Design and Instantiation of an Interactive Multidimensional Ontology for Game Design Elements – a Design and Behavioral Approach

Zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften

(Dr. rer. pol.)

von der
KIT-Fakultät für Wirtschaftswissenschaften
am Karlsruher Institut für Technologie (KIT)

genehmigte
DISSERTATION

von

Greta Luise Hoffmann (Dipl. Des.)

Tag der mündlichen Prüfung: 28.07.2022
Referent: Prof. Dr. Christof Weinhardt
Korreferent: Prof. Dr. Jella Pfeiffer

Karlsruhe, 2022

KARLSRUHE INSTITUTE OF TECHNOLOGY

Abstract

Department of Economics and Management
Institute of Information Systems and Marketing

Design and Instantiation of an Interactive Multidimensional Ontology for Game Design Elements – a Design and Behavioral Approach

by Greta Luise HOFFMANN

While games and play are commonly perceived as leisure tools, focus on the strategic implementation of isolated gameful elements outside of games has risen in recent years under the term gamification. Given their ease of implementation and impact in competitive games, a small set of game design elements, namely points, badges, and leaderboards, initially dominated research and practice. However, these elements reflect only a small group of components that game designers use to achieve positive outcomes in their systems. Current research has shifted towards focusing on the game design process instead of the isolated implementation of single elements under the term gameful design. But the problem of a tendency toward a monocultural selection of prominent design elements persists in-game and gameful design, preventing the method from reaching its full potential. This dissertation addresses this problem by designing and developing a digital, interactive game design element ontology that scholars and practitioners can use to make more informed and inspired decisions in creating gameful solutions to their problems.

The first part of this work is concerned with the collation and development of the digital ontology. First, two datasets were collated from game design and gamification literature (game design elements and playing motivations). Next, four explorative studies were conducted to add user-relevant metadata and connect their items into an ontological structure. The first two studies use card sorting to assess game theory frameworks regarding their suitability as foundational categories for the game design element dataset and to gain an overview of different viewpoints from which categorizations can be derived. The second set of studies builds on an explorative method of matching dataset entries via their descriptive keywords to arrive at a connected graph. The first of these studies connects items of the playing motivations dataset with themselves, while the second connects them with an additional dataset of human needs. The first part closes with the documentation of the design and development of the tool Kubun, reporting on the outcome of its evaluation via iterative expert interviews and a field study. The results suggest that the tool serves its preset goals of affording intuitive browsing for dedicated searches and serendipitous findings.

While the first part of this work reports on the top-down development process of the ontology and related navigation tool, the second part presents an in-depth research of specific learning-oriented game design elements to complement the overall research goal through a complementary bottom-up approach. Therein, two studies on learning-oriented game design elements are reported regarding their effect on performance, long-term learning outcome, and knowledge transfer. The studies are conducted with a game dedicated to teaching correct waste sorting. The first study focuses on a reward-based game design element in terms of its motivatory effect on perfect play. The second study evaluates two learning-enhancing game design elements, repeat, and look-up, in terms of their contribution to a long-term learning outcome. The comprehensive insights gained through the in-depth research manifest in the design of a module dedicated to reporting research outcomes in the ontology. The dissertation concludes with a discussion on the studies' varying limitations and an outlook on pathways for future research.

Acknowledgments

The first person I want to thank is my supervisor Prof. Dr. Christof Weinhardt, who gave me the chance to pursue my research endeavor despite my background as an outsider to the discipline and always supported me with most valuable advice, encouragement, and constructive criticism where necessary. I also want to express my deep gratitude to Prof. Dr. Jella Pfeiffer for her guidance, support, longstanding cooperation, and continuous goodwill, patience, and grace.

I also want to thank the other members of my Ph.D. committee: Prof. Dr. Gerhard Satzger, and Prof. Dr. Steffen Rebennack, for their time spent reviewing my thesis and their valuable comments and inspiring discussions before, during, and after my thesis defense.

The next set of thank you needs to be extended to my research colleagues, who welcomed me into their world, introduced me to essential foundational knowledge as well as pragmatic wisdom, supported me with all of my varying research efforts, and made me feel like part of the faculty family. I want to extend special shoutouts to Anke for her fantastic help in all the laboratory studies and her everlasting meticulousness, Wolfgang for his reading and correcting of various of my documents, his friendship and for introducing me to the Rocky Horror Show, Simon, David, and particularly Christian, for their additional supervision, encouragement, and advice, Marius for his support in the graph study and inspiring enthusiasm and dedication, Tom and Tim for warmly onboarding me at the beginning of my journey and Jonas for his help on the last miles.

I especially want to thank my family for supporting me throughout this whole process: My father for inspiring me to take up this challenge and suffering with me throughout the journey, my mother for a warm nest to escape to and loving gestures throughout the way, my sister Elsa for her open arms and always being a rock in times of need and my sister Mathilde for motivating me through her expressed admiration as well as her continuing work ethics.

