research has also demonstrated that the visual presentation of interactive products affects our judgment and experience of them (Hassenzahl, 2004). Games, however, more often combine the interactivity of digital products such as apps and websites, with the narrative depth of fiction and drama, creating unique demands on the visual design of video game characters.

Studies Visual Attributes of Characters in Video Games

In the context of our work, we focus primarily on the visual attributes of characters. Providing a deeper understanding of how people interpret game characters is relevant important because identifying with a representation increases the amount of time a game is played (Passmore et al., 2018), how deeply players comprehend information (Kao and Harrell, 2015), and overall engagement (Reinecke, 2009).

Previous work has established that we interpret characters values using a number of visual attributes, for example, character shape (Veronica, 2015), age (Schwind and Henze, 2018), gender (Schwind and Henze, 2018), and fashion (Klastrup and Tosca, 2009). Importantly, based on visual attributes we draw conclusions about characters' moral beliefs (Happ et al., 2013). Further, how we see a character affects in-game behavior. We tend to act in a way that we believe confirms a character's beliefs. For example, beating up prostitutes in *Grand Theft Auto* is not necessary, but people still do so, because it is in-line with the moral value system presented in the game (Happ et al., 2013).

There has been some work that has tried to tease apart visual properties of characters and how they are perceived. Schwind and Henze (2018) investigated gender and age differences in virtual faces, finding that in a character designing task that participants create villain faces as more masculine, unattractive, and with lower likeability. Villain faces have also been shown to be more related to features such as looking dead or zombie-like (Schwind et al., 2015).

To move toward a comprehensive understanding of the relationship visual attributes and experience, we need to investigate specific visual elements of character design systematically and empirically.

STUDY DESIGN

To provide an initial understanding about how the visual attributes of game antagonists can influence how people experience them, we carried out two studies. Our studies asked people to separately rate a stimulus set of 105 antagonist images, identifying their most salient visual attributes and to judge the characters' morality. Our two studies were conducted using the crowdsourcing platform Amazon Mechanical Turk (MTurk), using the same set of antagonists' images, but differing in the requested assessment of the images. In Study 1, participants rated the morality of each character based on their appearance using the five morality dimensions (Haidt and Joseph, 2008): "harm/care," "fairness/reciprocity," "ingroup/loyalty," "authority/respect," and "purity/sanctity." In Study 2, participants rated the prominence of character visual features (e.g., eyebrows, age, dermatological problems), defined as characteristics that "… relative to other characteristics, stand out and grab attention." Previous work has investigated the salience of eyes on fixation times (Birmingham et al., 2009), the role of eyebrows in face recognition (Sadr et al., 2003), and shown that we form opinions about a face within 100 ms (Willis and Todorov, 2006)—while prompting individuals to rate relative salience is conceptually fuzzy, it enabled us to guide attention and gauge participants subjective perception of a character. Our analysis connects these two different rating sets, using regression analysis and correlational analysis with the goal to build an understanding of correlations between visual attributes and how they might affect players' experiences of characters.

General Procedure

Both studies followed the same general procedure. Participants were recruited using the crowdsourcing platform Amazon Mechanical Turk. MTurk is a digital platform that acts as a broker between requesters (e.g., researchers looking for participants to rate character images) and workers (e.g., people willing to engage in a rating task for payment).

Our study procedure was reviewed by the Research Ethics Board of the University of New Brunswick and is on file as REB 2019-118. Before being asked to indicate their consent, participants were informed about the procedure, their payment, and the approximate time the task will take. To assure quality data, MTurk participants needed to be US-based and have successfully completed at least 500 tasks with an approval rating of at least 90%. Restricting eligibility combined with attentiveness measures (Study 1) and the screening of completion time reduced the likelihood of bot produced data in our data set, which is an increasing issue in crowdsourced research (Ahler et al., 2019).

Upon qualifying and accepting the MTurk task, participants accessed a website that guided them through the study. Participants were first presented an informed consent form, followed by a demographic questionnaire, and then given instructions on how to complete the rating task. Compensation was calculated at the rate of $7.50 USD/h, to be just above the national minimum wage in the US. We determined 35 min for Study 1 ($5 USD) and 70 min for Study 2 ($8 USD). Since rating 105 characters was a relatively long task, participants only needed to rate five characters to qualify for payment but could opt-in to rate more.

We conducted two separate studies that solicited moral judgments (Study 1) and visual attributes (Study 2) separately, because we anticipated sequence effects from asking moral judgments and visual attributes together. Further, creating two tasks greatly simplified each task and reduced the length of time to perform ratings on any individual character for MTurk participants.

Selecting and Presenting Antagonists

We selected a total of 105 antagonists using publicly available game ranking data using a pre-determined procedure. We determined four main criteria to select our character image repository. The repository should include images of characters that (1) are humanoid; (2) represent recent trends in game and character design; (3) are well-designed; and (4) represent the main antagonists of the games in which they appear. Our intent was to focus on carefully considered, well-designed characters that also represent many of the common visual attributes in their design. Further, we decided to focus on humanoid characters to ensure that the visual attributes that we asked about were present in the character, which assured that our insights are derived from a source that has found mainstream acceptance, covering a wide range of antagonists that have a presence in current games.

To identify antagonist characters, we first had to identify individual games that fit the criteria. To ensure that character designs were both of high quality and represented recent practices, we filtered the database of Gamerankings.com—a website that collects ratings from numerous sources to provide an average score. We selected games released from 2014 to 2019, with a rating above 80% from at least 20 reviews. These criteria allowed us to identify a set of games that met the criteria above, since it captured games, and, therefore, were most likely to contain characters, that were widely seen as "well-designed." Importantly, however, since changes to our criteria or the game database used could result in a different stimulus set, our stimulus set is likely not representative of all games. The resulting initial candidate game list featured 105 games that we filtered further based on the character specific criteria.

For each qualifying game, we identified the main antagonist or final boss using Fandom pages as our main source (https://www.fandom.com). Fandom provides background information and character images for many recent popular games, and all of the characters in our image set. After investigating each game individually, we removed games with non-humanoid antagonists and games without a clear antagonist (e.g., sports games usually do not feature an antagonist created by a game designer) from our initial pool. Our final list of suitable characters featured the main antagonists from 105 games.

Presenting Antagonists

To standardize our stimuli, we created composite images that include a body shot of the character and a close-up of their face. We know from previous work (Schwind et al., 2015; Schwind and Henze, 2018) that the face plays an important role in judging characters and needs to be fully visible for accurate judgments to be made. We removed any background from the images and placed the character on a plain gray background measuring 800 × 570 pixels. See **Figure 1**.

Our stimuli were then presented using a custom-built web application[1].

Figure 1 shows the presentation screen for Study 1 and Study 2, respectively, which were composed of the following five main elements. (1) Each screen displayed breadcrumbs to provide information to participants about their progression through the study. (2) The number of the current character rated over the total number of characters (*n*/105)—the system gave participants the option to stop the procedure after five images to avoid an extensive time commitment. (3) Additional instructions—participants could read instructions about the procedure at any time. (4) The character image drawn from a pool of 105 antagonists. We pseudo-randomized the presentation of images. Our image selection was automated to ensure that all images were presented with similar frequencies. To do this we grouped images by how often they had been previously rated by participants. Within the group of images that were rated the fewest times, we randomly selected five images and presented them in random order. The same procedure was performed for the next block of five images, omitting previously presented images out of the image pool. (5) A 7-point Likert-scale rating system for the morality scales used in Study 1, and a binary rating system for the salience of visual attributes in Study 2.

While still images are less rich in information than animated in-game characters, using still images and rating salient features in accordance with moral features strikes a balance between stimuli control, participant burden, and stimuli variance; i.e., displaying a large range of stimuli in a short amount of time. While limited when compared to experience of characters displayed in videogames, images sufficiently allow participants to identify visual character features that are perceived as salient.

Participants and Study-Specific Procedure

In Study 1 we assessed perceived character morality, and in Study 2 we collected binary ratings of the salience of visual attributes.

Study 1: Character Morality Ratings

For study 1, we recruited 99 participants. Four participants were removed from the data set, because they provided more than 25 ratings with a maximum variance ≤1, indicating a response pattern that was inattentive. Participants rated up to 21 sets of five images each. The first five images included additional demographic questions and were compensated by $1. The remaining 20 sets were compensated with 20 cents each, for a maximum of $5. In total, we obtained 5,963 ratings from 95 participants. Images received a minimum of 54 and a maximum of 60 ratings, mode = 57. For demographics, see **Table 1**.

[1]Our system was built in Python using the BOFS system (Johanson, 2019).