I want to deeply thank my friend and collaborator Brice for his dedication to the goal of making the world a better place through meticulous and loving design and engineering, as well as his repeated and continuous support. I also want to thank my friends Madelaine for supporting me with her amazing language skills and Lea for lending her sharp mind to help on the mathematical front.

And finally, my partner Raphael, without whom this work could not exist, who has supported me throughout every step of the journey and whose competence, dedication to quality, sharp skepticism, and humor continue to inspire me.

Contents

List of Figures

List of Tables

List of Abbreviations

AfA	Amt für Abfallwirtschaft
AR	Availability Requirement
BE	Backend
BFI-10	Big Five Inventory- Ten Items
CEGE	Core Elements of the Gaming Experience
D&D	Dungeons and Dragons
DMTI	Decision Making Tendency Inventory
EF	Errorful Learning
EL	Errorless Learning
GA	Google Analytics
GBL	Game Based Learning
GDE	Game Design Element
HCI	Human Computer Interaction
HfG	Hochschule für Gestaltung Karlsruhe
HN	Human Needs
IGDB	Internet Game Database
IS	Information Systems
k.NN	k-Nearest-Neighbour
KIT	Karlsruhe Institute of Technology
LFR	Look and Feel Requirement
MDA	Mechanics, Dynamics, Aesthetics
MfP	Motivations for Play
MMORPG	Massively Multiplayer Online Role-Playing Game
MR	Maintainability Requirement
MS	Maximization Scale
NAM	Narrative, Aesthetics, Mechanics
PBL	Points, Badges, Leaderboards
PCA	Principal Component Analysis
PCT	Personal Construct Theory

PID	Preference for Intuition and Deliberation
PM	Playing Motivations
PR	Performance Requirement
REI	Rational Experiential Inventory
RR	Reliability Requirement
RQ	Research Question
SUS	System Usability Scale
SVG	Scalable Vector Graphics
TR	Training Requirement
UX	User Experience
VAC	Taxonomy of Gamification Elements for Varying Anticipated Commitment
W3C	World Wide Web Consortium

Part I

Fundamentals

1. Chapter 1
Introduction and Motivation

"That's what games are, in the end. Teachers. Fun is just another word for learning."

Raph Koster, A Theory of Fun for Game Design

1.1. Motivation

While games and play are commonly perceived as tools of leisure and entertainment, they have long been applied in other domains, like education (Cruickshank & Telfer, 1980; Epworth Bells Crowle and Isle of Axholme Messenger, 1874), medical contexts (Beierwaltes, 1985), and the military (J. Kim, 2013). In recent years, a new focus on the effectivity of gameful elements in contexts outside of games has risen under new terms, namely gamification, game-based learning, and gameful design. Multi-disciplinary research has long shown that games have innate mechanisms that facilitate learning, making them successful educational tools and supplements (Van Eck et al., 2017). As far back as 1988, Fileni's studies on the educational and cognitive aspects of video games concluded that games could help students develop their skills and improve their learning. Since then, studies looking at educational use-cases with and around games have shown that they can be used for active teaching (Travis, 2011), increasing and strengthening conceptual understanding (Klopfer, 2008), process skills and cognitive practices (Steinkuehler & Duncan, 2008), and for performance and progress assessment (Bellotti et al., 2013; Shute et al., 2009). Overall, research studies in the domain of serious games find that by applying game design to a real-life context, education can be effectively enhanced (Bellotti et al., 2013). Equally, research studies in the domain of game-based learning (GBL) have consistently shown the effectiveness of commercial and educational games in classroom settings (Pivec et al., 2003) as well as in higher education (Burmester, 2006). However, most of the research in this domain focuses around measuring the efficacy of full game implementations and less on single game elements dedicated to achieving specific desired outcomes.

By establishing the term "gamification" as the application of game elements to non-game contexts, Deterding et al. (2011) set a pronounced focus on differentiating between full implementations of topic-based games and the strategic addition of small gameful elements onto an already existing system or service. This laid the groundwork for the permeation of gameful thinking into the domain of Information Systems (IS). This process was later formalized by Blohm and Leimeister (2013) as choosing suitable "service bundles" that are added to the respective core offer to create comprehensive, IT-based, and increasingly ubiquitous enhancing services. Since then, the number of studies focusing on the relationships between single game elements and changes in outcome variables (e. g., performance, motivation, and learning) has seen a sizeable increase (Rapp et al., 2019). Also, research has spread into a wide spectrum of application domains, starting from education (e. g., computer programming (Fotaris et al., 2016) and science classes (Sanmugam et al., 2016)) across fitness (e. g., weight loss (Bojd et al., 2022) and mobile exercise (Jang et al., 2018)) and health (e. g. child anxiety treatment (Pramana et al., 2018), alcohol interventions (Boyle et al. 2017) and participatory sensing (Zentek & Hoffmann, 2016)), over economy and finance (e. g. sales applications (Carignan & Lawler Kennedy, 2013), e-banking (Luis Filipe Rodrigues et al., 2013) and financial education (Hoffmann & Matysiak, 2019)) to topics of sustainability (e. g. sustainable transport education (Putz et al., 2018) and domestic energy engagement (Gustafsson et al., 2009)). Especially regarding effects on learning outcome and motivation, these studies reported positive indications for the successful enhancement of both – individually as well as interdependently (Buckley & Doyle, 2017)).