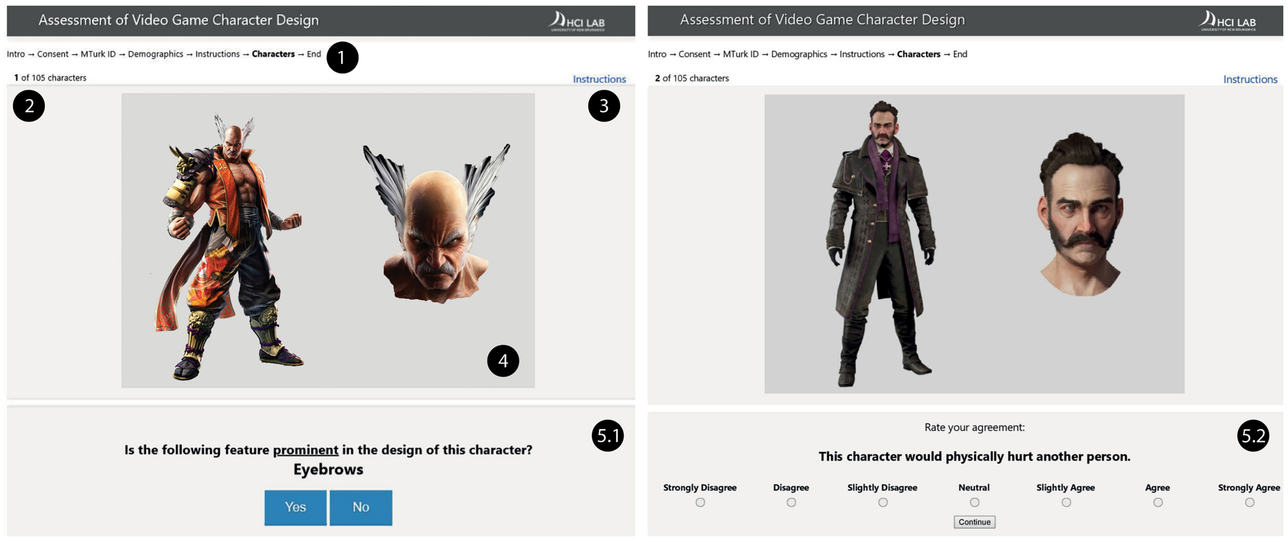


FIGURE 1 | Example character rating questions from character morality task interface used in Study 1 and the binary rating of salient visual attributes used in Study 2. Character images were adapted from fandom.com under the Creative Commons CC BY-SA license.

TABLE 1 | Demographics for Study 1 and Study 2.

	Study 1				Study 2			
Variables	***N***	**% (n/N)**	***M***	***SD***	***N***	**% (n/N)**	***M***	***SD***
Age	95		36.08	12.00	188	–	38.82	11.95
Gender	95				188	–		
Men		60% (57/95)				56.9% (107/188)		
Women		39% (37/95)				42% (79/188)		
Non-binary		0% (0/95)				0.5% (1/188)		
Prefer not to answer		1% (1/95)				0.5% (1/188)		
Playtime	95				188	–		
Everyday		40% (38/95)				36.7% (69/188)		
A few times per week		40% (38/95)				39.9% (75/188)		
A few times per month		18.9% (18/95)				6.4% (12/188)		
A few times per year		1.1% (1/95)				13.3% (25/188)		
Not at all		0% (0/95)				3.7% (7/188)		
Ethnicity	95				188	–		
Asian		9.5% (9/95)				7.4% (14/188)		
Black/African American		7.4% (7/95)				7.4% (14/188)		
Hispanic/Latino		6.3% (6/95)				6.4% (12/188)		
White		72% (69/95)				73.9% (139/188)		
Two or more categories		0% (0/95)				3.7 (7/188)		
Platforms	95				188	–		
Desktop		86.3% (82/95)				80.3% (151/188)		
Console		60% (57/95)				7.6% (127/188)		
Mobile		54% (56.8/95)				73.4% (138/188)		

Specific Procedure for Study 1: Ratings of Character Morality

To collect data on player's interpretation of characters, we used the short form of the Character Moral Foundations Questionnaire (CMFQ-S) (Grizzard et al., 2019), which is a validated short-form questionnaire based on Haidt and Joseph's (2008) five moral domains. Images in our stimuli set were rated on a 7-point Likert-scale, one time for each dimension of the five-dimensional CMFQ-S. The statements and morality domains were as follows, from the CFMQ-S (Grizzard et al., 2019):

- "Harm/care": This character would physically hurt another person.
- "Fairness/reciprocity": This character would deny others their rights.
- "Ingroup/loyalty": This character would betray their group.
- "Authority/respect": This character would cause chaos and disorder.
- "Purity/sanctity": This character would do something disgusting.

Seven-Point agreement ratings were converted to numbers prior to analysis where 1 means "Strongly Disagree" with one of the statements above; 4 is neutral; and, 7 "Strongly Agree." Following this a character who scores, say, a 1 for *harm/care* would be perceived to behave morally for the particular domain. Whereas, a character who scores 7 for *harm/care* would be perceived to behave strongly in an immoral way for the domain.

To control effects of familiarity, participants were instructed that their "[…] ratings should be made based on the appearance of the character only (in other words you should not use knowledge of the character to make your judgment)." Further, we asked participants to indicate whether they were familiar with a particular character, and to rate their overall familiarity with the character using a 100-point scale, by positioning a visual slider between "Not familiar at all" and "Very Familiar." In total, we collected 5,963 morality ratings from 95 participants. Images received a minimum of 54 and a maximum of 60 ratings, mode = 57.

Specific Procedure for Study 2: Identification of Prominent Visual Characteristics

In Study 2, we asked participants to indicate whether an attribute of an antagonist is salient or not. We decided to ask participants to rate individual character features instead of listing the most salient character features, because we were interested in a comprehensive analysis that allows for the evaluation of non-obvious visual character features that contribute to the overall perception of a character such as body alterations or age.

Attributes were derived from previous work on character attributes (McLaughlin, 2012) and extended by our own interpretation of relevant visual attributes of antagonists, resulting in a list of 24 visual attributes (see the full list in **Table 3**). A "salient" physical attribute was defined for participants as an attribute that "… relative to other characteristics stands out or grabs attention." Participants were prompted to make a binary decision to the statement "Is the following feature prominent in the design of this character?" followed by the name of the attribute (e.g., "eyes"). To stay away from overly scientific jargon, we used "prominence" instead of "salience" in our instructions to participants. For the most part these presented visual attributes simply stated the name of the attribute (e.g., "eyes," "hair," "nose," "mouth," etc.); however, some features required further explanation (i.e., "dermatological problems," "body weight," "build," "height," "head size," "skin exposure," "age," "stance," "clothing," "jewelry," "face cover," "body alterations"). In these cases, we provided a short description to provide clarification [e.g., "Dermatological problems (such as dark circles around the eyes, wrinkles, facial scars, warts, bulbous nose)"]; a full list of the visual features and descriptions has been provided in **Supplementary Material**.

For each character, participants initially responded to whether they were familiar with the character, and if so, how familiar (as in Study 1). Participants responded to all 24 attributes for each character, attributes were presented in serial, and participants were required to respond "yes" or "no" for each feature, before proceeding. To discourage participants from simply responding without considering each attribute, we imposed a brief 2 s delay before input would be accepted using the buttons.

For Study 2, we recruited 188 participants, who rated up to 21 sets of five images each (as previously described). Note the rating task in Study 2 took longer than in Study 1, hence more participants were recruited and each conducted fewer ratings on average. We received a total of 5,560 ratings from 176 participants. Images received a minimum of 48 ratings and a maximum of 59 ratings, mode = 54. See **Table 1** for demographics, and **Table 3** for a list of all 24 visual attributes. A limitation of Study 2 a lack of control for moral foundations, this is something we believe should be included in future work (we discuss further in limitations).

Analysis

All analysis was conducted using SPSS 25 (IBM, 2017).

Morality Score

Character Moral Foundations Questionnaire–Short (CMFQ-S) ratings (from Study 1) were transferred to score data (as described above), means were calculated by scale and data is presented in aggregate. The relationship between morality scores is evaluated using correlations and the average of all five scales is presented as a single "badness" score (see section Descriptive Statistics for Morality Ratings). Theoretically the morality dimensions are distinct, but the underlying assumption to either acting in line with a moral standard or not, is consistent across scales, and allows the scales to be combined if statistically internally consistent, as defined by Cronbach's alpha.

Salience Ratings

Salience ratings of visual attributes (from Study 2) were aggregated by calculating the percentage of participant responses that indicated a feature as being salient over the total number of responses by image. Ratings were normalized by the total number of responses to account for differences in the number of ratings received.

Broadly, we distinguish salience ratings in four blocks: (1) facial features and skin including features such as nose, mouth, or dermal problems; (2) body shape such as height or weight; (3) abstract features that depends on the viewer's judgment such as age or attractiveness; (4) accessories such as clothing or jewelery that could be removed.

Relationship Between Character Morality and Salient Features

To investigate the relationship between moral judgment and character features, we used correlations, and hierarchical

TABLE 2 | Pearson correlations for the five morality domains.