Studies such as the aforementioned have found positive outcomes in settings evaluating the effects of entire games or gamefully enhanced systems and services, showing that even a little gameful design can already have a positive impact on users. This concurs with literature on game-based motivation, where e. g. Deterding (2016), in an analysis of studies on make-believe, finds that as little as verbally or visually framing an activity as "game," "play," or "fun" (vs. "work" or "obligation") has a positive effect on motivation and performance. A similar effect was also observed by Csikszentmihalyi (2000), who found that participants were more likely to have flow experiences when work was approached as play. However, while these developments have led to a growing number of studies on game design elements, no comprehensive collection of such outcome-oriented game elements has yet been collated. The lack of such a database results in time- and work-intense research processes for researchers and practitioners that are looking for suitable elements. This can lead to inferior designs of gameful artifacts if the search for better-suited solutions is aborted prematurely (Van Eck et al., 2017).

One contributing factor to the lack of a satisfactory overview of serviceable game design elements can be identified in the fall-out of the hype that surrounded the term gamification following its establishment in 2011 and peak in 2013 (Fenn & LeHong, 2011; LeHong & Fenn, 2013). A lot of research intermediately focused on a small set of recurring game design elements, namely points, badges, and leaderboards (El-Khuffash, 2013), resulting in the term PBL (Points, Badges, Leaderboards) being used almost synonymously to the encompassing term "gamification" (Chou, 2016). Due to their compact design and the resulting ease of implementation, these elements quickly became the representative go-to elements for gamification and rose to high levels of visibility, leading to the perpetuation of their use through follow-up and copy studies. This, however, limited the exploration of other elements of interest. The set of game elements that are currently in use in the entertainment industry is already large and diverse and is continuously increasing with each new development. When further considering the relative youth of gamification as a research field, the perceived limitation to only a small set of elements in such an early stage artificially restricts the generation of novel insights and obscures the true complexity of effectively enhancing a system. It is also important to note that given a different context or audience, the same game elements can produce contradictory or even adverse outcomes: It has been shown in later studies on PBL that, based on context, some of these elements had adverse effects on certain groups of participants (Toda et al., 2018). Subsequentially, the dangers of this kind of mono-focus have been critically observed (Chou, 2016; L. E. Nacke & Deterding, 2017), and research has since expanded toward a broader scope of design elements. This development is particularly beneficial as a meta-analysis on digital games and learning (Clark et al. 2016) found that those games that had augmented their game mechanics, as well as their visual and narrative elements through research-based value addition, afforded significantly better learning outcomes in contrast to games that were not associated with augmented game designs.

A study conducted by the association of the German games industry found a growing demand for gamification and game-based service enhancements (GAME 2019). However, while industry experts usually have the necessary knowledge over a broad spectrum of design possibilities as well as the necessary experience to choose suitable elements for a gameful implementation, the process can be difficult for laymen that lack the necessary insights on cognitive and game-design principles (Greitzer et al., 2007). If a practitioner is looking for a suitable element for their gamification purpose, they have to either already be familiar with the name or term it is referred to or at least have a specific vision of what exactly they are looking for to implement. This can be difficult as there is no standardized identification or naming process of game design elements. Lacking a comprehensive dictionary, game design elements and respective instructions for implementations cannot be looked up easily. From a research perspective, this is also relevant, as a more shared understanding could be gained and redundancy prevented if results and effects of elements were reported under unified terms and principles. Further contributing to a lack of deviation from already established elements is the additional difficulty of incorporating additional, potentially exclusionary factors of suitability into the search process, such as details on target audiences, target medium, cultural facets, or necessary resources for implementation. These points highlight the utility that an ontology of linked data on game design elements would offer to research and practice. Practitioners would benefit i) from being able to gain an overview by browsing a large range of options as well as ii) from being able to make informed decisions regarding suitable elements for their specific needs based on the additional information given on single elements. On the other hand, researchers would benefit by iii) being able to identify relevant research gaps and iv) having more efficient browsing options for finding obscure elements. As such, our first research goal is to *create an ontology that provides for the user needs of scholars and practitioners by identifying, collating, and connecting user-relevant data*.