Domain	This character would…	1.	2.	3.	4.	5.
1. Care/harm	… physically hurt another person.	1	0.880**	0.751**	0.871**	0.782**
2. Fairness/reciprocity	… deny another person their rights.	0.880**	1	0.925**	0.968**	0.918**
3. Ingroup/loyalty	… betray his group.	0.751**	0.925**	1	0.926**	0.937**
4. Authority/respect	… cause chaos and disorder.	0.871**	0.968**	0.926**	1	0.906**
5. Purity/sanctity	… do something disgusting.	0.728**	0.918**	0.937**	0.906**	1

**$p < 0.01$.

TABLE 3 | Summary of hierarchical regression analysis for variables predicting badness ($N = 105$).

	Model 1 (Head)			Model 2 (+Body)			Model 3 (+Judgment)			Model 4 (+Accessories)		
Variable	***B***	***SE***	***β***	***B***	***SE***	***β***	***B***	***SE***	***β***	***B***	***SE***	***β***
Eye	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.02
Eyebrows	0.00	0.00	−0.03	0.00	0.00	−0.05	0.00	0.00	−0.06	0.00	0.00	0.02
Nose	0.01	0.01	0.11	0.01	0.01	0.12	0.01	0.01	0.13	0.01	0.01	0.15
Mouth	0.01	0.01	0.23*	0.01	0.00	0.19*	0.01	0.01	0.18*	0.01	0.00	0.19*
Ears	0.00	0.00	0.07	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.05
Skin problems	0.01	0.00	0.37*	0.01	0.00	0.37**	0.02	0.00	0.46**	0.02	0.00	0.43**
Facial hair	0.00	0.00	0.01	0.00	0.00	0.10	0.00	0.00	0.10	0.00	0.00	0.13
Hair	−0.01	0.00	−0.14	−0.01	0.00	−0.20*	−0.01	0.00	−0.14	0.00	0.00	−0.03
Weight				0.00	0.01	−0.08	0.00	0.01	−0.05	0.00	0.01	0.04
Build				0.01	0.01	0.10	0.01	0.01	0.11	0.00	0.01	0.05
Height				0.01	0.01	0.17	0.01	0.01	0.18	0.01	0.01	0.13
Head-body ratio				0.00	0.01	−0.06	−0.01	0.01	−0.10	−0.01	0.01	−0.11
Stance				0.01	0.00	0.32**	0.01	0.00	0.30**	0.01	0.00	0.22*
Skin color							−0.01	0.00	−0.11	−0.01	0.00	−0.09
Masculinity/Femininity							0.00	0.01	−0.01	0.00	0.01	0.05
Attractiveness							−0.01	0.00	−0.10	0.00	0.00	−0.07
Skin exposure							0.00	0.01	0.01	0.00	0.01	−0.02
Age							−0.01	0.00	−0.20*	−0.01	0.00	−0.15
Clothing										0.00	0.00	0.08
Jewelery										0.00	0.00	−0.07
Face cover										0.01	0.00	0.19
Tattoos										0.00	0.01	0.02
Weapon										0.01	0.00	0.21*
Body alterations										0.00	0.00	0.01
R^2		0.390			0.556			0.603			0.669	
F for change in R^2		7.68**			6.80**			2.06			3.19*	

*$p < 0.05$, **$p < 0.01$.

regression analysis to identify our final set of most predictive character features.

We present further details on the statistical procedures used at the beginning of each subsection in the Results section.

Results

We present the results of both studies together for simplicity, and since much of our analysis examines correlation between the ratings collected in each study. We refer specifically to morality ratings (gathered in Study 1) and visual attribute salience (gathered in Study 2) in order to reference the source of the data.

The number of ratings collected per image varied slightly between images in both studies; *Study 1*: min = 54, max = 60; and, *Study 2*: min = 52, max = 62.

Descriptive Statistics for Morality Ratings

For Study 1, mean and standard deviation for each moral domain were calculated per image. Recall that ratings were made on a 7-point Likert-scale ranging from strongly disagree (1) to strongly agree (7) with a "lack of morality statement," e.g., "This character would hurt another person."—high scores suggest perceived immorality.

In the morality rating task, participants were only familiar with 16.454%, and of those characters they were familiar with, they rated their familiarity low ($M = 68.056$, $SD = 10.113$).

For our overall stimulus set, we have the following results for each moral domain: "Care/harm" ($M = 5.030$, $SD = 1.110$), "Fairness/reciprocity" ($M = 4.618$, $SD = 1.042$), "Ingroup/loyalty" ($M = 4.185$, $SD = 0.830$), "Authority/respect" ($M = 4.675$, $SD = 1.082$), "Purity/sanctity" ($M = 4.100$, $SD = 1.021$). Surprisingly this might indicate that current (human) villain designs in video games do not convey that they are clearly immoral (mean ratings for each or only slightly on the immoral side of neutral). This is in contrast, perhaps, to previous studies of animated Disney villains who are unmistakably evil-looking (Hoerrner, 1996).

To better understand how the different morality domains might be related, we looked for correlations between the five items and found high correlations between domains; see **Table 2**. After evaluating the reliability of across scales (Cronbach's-$\alpha = 0.972$), we calculated a single "morality" score for each character by taking the mean across of all five dimensions ($min = 2.20$, $max = 6.16$, $mean = 4.52$, $SD = 0.969$). The distribution of "morality" is negatively skewed ($skewness = -0.377$), which is in line with our expectations, considering that the source of our images are examples of video game villains.

Character Examples of Moral Dimensions

To explore how individual characters moral dimensions compare to the data set at large, we calculated categories based on standard deviations for each domain. We created four categories reflecting the morality for each domain: moral (<-2 SD), slightly moral (−1 SD to 0), slightly immoral (0 to +1 SD), and immoral (> +2SD), we binned morality to identify divergence from the mean and variation between morality domains; i.e., being corrupted in one domain, but uncorrupted on all other domains. In **Figure 2**, categories were solely created for illustration purposes and are not used in any further analysis.

To exemplify the presence of different moral characteristics, we present twelve characters with varying pronunciations in the five moral domains (see **Figure 2**). Tsumugi Shirogane was viewed consistently as being uncorrupt ("Care/harm": moral, "Fairness/reciprocity": moral, "Ingroup/loyalty": moral, "Authority/respect": moral, "Purity/sanctity": moral). These categories mean that this character is perceived as a character that would **not** physically hurt others, would **not** deny another person's rights, will **not** betray her group, would cause chaos and disorder, and is **not** disgusting. Ryuji Goda on the other hand, is almost completely the opposite ("Care/harm": immoral, "Fairness/reciprocity": immoral, "Ingroup/loyalty": slightly immoral, "Authority/respect": immoral, "Purity/sanctity": immoral); this character is perceived as strongly immoral.

Some characters show interesting patterns where they score differently on different moral dimensions. Aaron Keener ("Care/harm": immoral, "Fairness/reciprocity": slightly immoral, "Ingroup/loyalty": slightly moral, "Authority/respect": slightly immoral, "Purity/sanctity": slightly moral), who is perceived as careless, slightly denying other their rights, and willing to cause chaos and disorder. But he is not perceived as disloyal or disgusting. Yunica and Heiss follow similar patterns. Our results show how the dimensions of badness can be used to analyze and compare characters and find character designs that provide both strong and nuanced perceptions of morality. We leave further commentary and interpretation on these examples to the Discussion, after the results regarding visually salient features have been introduced.

Predicting the Morality Through Aesthetic Characteristics

To investigate the relationship between aesthetic features and a character's perceived morality, we performed hierarchical regression analysis with our 24 visual attributes grouped into four blocks: head, body, interpretative characteristics (e.g., age, masculinity-femininity), and presentative characteristics (e.g., clothes, tattoos); see **Table 3** for the full list of visual attributes in each of the blocks. Blocks were entered as predictor variables of perceived badness. Our results show that head characteristics, body characteristics, and presentative characteristics have the most predictive value when predicting badness ($p = 0.011$, $R^2 = 0.669$). As displayed in **Table 3**, the salience of the mouth, skin problems, the stance, and weapons, are the best predictor of badness.

Our final model (Model 4), shows that a combination of the mouth ($\beta = 0.19$), skin problems ($\beta = 0.43$), stance ($\beta = 0.22$), and a weapon ($\beta = 0.21$), are the strongest predictors of morality ($R^2 = 0.669$). While these visual attributes are the most predictive for morality, related visual attributes should also be considered when analyzing characters or planning the visual design of an immoral character.

The Relationship Between Aesthetic Elements

We next analyzed our data for trends that demonstrate which aesthetics elements are perceived as most salient together in antagonist designs. We showed that the mouth, skin problems, stance, and weapon, are the most predictive variables for morality. However, several of the visual attributes in the model show interdependencies; e.g., when a facemask is present, the mouth cannot be seen, or stance might be related to the presence of a weapon. We calculated correlations between visual attributes to discover attribute that are closely related to the most predictive visual attributes. Considering that there are many minor correlations between the salience of different visual attributes, we only discuss visual attributes that correlate $r > 0.25$ with visual attributes relevant for the prediction of morality. See **Table 4** for the correlation table, and **Figure 2** for examples of characters with different salient visual attributes.