There have been several research efforts by different groups of researchers that are working on game design element classifications and ontologies (most notable are the typology by Elverdam and Aarseth (2007) and the gameplay design pattern collection by Björk and Holopainen (2005)). However, the typology by Elverdam and Aarseth (2007) consists of a theoretical, abstract structure intended by the authors to be used and understood as a game design "grammar" and does not offer linked examples of game design elements within the derived categories. On the other hand, while the gameplay design pattern collection by Björk and Holopainen (2005) provides detailed insights into the usage and potential consequences of incorporating each element, it does not afford all utilities that a game design ontology could offer. First, the collated elements and patterns are excerpted from existing games; as such, they lack information on research outcomes and context-related effects concerning different user groups, contexts, and technologies. Second, while the gameplay design pattern collection is embedded in a digital Wiki (Björk, 2012), it lacks in terms of usability. Entries can be browsed by name via an overview page as well as through a categorical structure, however, this process presumes preexisting knowledge of games and context. It is therefore not yet optimized to support efficient research and design processes – especially for users that do not already have expert-level knowledge. This focus on game design experts is further reflected in the high frequency of use of insider terms and concepts that need to be understood before being able to fully contextualize single elements. Addressing these issues, our second research goal is to *systematically design and develop a digital, interactive interface for user-friendly navigation of the assembled ontology*. This includes the implementation and assessment of design components for quick and easy ontology navigation and the collection and integration of missing data relevant to users, e. g. targeted audience, targeted outcome, and examples from games. Once established, the ontology contributes to research and practice by affording much easier and quicker, need-centric access to existing information. Furthermore, by highlighting missing meta-data on already existing entries, it can be easier for game design element-based research to quickly identify and close existing research gaps. Finally, through data analytics on unsuccessful search queries, it can be possible to identify missing entries as well as alternative nominations, thus affording the systematic build-up of an exhaustive ontology and a set of common names.

Game and play are integral elements during early childhood development and growth and continue to play a big role in secondary education (Mayer, 2019), making education one of the biggest domains for gameful design. However, when looking at the existing collections of game design elements, we find a notable gap in elements or patterns related to learning. As most collections draw from the most successful and well-known entertainment games, elements from serious games and game-based learning applications are scarcely included in the aggregated datasets. While there are some classification efforts pertaining specifically to learning-oriented game design elements (Challco et al., 2014; O'Neil et al., 2005), they are kept on a very broad level and do not contain specific design elements, let alone additional information on existing research outcomes. To ensure that the design process of the ontology is not only conducted from a top-down approach but also backed up by a user-centric bottom-up view, we accompany the creation process of the ontology via a set of in-depth experimental studies that focus on the analysis of the effect of specific, learning-oriented game design elements on playing behavior and learning outcome. Our third research goal for this is to *add insights from a user-based perspective to the development process.*

This in-depth research is conducted in the topical domain of waste sorting, a subtopic in the domain of sustainability that is given little attention in terms of learning enhancing tools, despite a dire need for education (Luo et al., 2019). Recent studies show that global progress in waste reduction and recycling is slow, partly due to a lack of knowledge about what goes into which bin (Filho et al., 2016; Luo et al., 2019; Schultz et al., 1995; Thomas & Sharp, 2013). Many recycling and waste sorting facilities are as yet unable to reach maximum efficiency without pre-sorting measures (Bucciol et al., 2015; Hawlitschek, 2020), an issue that has been tackled in countries like Germany, Austria, and Switzerland by making domestic pre-sorting a citizen's responsibility (Buclet & Godard, 2000). However, incentivizing citizens to dispose of their household waste correctly and consistently continues to be a challenge for society, as it is a task that requires citizens to know how to fulfill the required task. In terms of learning, this topic domain is particularly challenging, as it is often perceived negatively, incentives are few, and outdated measures of communication and information like analog, paper-based flyers, and informational material still dominate as a means to educate the public (Luo et al., 2019). The object of analysis for this part of the research process is the waste sorting training game "Die Müll AG" (Engl. Name: "Trashmonsters") that we created in cooperation with the department of recycling and waste disposal Amt für Abfallwirtschaft (AfA) in Karlsruhe to incentivize the learning as well as real-life conduct of correct waste sorting (Clocher & Hoffmann, 2014). The studies' outcomes will be used to enhance the ontologies usability through the additional in-depth perspective of a researcher's perspective and needs. We further use this research to enrich the final ontology with additional, learning-specific elements.

1.2. Research Agenda and Research Questions

To systematically address the three goals outlined in the motivation section, I structured this research into two overarching segments, with respectively four and two research questions. In the first segment, Part II of this thesis, I use an explorative top-down approach to address the first two research goals pertaining to the development of a user-friendly, digital ontology that affords serendipitous findings as well as problem-specific solutions for a target audience of researchers as well as practitioners. Therein, the first three research questions address matters of classification and labeling as well as connecting the collated dataset of game design elements with related datasets of playing motivations and human needs to arrive at a satisfying ontological structure. In the fourth research question, I address matters of usability regarding the design and development of the digital interface for navigating ontological data. In Part III, I address the third research goal of refining the final ontology through insights from an in-depth, user-based perspective via two research questions relating to outcomes of specific learning enhancing design elements.

I start the ontological research process with the collation of a dataset of game design elements drawn from game design literature. To enhance this dataset towards a fully functional ontology, I start by exploring different methods to aggregate and evaluate user- and context-relevant metadata. Classification structures offer affordances for orientation and context regarding the similarity of the elements among themselves (Jacob, 2004). With regard to classifying game design elements, several research efforts have been made (Elverdam & Aarseth, 2007; Robinson & Bellotti, 2013; Tondello et al., 2016). However, these structures operate on an abstract level and lack experimental testing concerning their usability as labels for user-based searches. To ensure that user needs are reflected within the future structure of the ontology, I started the classification process by testing a selection of preexisting established frameworks by conducting a card-sorting experiment with a representative list of elements. This experiment followed the research question:

RQ1. *How compatible are established frameworks from game design theory with a dataset of diversely aggregated game design elements?*

Given a multi-parameter structure, different viable viewpoints can emerge for establishing a classification structure (Green, 1996). To gain insights into the viewpoints that practitioners would expect to find within a game design element ontology, I conducted a second explorative card sort as an open sort (no predetermined categories) with a group of game design experts. This was done to gain qualitative insights into underlying, intuitively derived category dimensions and gain insights into further viable classification viewpoints. The experiment was conducted to answer the following research question:

RQ2. *What categorial viewpoints of game design elements can be identified from an expert perspective?*

While the outcomes of such card sorts offer valuable metadata in terms of derived hierarchical structures and labels, they are costly in terms of time and effort. To arrive at additional, user-relevant metadata more efficiently, I explored options for automatically linking thematically adjacent datasets to the game design element ontology. Research on player types has shown that different genres of games address different target audiences (Tuunanen & Hamari, 2012). Through systematic clustering of distinct playing behaviors within the same game, studies have further shown that different player types can be linked to certain preferred mechanics or clusters of game design elements (Ferro et al., 2013). The underlying assumption of these studies is that personality is a strong influencing factor on which of the elements/mechanics lead to which behavioral outcomes (Busch et al., 2016). To enable researchers and practitioners to better consider their varying target audiences in their implementations, I collated additional datasets of playing motivations aggregated from player type and playing motivation literature, as well as a dataset of human needs drawn from psychological frameworks. To achieve an efficient and effective linking process of the game design element dataset with related datasets, I explore a keyword-based matching algorithm under the following research question:

RQ3. *How meaningfully can datasets be connected through algorithmic keyword-based matching?*

Following the aggregation process of the ontology, I focused on the second research goal – the development of a user-friendly interface for the navigation of the aggregated and enhanced datasets. I started with a systematic analysis of existing database navigation tools to identify requirements and gain an overview of existing solutions. While there exists a large number of database services for all diverse fields and topics, with many of them offering massive amounts of data, current tools do not afford quick and intuitive navigation of data (Stelmaszewska & Blandford, 2004), particularly lacking features that allow for low threshold navigation (Scott & O'Sullivan, 2005). An additional unsolved challenge lies in the design of visual interfaces that allow users to obtain a sensible overview of the spectrum of relevant items. Given the usability-related insufficiencies of the currently available technology, I conducted the design and development of the presented artifact based on the following, overarching research question:

RQ4. *What design factors afford user-friendly browsing processes for database navigation?*

In the second segment of my research (Part III of this thesis), I focus on refining the ontological structure by identifying further user needs that emerge from research practice.

During the aggregation process of the original dataset of game design elements, it became apparent that the foundational literature particularly lacked in terms of learning-outcome-oriented game design elements. Intending to add a user-based perspective to the development process, I conducted a set of studies on learning-enhancing game design elements. I chose the enhancement of learning outcome on correct waste sorting as the object of analysis due to its topical urgency (Filho et al., 2016). The learning data is embedded in a game where the core gameplay is to correctly sort incoming waste into the correct bin. In an interest in incentivizing perfect play to afford optimal learning, I first analyzed the effect of a "perfect-reward" game design element on playing performance (the number of waste items correctly sorted into their respective waste bin during a playthrough). The research question underlying this research study is:

RQ5. *What effect does a reward for perfection have on playing performance?*

Apart from perfect play, learning outcomes can be enhanced through different cognitive learning strategies (Friedrich & Mandl, 2006). Particularly the overall learning benefits of repetition are well documented across different learning domains (Ahmadian, 2012; Bygate, 1996). However, if composed as a design element, the inherent tediousness of repetition could negatively interfere with the game fun. Thus, it is important to gain deeper insights into the potential risks and benefits of including such a design element. This could be alleviated by the addition of an index-based element, where the correct answer can be looked up penalty-free during the core gameplay (look-up element). A literature review on expectable learning outcomes on such an index-based design element showed that results were contradictory (e. g. Größler et al., 2000), dependent on learners' prior knowledge (Hirsch Jr., 2000; Miller & Gildea, 1987) or not directly translatable to the context of the game. To better understand expectable outcomes of the addition of one or both of such elements on learning outcome, an experiment was designed and conducted to answer the following research question:

RQ6. *What effect does a repetition-based and a look-up-based game design element have on long-term learning of correct sorting of waste items into their target bin?*

Given the explorative, constructive nature of our overarching research goals, we employ a variety of methods for our various studies. With its explorative, outcome-oriented focus, Part II of this work is conducted using qualitative, experimental research methodologies: RQ1 and RQ2 are evaluated using the qualitative, user-focused methodology of card sorting, while RQ3 follows an algorithmic keyword-matching approach that we evaluate via graph analysis. For the development of the artifact, I follow a human design process and evaluate RQ4 through semi-structured interviews and a field test. In contrast, for the in-depth research in Part III, RQ5 and RQ6 were answered using quantitative empirical research in the form of laboratory and online experiments.