Skin problems are correlated with the salience of mouth ($r = 0.58$), nose ($r = 0.41$), ears ($r = 0.28$), hair ($r = 0.28$), and tattoos ($r = 0.34$). Suggesting that skin problems appear in prominent parts such as the face and are used in combination with other facial features as we can see in a character like Vitalis.

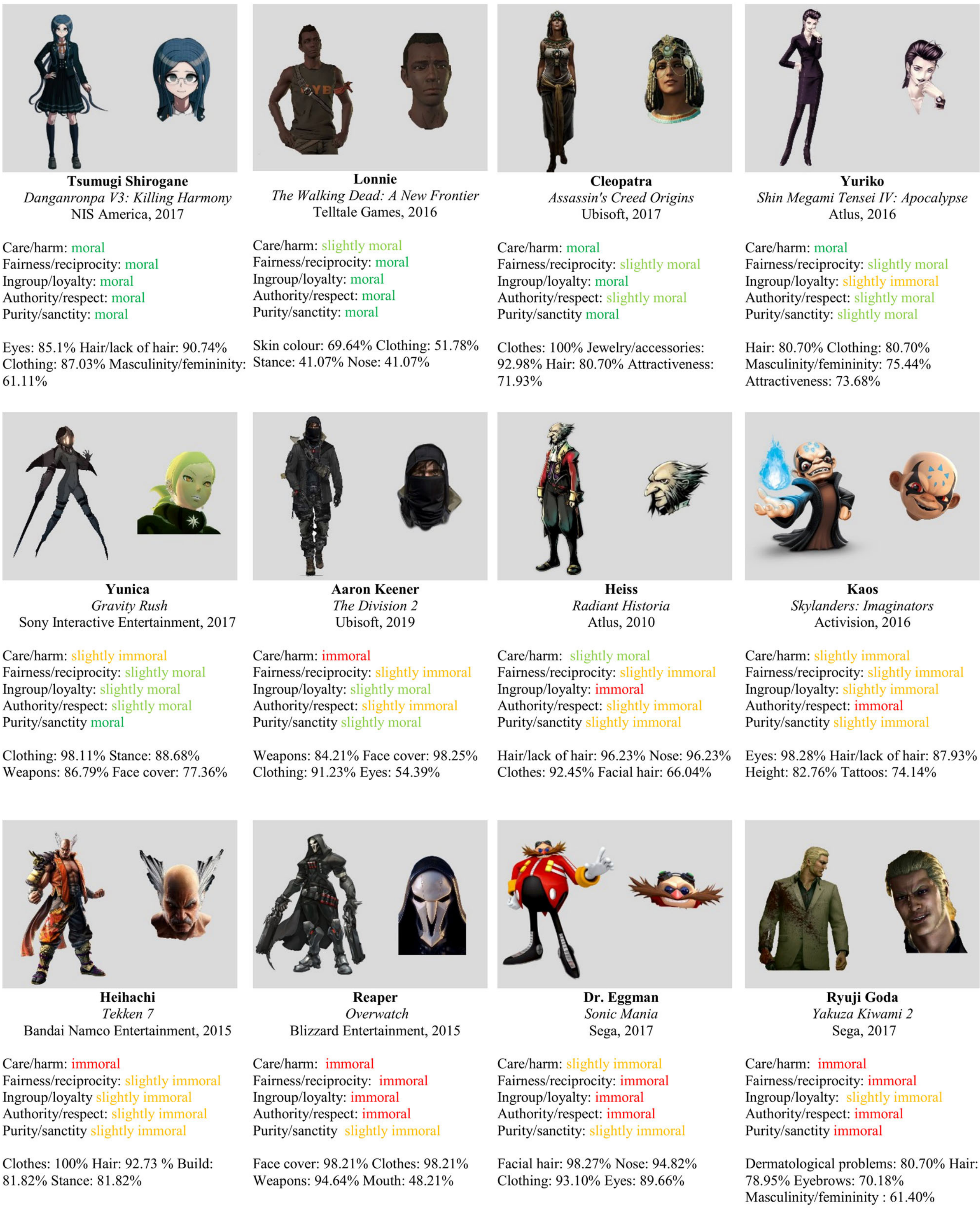


FIGURE 2 | Examples of video game characters and their morality by domain. Expression by domains is color coded from red (immoral) to green (moral). Character images were adapted from fandom.com under the Creative Commons CC BY SA license.

TABLE 4 | Pearson correlation of attributes.

Variable	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.	23.	24.
1. Eyes	1.00																							
2. Eyebrows	0.14	1.00																						
3. Nose	0.10	0.22	1.00																					
4. Mouth	**0.30**	0.18	**0.30**	1.00																				
5. Ears	**0.29**	0.24	**0.26**	0.10	1.00																			
6. Skin problems	0.17	0.20	**0.58**	**0.41**	**0.28**	1.00																		
7. Facial hair	−0.02	**0.37**	**0.33**	0.08	0.05	0.18	1.00																	
8. Hair	**0.27**	0.17	−0.01	0.04	**0.28**	−0.05	0.12	1.00																
9. Weight	0.20	0.12	0.13	0.19	−0.02	0.05	0.16	−0.06	1.00															
10. Build	0.20	0.15	**0.25**	0.20	0.13	0.21	0.05	0.09	**0.55**	1.00														
11. Height	0.30	0.15	0.03	0.21	0.20	0.06	−0.01	0.18	**0.51**	**0.54**	1.00													
12. Head-body ratio	**0.45**	0.13	0.12	**0.29**	0.13	0.05	0.09	0.18	**0.54**	**0.39**	**0.58**	1.00												
13. Stance	0.24	0.00	−0.06	0.11	0.15	−0.02	−0.19	0.16	0.16	**0.33**	**0.48**	0.16	1.00											
14. Skin color	**0.26**	0.10	0.12	**0.41**	0.11	**0.25**	−0.02	0.13	0.12	**0.34**	0.12	0.09	0.13	1.00										
15. Masc.-Fem.	0.12	0.21	−0.07	0.06	0.02	0.03	0.05	0.22	0.14	**0.45**	0.13	0.04	0.20	0.18	1.00									
16. Attractiveness	0.24	0.04	0.07	0.24	0.09	0.11	−0.12	**0.27**	−0.01	0.04	0.14	−0.01	0.17	0.18	**0.30**	1.00								
17. Skin exposure	0.20	−0.10	−0.08	0.02	0.15	−0.01	−0.19	0.12	0.01	**0.40**	0.13	0.10	0.23	0.23	**0.37**	0.16	1.00							
18. Age	0.16	0.03	0.18	−0.12	0.10	**0.25**	0.16	0.09	0.07	0.02	0.03	0.08	−0.09	−0.10	0.04	0.01	0.06	1.00						
19. Cloth	0.18	0.05	−0.20	0.03	0.01	−0.07	−0.15	0.12	0.12	0.14	**0.36**	0.04	**0.48**	−0.05	0.10	0.16	0.09	−0.05	1.00					
20. Jewelery	**0.31**	0.12	−0.15	0.08	−0.01	−0.11	0.00	0.01	0.11	0.14	0.12	0.01	0.23	0.22	**0.28**	**0.28**	0.19	−0.01	**0.50**	1.00				
21. Face cover	0.07	**−0.33**	−0.20	0.03	−0.15	−0.09	**−0.25**	**−0.39**	0.02	0.09	0.09	0.10	0.13	0.03	−0.21	−0.14	0.23	−0.17	**0.27**	0.23	1.00			
22. Tattoos	0.20	0.14	0.19	0.22	**0.34**	0.23	−0.01	0.18	−0.04	0.19	**0.29**	0.15	0.17	**0.27**	0.09	0.05	**0.27**	−0.06	0.05	0.09	0.05	1.00		
23. Weapons	0.00	**−0.28**	−0.08	−0.09	−0.18	−0.07	**−0.29**	**−0.26**	−0.12	0.03	0.04	−0.09	**0.25**	−0.03	−0.03	−0.09	0.01	−0.18	0.02	0.02	**0.41**	0.06	1.00	
24.Body alterations	0.11	−0.20	−0.03	−0.03	0.04	−0.03	−0.20	−0.09	0.11	**0.29**	0.13	0.03	**0.30**	0.04	0.01	0.01	0.12	−0.12	0.22	0.13	**0.44**	0.06	**0.35**	1.00

$p < 0.01$ is highlighted in bold.

Stance correlates with build ($r = 0.33$), height ($r = 0.48$), clothes ($r = 0.48$), and weapons ($r = 0.25$). Stance highlights a character's build and height and makes impressive or distinct physiques standout or appear intimidating—an approach to convey morality. Clothing can be used to further highlight certain aspect of the character's physique, such as partially revealing skin, Heihachi's power pose, for example, is underlined by his clothing that partially reveals his impressive muscles that suggest that he is able to inflict physical pain. Characters with weapons are often presented in stances that relate to the weapon's fighting style, e.g., Yunica shows a fencing posture while holding a sword-like weapon (see **Figure 2**).