1.3. Structure of the Dissertation

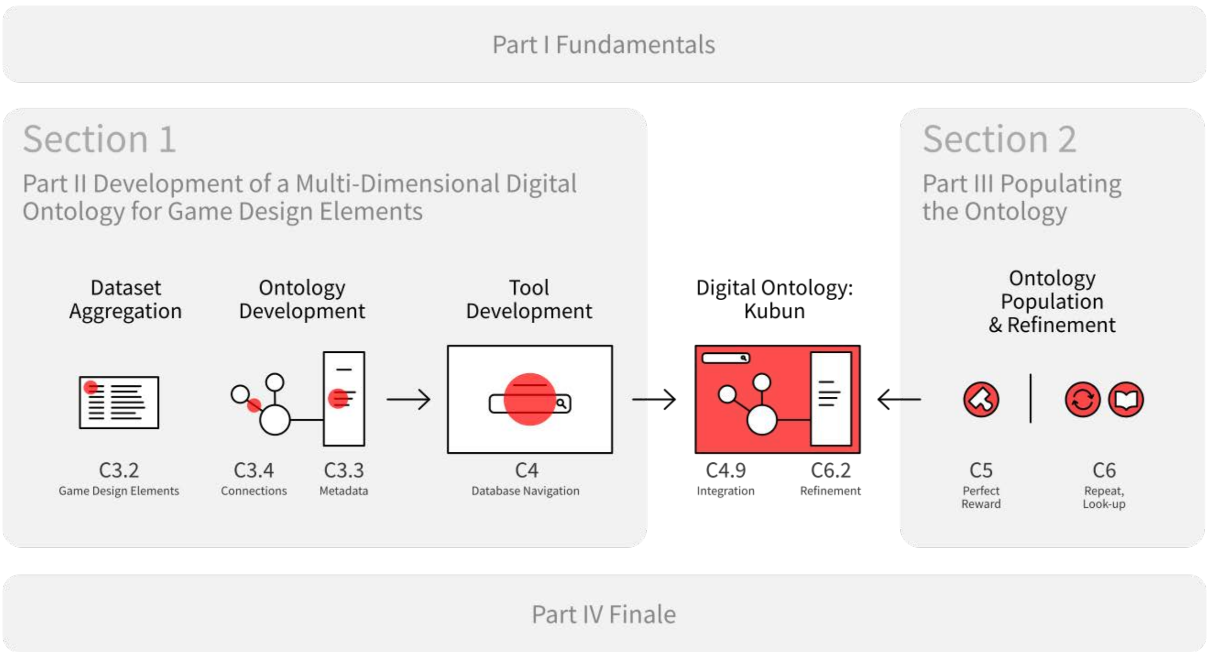


Figure 1 – Overview Structure Dissertation

This thesis is split into four main parts. Part I consists of two chapters, Chapter 1, where I present the motivation, research agenda, and structure. Following this introductory chapter, I lay out elementary terms and definitions and the theoretical foundations underlying this work in Chapter 2. In Chapter 2.1, I establish a definition for the terms game design element and playing motivation for the contexts of this work and elaborate on the choice of using the expression "gameful design" in contrast to "gamification." Following this, in Chapter 2.2, I introduce the theoretical foundations this work is built upon - game design theory, human-computer interaction, cognitive sciences, and classifications.

In Part II of this work, I present the systematic research and development process of the digital, multi-dimensional game design element ontology. In Chapter 3, I present the aggregation of the ontology as well as four explorative studies conducted for its enhancement in terms of metadata. After the introduction, I document the preparations for the subsequent studies – the dataset aggregations of game design and gamification elements as well as the player types and playing motivations in Chapter 3.2. In Chapter 3.3, I present the development process of a tool for digital card sorting as well as two card sort studies. In the first study, I use a closed card sorting process to map representative items from the dataset with prominent game design classifications from literature. In the second study, I build on expert knowledge to conduct a multi-dimensional categorization through an open card sort. Finally, in Chapter 3.4, I present two studies in which I explore the possibility of enhancing the dataset by connecting it with adjacent datasets via an experimental keyword matching. In the first study, I conduct an internal keyword matching on a dataset of playing motivations and player types to gain an understanding of the layout and distribution of the emerging graph. In the second study, I synthesize the playing motivations dataset with an additional dataset of human needs to research the internal connections between playing motivations and their links to external concepts/needs. The generated insights from the studies presented in Chapter 3 culminate in an enriched multi-dimensional ontology of game design elements.

Chapter 4 presents the design and development of the digital tool for the navigation of the collated ontology, starting with a requirement and market analysis, continuing with the design and development process, and ending with the evaluation through user tests and a field study. In the final chapter of Part II (Chapter 4.9), I document the final steps for integrating the ontology into the developed tool and present the final product.

In Part III of this work, I present a set of supplemental in-depth studies. This part of my research focuses on the view of the prospective user of the designed ontology. To this end, I conducted two studies with a specific focus on playing behaviors and game design elements that foster learning. In Chapter 5, I start by presenting additional foundations about the specific topic of waste sorting and learning as well as outlining the game artifact and its design rationale (Chapters 5.1, 5.2, 5.3). Following these foundational chapters, I present the two studies where I analyze three game elements with regards to motivational and learning outcomes: in Chapters 5.4, 5.5 and I present a pre-study and an experiment where I assess a perfect reward element with regard to its effectivity in motivating perfect play and analyze related playing behaviors. Chapter 5.6 presents the results of a post-study, where I conduct an explorative evaluation of data gathered from a field study to gain additional insights into playing behaviors relating to two learning-enhancing game design elements: repeat and look-up. Chapter 6.1 I present the design, procedure, and results of the second laboratory experiment where I measure the effects of these two game design elements (repetition and look-up) concerning learning outcome in comparison to common teaching materials on waste sorting. The overall findings and insights of these in-depth studies toward the game design element ontology are discussed in Chapter 6.2, concluding Part III.

Finally, the thesis concludes with Part IV, comprising Chapter 7, where I present conclusions, limitations, and promising avenues for future research. An overview of the structure of this thesis can be seen in Figure 1.

2. Chapter 2
Theoretical Foundations

"Work consists of whatever a body is obliged to do, and [. .].
Play consists of whatever a body is not obliged to do."

Mark Twain, The Adventures of Tom Sawyer

2.1. Elementary Terms and Definitions

In the following chapter, we present a discussion and derivation of the terms and definitions relevant to the contents of this work.

2.1.1. Game Design Element

For building an ontology of game design elements, we start by establishing a working definition of the term. Over time and domains, different definitions of the term game (design) element have emerged, differing in their wording, meaning, and implications depending on their research context. In their mapping study on gamification, Dichev et al. (2015) use the example of the game design element "badge" to highlight how it is referred to in different terms across domains and contexts: *an interface design pattern* (Deterding et al., 2011), a *game mechanic* (Zichermann & Cunningham, 2011), a *game dynamic* (Iosup & Epema, 2014), a *motivational affordance* (Hamari et al., 2014), and a *game component* (a specific instantiation of mechanics or dynamics) (Werbach & Hunter, 2012). Other authors use the term "game design element" to classify design elements into aesthetics, dynamics, and mechanics). In the light of the varying terms and definitions across the different fields of research, the following chapter will present the derivation process we used to extract a nominal definition that best serves the context of this work.[1]

In 2011, Deterding et al. published a defining manuscript on the rising trend-term gamification, which they define as "*the use of game design elements in non-game contexts*" (Deterding, 2011, p.10). They describe the boundaries of each hyponym of their definition (game, element, design, and non-game context) in detail. They further approach their definitory process for the compound "game design element," establishing a reasonable middle ground between "any element that can be found in any game" and "game elements being unique or specific to games" as the former is too restrictive, and the latter lacks sensible boundaries. They conclude with two definitions for game design elements: "*game elements as building blocks or features shared by games*" (Deterding 2011, p.12) and *"elements that are characteristic to games"* (Deterding, 2011, p.12). These two definitions can be broken down into two abstract parts relating to the component (element) and its relational origin (games): a *container* that holds *features characteristic of games*.

[1] The graphics underlying the definitions were produced by us as visual translations of our understanding of the definitions in the context of this work (see Figures 2-9)

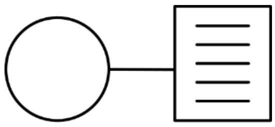


Figure 2 – Game Design Element, Container and Features (*Deterding et al. 2011*)

However, in their manuscript, an additional chart of examples for levels of game design elements expands the scope of their definition: *game interfaces* (e. g., dialogues, menus), *design patterns* (e. g., badge, leaderboard, level), *game design patterns* and *mechanics* (e. g. time constraint, limited resources, turns), *game design principles* and *heuristics* (e. g. enduring play, clear goals, variety of game styles), *game models* (MDA; challenge, fantasy, curiosity; game design atoms; CEGE), *game design methods* (e. g. playtesting, play centric & value conscious game design). As these examples show, the authors use the term element to describe processes surrounding game design practice (e. g., game design methods), motivations (e. g. see, game models), as much as components (e. g., game design patterns).