Weapons are also correlated with face coverings ($r = 0.41$) and body alterations ($r = 0.35$). One class of characters using weapons are assassins such as Reaper or Aaron Keener, who cover their faces to avoid recognition. However, face coverings can also prevent clearly visible facial expressions, which means that other visual attributes must convey morality; e.g., the stance and weapons of a character can display aggression and the ability and willingness to harm others, such as Yunica and Aaron Keener (see **Figure 2**).

While there are many visual attributes, representing archetypes, eliciting a specific perception of a character, and how we experience characters aesthetically leads to relationships between attributes that can be further explored to understand the visual construction of antagonists. In **Table 5** we present the relationships of salience and badness for each visual attribute. The table contains the regression coefficients for either a linear or quadratic relationship between salience and badness, and a sparkline visualizations that displays an overview of the relationship. The left end of the x-axis for the sparkline is more badness, the right end is less badness; higher on the y-axis indicates higher salience for the given level of badness.

DISCUSSION

The results of our analysis provide an exploration and data-driven insights in players perceive perception of characters morality.

Our work explores the relationship between salient visual attributes of villains and their perceived morality. We provide insights into the relationship between salient visual features, showing which aspects are combined to define perceived morality, and which features predict perceived badness of a character.

In the subsections below, we organize our discussion around how people perceive game antagonists and how designers might leverage our results in their design practices. We then discuss limitations of our current studies and the new directions that our work makes possible for future work.

The Design and Perception of Game Antagonists

Overall, the game antagonists in our stimulus set were viewed as only slightly corrupt, or just slightly more immoral than a neutral rating. The binning of characters by their overall morality, illustrates an only slightly negative skew toward immorality, and in fact roughly 47 of the 105 characters were viewed as being moral than immoral (see section Character Examples of Moral Dimensions). While we had some examples where characters were viewed consistently and strongly as immoral, the tendency of our sample is only subtly evil characters. We do not believe this means that our stimulus set is somehow limited or that games do not represent antagonists who are as "bad" as in other media. Rather, we believe that this shows that games often provide more nuanced visuals and storytelling when it comes to their antagonists, which can be easily seen in some of our examples (see **Figure 2**). Our stimulus set is in stark contrast to other image stimulus sets used for understanding perceptions of character morality. For example, Disney villains were uniformly viewed as strongly evil (Hoerrner, 1996). In this example, however, the approach is to communicate very clearly to the (sometimes young) audience exactly who the villains are.

While we did not study behavior and acts of evil or immoral behavior, the fact that game designers often explore less overt visual representations of "bad" is interesting, because interactivity provides other means to display evil behavior and the time spent with a game is significantly longer than watching a movie, which allows to discover the evil side of a character over time—similar to TV shows. Of course, and importantly, antagonists in stories are not always "evil" (Martin Del Campo, 2017). That being said, we did review the characters in our stimulus set, and our interpretation of the back stories of almost all, if not all, characters suggested that these characters did indeed take actions that harmed others, were disgusting, were betrayals to their group, etc.; i.e., they were immoral in action, even if their visual design did not suggest it. The fact that the visuals of characters do not always portray outward and strong aggression, for example, reflects the range of ways that game designers tell stories and they support those stories visually through their character designs. Indeed, it is a surprise in the story of the game *Danganropa* that a particularly friendly looking character (Tsumugi Shirogane) is indeed the antagonist. In contrast, other stories might present characters who have seemed very immoral, in both appearance and action, but might still perform good acts. For example, disillusioned with the evil covenant, The Arbiter (a grotesque alien) switches sides joining forces with Master Chief, the main protagonist in *Halo 2* (Bungie, 2004).

Our analysis of character ratings in each of the five morality domains are highly correlated with one another. This suggests that when we judge a character as morally corrupt the distinction between moral dimensions is often unclear—evil is evil. This is, however, out of line with previous work in assessing idealized protagonists (and not game protagonists) using the CFMQ-S instrument (Grizzard et al., 2019), where the domains were not strongly correlated. The strong correlations could mean that game designers more uniformly represent all domains of morality when creating antagonists. We discuss how this opens up possibilities for designers below.

The Constructions of Villainous Stereotypes

Tamborini's model of intuitive morality and exemplars (MIME) (Tamborini, 2011) provides a short and long-term perspective on morality and adds intuitive and emotional aspects to the

TABLE 5 | Non-standardized regression coefficients for individual characteristics and badness, and sparkline visualizations for the highest significant order effect.

Characteristic	Linear	Quadratic	Unstandardized coefficients			Sparkline
	p	*p*	*Intercept*	*a*	*b*	
Eyes	0.07	0.026*	110.679	−34.64	4.485	
Eye brows	0.236	0.147	15.926	2.705	–	
Nose	<0.001**	<0.001**	20.162	−2.588	1.257	
Mouth	<0.001**	<0.001**	92.957	−39.649	5.515	
Ears	0.045*	0.123	−0.998	4.072	–	
Dermatological problems	<0.001**	<0.001**	14.398	−10.97	3.16	
Facial hair	0.303	0.088	13.037	4.08	–	
Hair	0.277	0.483	76.133	−3.412	–	
Weight	0.078	0.015*	73.933	−37.654	4.024	
Build	<0.001**	<0.001**	63.921	−28.102	4.031	
Height	0.001**	0.003**	15.437	−3.976	1.037	
Head–body ratio	0.139	0.005**	89.735	−39.588	4.868	
Skin color	0.086	0.012**	89.278	−33.819	4.262	
Masc.–Fem.	0.639	0.891	39.602	0.768	–	
Attractiveness	0.821	0.577	38.511	0.433	–	
Skin exposure	0.624	0.117	6.542	0.758	–	
Age	0.345	0.584	33.552	−1.75	–	
Stance	<0.001**	0.001**	11.13	10.072	−0.24	
Clothes	0.057	0.156	57.771	4.335	–	
Jewelery	0.839	0.979	26.187	0.468	–	
Face cover	0.016*	0.008**	75.715	−35.959	4.864	
Tattoo	0.005**	0.009**	19.196	−11.624	1.82	
Weapons	0.003**	0.012**	−33.225	15.743	−0.71	
Body alterations	0.015*	0.01**	33.329	−16.569	2.288	

The left end of the x-axis for each sparkline is more badness, the right end is less badness; higher on the y-axis indicates higher salience for the given level of badness. *$p < 0.05$, **$p < 0.01$.

processing of character judgments. When applying MIME to interpret our stimulus set it is important to keep in mind that, following exemplification theory (Zillmann, 1999) frequent exposure to moral examples increase the effect of media exposure, e.g., the frequently displayed character with baggy pants, muscle shirt, and bandana who kills, robs, and sells drugs, has created the powerful iconic image of the ghetto gangster.

Considering the most predictive characteristics in our stimulus set (i.e., weapon, dermatological problems, stance, and the character's mouth), we can consider how these characteristics contribute to villainous stereotypes. While the relationship between carrying a weapon and the stance of a character can be directly linked to "badness" through social norms—a weapon suggests hostility, and a powerful or combative stance demonstrates aggressiveness the role of dermatological problems and the mouth, are unexpected and culturally insightful. The mouth plays an important role in communicating emotions or intent in western cultures (Yuki et al., 2007), e.g., signaling approach-ability by smiling, or emotions such as anger, fear, or disgust. For character design, the mouth opens up opportunities to communicate the internal state a character, e.g., the evil grin of the Disney character Ursula. Model of intuitive morality and exemplars would suggest that when dermatological issues are used to depict villains, an automatic negative response to other characters with skin problems would result. This means that villains with acne, burn scars, etc., might lead to other uses of those same visual attributes, leading to coherence between domain and exemplar salience. The consequence of this can result in real consequences for individuals in their day-to-day lives with skin problems, who could be more likely perceived as villains (Funk and Todorov, 2013). Such stereotypes would need to be counteracted by creating content where domain and exemplar salience conflict; e.g., characters with skin problems that are inherently good. An example character that already shows a manifestation of such conflict is Marvel's anti-hero Deadpool, who, under his mask, has substantial scarring. Deadpool, fights for good, but is also tortured and mischievous, his scarred face underlies his ongoing conflict with society that find his appearance repugnant. It is important to note that different skin issues are perceived differently, while acne, pock marks, or scars have been connected to criminal stereotypes before (MacLin and Herrera, 2006), more fine-grained analysis show that individuals with acne are perceived as shy or insecure (Dréno et al., 2016).

In games, dermatological issues can be used to conjure associations with badness, but especially scars provide opportunity to reshape the perception of scar tissue—for example, when used as aesthetical signifiers or to memorize special events. Scars could, for example, be visible on characters as a badge of defeating a difficult final boss or for taking part in a challenging battle. In different cultures, scars also have different meanings. Scarification—the deliberate act of scaring someone for aesthetical purposes—has roots in traditions of African tribes and has found its way into body modification culture. Directions that games could use, for example, to provide new character options increasing diversity through customizability (Dickerman et al., 2008; Birk et al., 2016; Passmore et al., 2018), by allowing characters to be created that defy negative stereotype associated with dermatological issues, or to create a visual language around the beauty of scars.

While dermatological problems tie into stereotypes and negative expectations, the mouth is one of the most important features used in facial expression and to communicate non-verbally. Smiling, baring teeth, or pulling the corners of our mouth down, are facial expression that can be inviting, display aggression, or disdain. In our analysis the mouth is a strong predictor of badness, showing a reversed u-shaped relationship between the relative salience of the mouth prominence and badness, i.e., the mouth is salient for those who rated the character as being the most moral (good) and the least moral (bad), but in-between the extremes the mouth tends not to be salient. Considering the importance of facial expressions to communicate intentions, e.g., aggression vs. friendliness, we can assume—and the quadratic relationship confirms this—that the mouth will also play an important role to judge morally "good" characters.

How Game Designers Can Use the Results

Our results expose new ways that designers can try to push their designs to leverage commonly used visual attributes in order to get a reliable and effective morality interpretation for their antagonists. Designers might also use our results to identify new design alternatives that have not been previously well-explored. In this section we speculate how our results can be used by game designers.

Characters who people perceived as the most immoral leveraged many common visual features. The most immoral characters disproportionately featured salient eyes, noses, mouths, ears, skin problems, builds, head-to-body ratios, ages (especially appearing older), clothing, face coverings, tattoos, and weapons. These results highlight a large number of features that designers can leverage and try to strengthen and make more salient in their designs to make certain a character is perceived as immoral (based on their visual appearance).

Characters seen as the most moral did not leverage many salient visual attributes. This makes sense as characters who are viewed as immoral leveraged visual attributes often in combination or exaggerated ways, making them standout (e.g., consider the exaggerated head-to-body ratio of Neo Cortex who was viewed as strongly immoral). Attractiveness was the only physical attribute we found that was used disproportionately more for the most moral characters (roughly at the same rate as the most immoral characters, but disproportionately more than other characters). Given that people tend to consider "beautiful" people as "good" (Diessner et al., 2008), it is insightful that many participants rated beautiful characters, that had fewer other salient features, as more moral. This highlights that designers might consider designing antagonists that players view as being moral, while providing a salient physical attribute commonly associated with immoral characters. For example, participants who rated characters as having salient tattoos did not rate the same antagonists as having attractiveness as a salient feature (e.g., Kaos). So, designers might explore the combination of both attractiveness and a feature like tattoos in antagonists. It is important to note that attractiveness was not a pre-requisite for being viewed as moral (e.g., Lonnie was perceived to be moral, but was rated relatively low in terms of having salient attractiveness; see **Figure 2**).

As previously described, our analysis of character ratings in each of the five morality domains were highly correlated with one another. This means that designers tend to present characters that are uniform across all domains. However, our analysis revealed interesting exceptions to this trend. For example, the weapons and stances of Yunica (**Figure 2**) strongly suggest that they are willing and ready to harm others, while ratings of the other morality domains suggested that they would be unlikely to do something disgusting and would remain loyal to their groups. Similarly, Yuriko's slicked hair and businesswoman attire, suggested to people that she appeared less loyal, but unlikely to physically hurt people. Exploring ways that characters could be designed to create other variations of morality across the domains may provide interesting possibilities and directions for designers.

Finally, designers might leverage our results directly to plan and gauge how their planned visual attributes for an antagonist would be perceived by players. This might be important for game designers who seek to include a range of visual attributes. While we have not provided a predictive model, below we elaborate on our planned future work to build tools to support and evaluate character design activities that leverage predictive models that can be built based on work such as ours.

Limitations

Our choices for inclusion/exclusion when creating our stimulus set likely had an effect on our results. Our stimulus set does not represent all antagonists in all games. Recall that we excluded non-humanoid characters, and indie game characters, games that were not among the most popular games, and games published before 2014. Together these decisions likely had some influence on our observations and models. Firstly, as described, we did this to ensure that we had games in our stimulus set that are of a high quality and represent current trends in the industry. Second, this still represents a large set of games and/or games that are exemplary for the investigated time frame. Any attempt to operationalize current practices in a large space will necessarily need to make trade-offs. We believe that our results provide both important new insights for designers and researchers and provide a template of a new style of study for modeling the aesthetic practices in game design that can have important applications.

We see another important limitation of our work to be the use of static images. In many of the games players gather further impressions of characters through the way they move. Our use of static images, while drawn from a wide variety of games, does not fully capture other aspects of the visual design of antagonists. In particular, body language, movement, or speech might be used by artists and animators to more fully communicate information. For example, a character's stance is often tightly integrated with animation, to help convey tensed or relaxed muscles. Further, we provided only two images upon which judgments can be based. Even in games that do not use animation, different graphical still shots are used to display expressions of emotion (e.g., anger, happiness, surprise, aggression, etc.). Future work should consider displaying a richer set of media to solicit judgments from raters.

Further, in our analysis we do not consider the behavior or actions of characters, which obviously play prominently in how people would perceive their morality. However, this type of analysis is out of scope of our current research. We were focused purely on how game designers embody their character's morality through visual design. An interesting, but extremely challenging, line of research might explore common story telling techniques around characters to understand how these impact key elements of player perception of those characters.

Finally, we see our participant sample as a potential limitation of our work. While we believe our sample did achieve a reasonable mixture of gaming backgrounds, it could be that this demographic may not uniformly represent the cultural views and experiences that readily exist amongst gamers, or in gaming culture. That is, it seems likely that people familiar with games might carry their pre-existing knowledge of archetypes, tropes, stereotypes, running narratives, etc. that exist between games, and that people who are more familiar (enculturated) with gaming culture, might reveal completely different and, perhaps, more nuanced views and understanding of characters. While we have found no evidence to suggest that this is the case, future work might also incorporate perspectives of gaming culture and how it might affect perceptions of character design and moral judgments. Additionally, we did not control for moral leanings in our convenience sample from Mechanical Turk. Previous work has found that MTurk samples tend to be similar to student samples regarding political leaning, but proportionally more secular (Lewis et al., 2015). Nevertheless, given samples similarly sized to ours (186 participants in Study 2) future work should capture moral foundations to provide a better understanding of the respondents and how their moral leanings might have influenced their ratings.

Future Work

In this work we focused on antagonists, since they are underexamined yet play a critical role in many modern games. Our work is the first that we are aware of to take this particular approach of formulating a stimulus set that captures visual design practices, gather data describing people's perceptions of key design features and interpretation of those feature based on the stimulus set, and to characterize it using descriptive analysis and correlational models. We believe this work demonstrates an approach to an exciting direction of research that aims to build an understanding of game design practices and to make new computational support tools for game design possible. This is similar to the goals in the field of computational aesthetics, yet we believe that rather than automating many of these classically human-led endeavors, we wish to conduct research that will better support current practices and provide new directions for game design.

Along with the broader goal of exploring game designs and aesthetics through computational approaches, we believe there are a number of direct next steps that our work offers. First, we focused on antagonists in this work; however, previous work in media studies focused on idealized, animated protagonists; however, this other previous work did not focus on current practices in video games. We believe it would be extremely interesting to repeat our study with both game protagonists and antagonists, and to draw comparisons across studies. We would also like to expand and mature our work on antagonist design (and video game character design more generally), providing guidelines for designing formidable video game villains, protagonists, and non-player characters. Our work accounts for physical appearances and not character actions or mechanics in the game, we believe we can explore game mechanics and behavior and actions in story elements of the game to more holistically describe character designs. Finally, we would like to explore how our statistical approach might inform the design of tools to assist designers in assessing designs. These could take the form of predictive tools to provide informed estimates about the potential morality of a particular design and/or enrich data for modeling through crowdsourcing perception of visual attributes and morality.

CONCLUSIONS

Antagonists are critical elements of many games, but as of yet no previous work has explored one of the key ways that players experience characters, through their visual design. Our work provides a first empirical characterization of how game designers represent game antagonists and how people perceive these characters in terms of their morality. To do this we conducted two studies on Mechanical Turk to solicit ratings. The first study collected people's judgments on the perceived morality based solely on the visual design for each of the 105 characters in our stimulus set. The second study gathered judgments on which visual attributes are more salient. Our analysis provides a valuable characterization of current design practices and how players perceive game antagonists, and provides a number of key ways that designers can strengthen their antagonist's visuals and ways that they can break from current trends to explore new ways to visually represent their characters. Our research extends current research practices that seek to build an understanding of game design practices, and provides exciting directions exploring how design and aesthetic practices can be better studied and supported.

DATA AVAILABILITY STATEMENT

The raw data sets collected from the populations in both studies have been made publicly available at: https://dx.doi.org/10.17605/OSF.IO/V2B7A.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Research Ethics Board University of New Brunswick (REB 2019-118). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

The creation of the stimulus set and initial literature review was led by RP, with the support of MB and SB. The experimental data collection system and experimental system was led by SB, with assistance from MB and RP. The statistical analysis was led by MB, with assistance from SB and RP. All authors contributed to the article and approved the submitted version.