A definition for "*game elements*" given by Ring (2013) adds further nuances to the term:

"A game element is a characteristic that is contributing to the game feeling and is frequently occurring in games. Individual game design elements are however neither necessary nor sufficient to create a game on their own." (Ring, 2013, p.51).

This definition adds two new facets. By introducing the term "game feeling," a formerly unmentioned aspect of the element construct is highlighted: the receiving entity of the game design element (*the playing entity*) and the underlying goal to affect their emotional state.

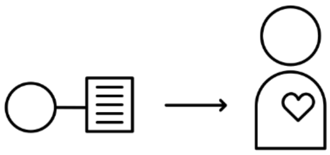


Figure 3 – Game Design Element (*Ring 2013*)

Another facet of Ring's definition is the distinction of a game (design) element as a component that can only exist as a part of the whole, not as the whole itself. However, we would argue that this second facet is refutable because games are occasionally used as game design elements in the form of "minigames" (e. g. in the game "Machinarium" (Amanita Design, 2009). These games typically feature different gameplay from the game they are embedded in and are placed within the larger context of the game. Thus, individual game design elements can consist of fully functional games in themselves and still be used as components for games, and as such, we will dismiss this second facet for our understanding of the term.

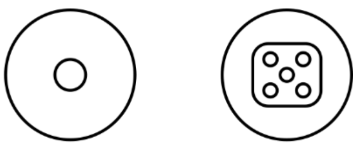


Figure 4 – Game Design Element: Being a Component vs. a Whole (*Ring 2013*)

However, while full games can serve as game design elements in specific contexts, there are fundamental components of game elements ("the atoms of games" (Koster, 2005)) that are not functional as game (design) elements by themselves but smaller, isolated concepts or ideas. They are called *ludemes*: "*a ludeme or "ludic meme" is a fundamental unit of play, often equivalent to a "rule" of play; the conceptual equivalent of a material component of a game.*" (Parlett 2016, p.81).[2] We use this distinction to separate our understanding of a game (design) element as a functional unit to be added to a system composed of the isolated abstract concepts represented by ludemes.

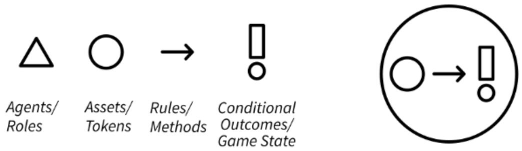


Figure 5 – Ludemes vs. Game Design Element (*Parlett 2016*)

Given the variety mentioned above of terms referring to similar or specified meanings of "game design element," we want to explore some of the most popular terms used in adjacent domains. Publications in game design practice commonly use the term "*game mechanic*" (Adams & Dormans, 2012; Zichermann & Cunningham, 2011). According to Hunicke et al. (2004), this term lacks a single and precise definition. Contributing to this might stem from differences in perspective when it comes to affected components: while referring to the same result, definitions vary in terminology when given from a computer science perspective:

"[Game mechanics are] methods invoked by agents, designed for interaction with the game state" (Sicart, 2008, p.7)

or a design perspective:

"[game mechanics are] semantically viable (that is, meaningful) combinations of tokens and rules." (Sellers, 2017, p.101)

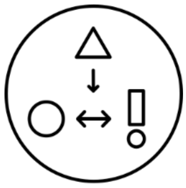


Figure 6 – Game Mechanic
(*Sicart 2008, Sellers 2017*)

Overall, we conclude while differing in terminology, the term game mechanic is not logically different from the established definitions for game design elements. However, we will continue to use the latter, as it is the established term in gamification/gameful design literature.

Inspired by the method of design patterns (Alexander, 1977), another similar term in use is "*game design pattern*." While originally ideated by Kreimeier (2002), the concept was elaborated on and publicized by Björk & Holopainen (2005) within the research effort to create a taxonomy of game design patterns. While

[2] „For example, whereas the material piece shaped like a horse and designated "knight" is a component of the game, the distinctively skewed move of a knight is a ludeme of the class "rule of movement". But other types of ludemes also exist. For example, the name, referend and associated connotations of "knight" - those of a chivalric courtier - may be said to constitute a thematic ludeme." (Parlett 2016, p.83)

the original definition of design patterns as collections of reusable solutions to solve recurring problems strongly leans on their purpose as problem-solving tools (Alexander 1977), Björk and Holopainen (2005) distance themselves from this problem-oriented focus and define them as tools to support creative design work. While only structurally defining their patterns in their first work (name, description, consequences, using the pattern, relations), they add a definition in their later work:

"semi-formal inter-dependent descriptions of commonly reoccurring parts of the design of a game that concern gameplay" (Olsson et al. 2014, p.2)

Their definition shows a close relation to the game mechanics definition by Sellers (2017), as well as game design elements (as defined by Deterding et al., (2011)). Still, it adds the facet of the interdependency of the respective elements.

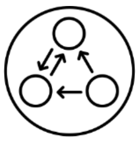


Figure 7 – Game Design Patterns
(*Björk and Holopainen