FUNDING

This work was partially supported by the Natural Science and Engineering Research Council Canada and the New Brunswick Innovation Foundation.

ACKNOWLEDGMENTS

We would like to thank the reviewers of our manuscript who provided extensive constructive feedback that substantially improved the analysis and presentation of our work. We would also like to thank the anonymous participants in our research from the Mechanical Turk platform.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomp.2021.531713/full#supplementary-material

REFERENCES

Ahler, D. J., Roush, C. E., and Sood, G. (2019). *Micro-Task Market for Lemons*. 40. Chicago, IL.

Bakkes, S., Tan, C. T., and Pisan, Y. (2012). Personalised gaming: a motivation and overview of literature. *J. Creat. Technol.* 4, 1–10. doi: 10.1145/2336727.2336731

Bar, M., Neta, M., and Linz, H. (2006). Very first impressions. *Emotion* 6, 269–278. doi: 10.1037/1528-3542.6.2.269

Baranowski, T., Buday, R., Thompson, D. I., and Baranowski, J. (2008). Playing for real: video games and stories for health-related behavior change. *Amer. J. Prevent. Med.* 34, 74.e10–82.e10. doi: 10.1016/j.amepre.2007.09.027

Birk, M. V., Atkins, C., Bowey, J. T., and Mandryk, R. L. (2016). "Fostering intrinsic motivation through avatar identification in digital games," *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI'16* (San Jose, CA), 2982–2995. doi: 10.1145/2858036.2858062

Birmingham, E., Bischof, W. F., and Kingstone, A. (2009). Saliency does not account for fixations to eyes within social scenes. *Vis Res.* 49, 2992–3000. doi: 10.1016/j.visres.2009.09.014

Boltanski, L., and Thévenot, L. (2000). The reality of moral expectations: a sociology of situated judgement. *Philos. Explor.* 3, 208–231. doi: 10.1080/13869790008523332

Boyan, A., Grizzard, M., and Bowman, N. (2015). A massively moral game? Mass effect as a case study to understand the influence of players' moral intuitions on adherence to hero or antihero play styles. *J. Gam. Virt. Worlds* 7, 41–57. doi: 10.1386/jgvw.7.1.41_1

Bull, R. H. C., and Green, J. (1980). The relationship between physical appearance and criminality. *Med. Sci. Law* 20, 79–83. doi: 10.1177/002580248002000202

Classic Villain—TV Tropes, n.d. available online at: https://tvtropes.org/pmwiki/pmwiki.php/Main/ClassicVillain (accessed January 31, 2020).

Croley, J., Reese, V., and Wagner, R. (2017). Dermatologic features of classic movie villains: the face of evil. *JAMA Dermatol.* 153, 559–564. doi: 10.1001/jamadermatol.2016.5979

DeVane, B., and Squire, K. D. (2008). The meaning of race and violence in grand theft auto: San Andreas. *Games Cult.* 3, 264–285. doi: 10.1177/1555412008317308

Dickerman, C., Christensen, J., and Kerl-McClain, S. B. (2008). Big breasts and bad guys: depictions of gender and race in video games. *J. Creat. Mental Health* 3, 20–29. doi: 10.1080/15401380801995076

Diessner, R., Solom, R. C., Frost, N. K., Parsons, L., and Davidson, J. (2008). Engagement with beauty: appreciating natural, artistic, and moral beauty. *J. Psychol.* 142, 303–329. doi: 10.3200/JRLP.142.3.303-332

Doris, J., Stich, J., Phillips, S., and Walmsley, L. (2020). "Moral psychology: empirical approaches," in *The Stanford Encyclopedia of Philosophy* (Spring 2020), ed E. N. Zalta (Metaphysics Research Lab, Stanford University). Available online at: https://plato.stanford.edu/archives/spr2020/entries/moral-psych-emp/ (accessed March 15, 2021).

Dréno, B., Tan, J., Kang, S., Rueda, M.-J., Torres Lozada, V., Bettoli, V., et al. (2016). How people with facial acne scars are perceived in society: an online survey. *Dermatol. Ther.* 6, 207–218. doi: 10.1007/s13555-016-0113-x

Eden, A., Daalmans, S., and Johnson, B. K. (2017). Morality predicts enjoyment but not appreciation of morally ambiguous characters. *Media Psychol.* 20, 349–373. doi: 10.1080/15213269.2016.1182030

Eden, A., Oliver, M. B., Tamborini, R., Limperos, A., and Woolley, J. (2015). Perceptions of moral violations and personality traits among heroes and villains. *Mass Commun. Soc.* 18, 186–208. doi: 10.1080/15205436.2014.923462

Eden, A., Tamborini, R., Grizzard, M., Lewis, R., Weber, R., and Prabhu, S. (2014). Repeated exposure to narrative entertainment and the salience of moral intuitions. *J. Commun.* 64, 501–520. doi: 10.1111/jcom.12098

Ekström, H. (2013). *How Can a Character's Personality be Conveyed Visually, Through Shape*. 54. Visby.

Fahraeus, A., and Yakali Çamoglu, D. (2011). *Villains and Villainy: Embodiments of Evil in Literature, Popular Culture and Media,* Vol. 76. Rodopi. Available online at: http://urn.kb.se/resolve?urn=urn:nbn:se:hh:diva-19409 (accessed March 15, 2021).

Ferguson, C. J. (2008). The school shooting/violent video game link: causal relationship or moral panic? *J. Investig. Psychol. Offend. Profil.* 5, 25–37. doi: 10.1002/jip.76

Fiske, S. T., Cuddy, A. J. C., and Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends Cognit. Sci.* 11, 77–83. doi: 10.1016/j.tics.2006.11.005

Flowe, H. D. (2012). Do characteristics of faces that convey trustworthiness and dominance underlie perceptions of criminality? *PLOS ONE* 7:e37253. doi: 10.1371/journal.pone.0037253

Funk, F., and Todorov, A. (2013). Criminal stereotypes in the courtroom: facial tattoos affect guilt and punishment differently. *Psychol. Public Pol. Law* 19, 466–478. doi: 10.1037/a0034736

Goldstein, A. G., Chance, J. E., and Gilbert, B. (1984). Facial stereotypes of good guys and bad guys: a replication and extension. *Bull. Psychon. Soc.* 22, 549–552. doi: 10.3758/BF03333904

Grizzard, M., Fitzgerald, K., Francemone, C. J., Ahn, C., Huang, J., Walton, J., et al. (2019). Validating the extended character morality questionnaire. *Media Psychol.* 43, 107–130. doi: 10.1080/15213269.2019.1572523

Grizzard, M., Huang, J., Fitzgerald, K., Ahn, C., and Chu, H. (2018). Sensing heroes and villains: character-schema and the disposition formation process. *Commun. Res.* 45, 479–501. doi: 10.1177/0093650217699934

Haake, M., and Gulz, A. (2008). Visual stereotypes and virtual pedagogical agents. *J. duc. Technol. Soc.* 11, 1–15. Available online at: https://www.j-ets.net/collection/published-issues/11_4

Haidt, J., and Joseph, C. (2008). "The moral mind: how five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules," in *Evolution and Cognition. The Innate Mind*, Vol. 3, *Foundations and the Future*, eds P. Carruthers, S. Laurence, and S. Stich (Oxford: Oxford University Press), 367–391. doi: 10.1093/acprof:oso/9780195332834.003.0019

Happ, C., Melzer, A., and Steffgen, G. (2013). Superman vs. BAD man? The effects of empathy and game character in violent video games. *Cyberpsychol. Behav. Soc. Network.* 16, 774–778. doi: 10.1089/cyber.2012.0695

Hassenzahl, M. (2004). The interplay of beauty, goodness, and usability in interactive products. *Hum. Comput. Interact.* 19, 319–349. doi: 10.1207/s15327051hci1904_2

Hoerrner, K. L. (1996). Gender roles in disney films: analyzing behaviors from snow white to simba. *Womens Stud. Commun.* 19, 213–228. doi: 10.1080/07491409.1996.11089813

Hoffner, C., and Cantor, J. (1991). "Perceiving and responding to mass media characters," in *Responding to the Screen: Reception and Reaction Processes* (Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.), 63–101.

Ivory, J. D. (2006). Still a man's game: gender representation in online reviews of video games. *Mass Commun. Soc.* 9, 103–114. doi: 10.1207/s15327825mcs0901_6

Jacobsen, T. (2006). Bridging the arts and sciences: a framework for the psychology of aesthetics. *Leonardo*, 39, 155–162. doi: 10.1162/leon.2006.39.2.155

Joeckel, S., Bowman, N. D., and Dogruel, L. (2012). Gut or game? The influence of moral intuitions on decisions in video games. *Media Psychol.* 15, 460–485. doi: 10.1080/15213269.2012.727218

Johanson, C. (2019). Bride of Frankensystem. *Zenodo*. doi: 10.5281/zenodo.3544693

Kao, D., and Harrell, D. F. (2015). "Exploring the impact of role model avatars on game experience in educational games" in *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY'15* (London), 571–576. doi: 10.1145/2793107.2810291

Klapp, O. E. (1954). Heroes, villains and fools, as agents of social control. *Amer. Sociol. Rev.* 19, 56–62. doi: 10.2307/2088173

Klapper, J. T. (1960). *The Effects of Mass Communication*. Glencoe: Free Press.

Klastrup, L., and Tosca, S. (2009). "Because it just looks cool!" Fashion as character performance: the case of WoW. *J. Virtual Worlds Res.* 1, 3–17. doi: 10.4101/JVWR.V1I3.305

Kohlberg, L. (1971). Stages of moral development. *Moral Educ.* 1, 23–92.

Lane, R. (2017, February 20). What makes a great videogame villain? *PC Gamer*. Available online at: https://www.pcgamer.com/what-makes-a-great-videogame-villain/ (accessed March 15, 2021).

Leder, H., Belke, B., Oeberst, A., and Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *Brit. J. Psychol.* 95, 489–508. doi: 10.1348/0007126042369811

Lewis, A. R., Djupe, P. A., Mockabee, S. T., and Wu, J. S.-Y. (2015). The (non) religion of mechanical turk workers. *J. Scient. Study Relig.* 54, 419–428. doi: 10.1111/jssr.12184

Lindner, D., Trible, M., Pilato, I., and Ferguson, C. J. (2019). Examining the effects of exposure to a sexualized female video game protagonist on women's body image. *Psychol. Popul. Media Cult.* 553–560. doi: 10.1037/ppm0000251

Long, A. A., Sedley, D. N., and Sedley, D. M. (1987). *The Hellenistic Philosophers: Volume 1, Translations of the Principal Sources with Philosophical Commentary*. Cambridge: Cambridge University Press.

Lundwall, C. (2017). Creating *Guidelines* for *Game Character Designs*. Bachelor thesis in the Subject of Computer Graphics Arts, Luleå University of Technology.

MacLin, M. K., and Herrera, V. (2006). The criminal stereotype. *North Amer. J. Psychol.* 8, 197–208.

Manninen, T., and Kujanpää, T. (2007). The value of virtual assets: the role of game characters in MMOGs. *Int. J. Bus. Sci. Appl. Manage.* 2, 21–33.

Martin Del Campo, M. (2017). *Sympathy for the Devils: An Analysis of the Villain Archetype Since the Nineteenth Century*. Thesis, Texas AandM International University. Available online at: https://tamiu-ir.tdl.org/bitstream/handle/2152.4/115/MARTINDELCAMPO-THESIS-2017.pdf?sequence=1andisAllowed=y (accessed March 15, 2021).

McLaughlin, T. (2012). *Taxonomy of Digital Creatures: Interpreting Character Designs as Computer Graphics Techniques*. New York, NY. 3–10.

Mohd Tuah, N., Wanick, V., Ranchhod, A., and Wills, G. (2017). Exploring avatar roles for motivational effects in gameful environments. *EAI Endor. Trans. Creat. Technol.* 4:153055. doi: 10.4108/eai.4-9-2017.153055

Passmore, C. J., Birk, M. V., and Mandryk, R. L. (2018). "The privilege of immersion: racial and ethnic experiences, perceptions, and beliefs in digital gaming," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Vol. 383 (New York, NY), 1–383. doi: 10.1145/3173574.3173957

Przybylski, A. K., Rigby, C. S., and Ryan, R. M. (2010). A motivational model of video game engagement. *Rev. Gen. Psychol.* 14, 154–166. doi: 10.1037/a0019440

Raney, A. A. (2004). Expanding disposition theory: reconsidering character liking, moral evaluations, and enjoyment. *Commun. Theory* 14, 348–369. doi: 10.1111/j.1468-2885.2004.tb00319.x

Raney, A. A. (2006). "The psychology of disposition-based theories of media enjoyment," in *Psychology of Entertainment* (Hillsdale, NJ: Lawrence Erlbaum Associates Publishers), 137–150.

Raz, J. (1995). *Ethics in the Public Domain: Essays in the Morality of Law and Politics*. Oxford University Press. Available online at: https://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198260691.001.0001/acprof-9780198260691 (accessed March 15, 2021).

Reinecke, K., Yeh, T., Miratrix, L., Mardiko, R., Zhao, Y., Liu, J., et al. (2013). "Predicting users' first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI'13* (France). doi: 10.1145/2470654.2481281

Reinecke, L. (2009). Games and recovery: the use of video and computer games to recuperate from stress and strain. *J. Media Psychol.* 21, 126–142. doi: 10.1027/1864-1105.21.3.126

Sadr, J., Jarudi, I., and Sinha, P. (2003). The role of eyebrows in face recognition. *Perception* 32, 285–293. doi: 10.1068/p5027

Sanders, M. S., and Tsay-Vogel, M. (2016). Beyond heroes and villains: examining explanatory mechanisms underlying moral disengagement. *Mass Commun. Soc.* 19, 230–252. doi: 10.1080/15205436.2015.1096944

Schell, J. (2008). *The Art of Game Design: A Book of Lenses.* Amsterdam; Massachusetts, MA: Elsevier/Morgan Kaufmann.

Schwind, V., and Henze, N. (2018). "Gender- and age-related differences in designing the characteristics of stereotypical virtual faces," in *The Annual Symposium on Computer-Human Interaction in Play Extended Abstracts - CHI PLAY'18* (Melbourne, VIC), 463–475. doi: 10.1145/3242671.3242692

Schwind, V., Wolf, K., Henze, N., and Korn, O. (2015). "Determining the characteristics of preferred virtual faces using an avatar generator," in *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY'15* (London), 221–230. doi: 10.1145/2793107.2793116

Secord, P. F., Bevan, W. Jr., and Dukes, W. F. (1953). Occupational and physiognomic stereotypes in the perception of photographs. *J. Soc. Psychol.* 37, 261–270.

Tamborini, R. (2011). Moral intuition and media entertainment. *J. Media Psychol. Theor. Methods Appl.* 23, 39–45. doi: 10.1027/1864-1105/a000031

Tamborini, R., Eden, A., Bowman, N. D., Grizzard, M., Weber, R., and Lewis, R. J. (2013). Predicting media appeal from instinctive moral values. *Mass Commun. Soc.* 16, 325–346. doi: 10.1080/15205436.2012.703285

Tamborini, R., Grizzard, M., Bowman, N. D., Reinecke, L., Lewis, R. J., and Eden, A. (2011). Media enjoyment as need satisfaction: the contribution of hedonic and nonhedonic needs. *J. Commun.* 61, 1025–1042. doi: 10.1111/j.1460-2466.2011.01593.x

Tamborini, R., Prabhu, S., Lewis, R., Grizzard, M., and Eden, A. (2017). The influence of media exposure on the accessibility of moral intuitions and associated affect. *J. Media Psychol. Theor. Methods Appl.* 30, 79–90. doi: 10.1027/1864-1105/a000183

Tuch, A. N., Presslaber, E. E., Stöcklin, M., Opwis, K., and Bargas-Avila, J. A. (2012). The role of visual complexity and prototypicality regarding first impression of websites: working towards understanding aesthetic judgments. *Int. J. Hum. Comput. Stud.* 70, 794–811. doi: 10.1016/j.ijhcs.2012.06.003

Veronica, J. (2015). *Character Shape and Visceral Response: Can a Figure's Shape Alone Generate a Reaction Regarding its Nature?* Project, University of Skövde.

Vorderer, P., Hartmann, T., and Klimmt, C. (2003). "Explaining the enjoyment of playing video games: the role of competition," in *Proceedings of the Second International Conference on Entertainment Computing* (Pittsburgh, PA), 1–9.

Willis, J., and Todorov, A. (2006). First impressions: making up your mind after a 100-ms exposure to a face. *Psychol. Sci.* 17, 592–598. doi: 10.1111/j.1467-9280.2006.01750.x

Yarmey, A. D. (1993). Stereotypes and recognition memory for faces and voices of good guys and bad guys. *Appl. Cognit. Psychol.* 7, 419–431. doi: 10.1002/acp.2350070505

Yuki, M., Maddux, W. W., and Masuda, T. (2007). Are the windows to the soul the same in the East and West? Cultural differences in using the eyes and mouth as cues to recognize emotions in Japan and the United States. *J. Exp. Soc. Psychol.* 43, 303–311. doi: 10.1016/j.jesp.2006.02.004

Zillmann, D. (1996). "Sequential dependencies in emotional experience and behavior," in *Emotion: Interdisciplinary Perspectives* (New York, NY: Psychology Press), 243–272.

Zillmann, D. (1999). Exemplification theory: judging the whole by some of its parts. *Media Psychol.* 1, 69–94. doi: 10.1207/s1532785xmep0101_5

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership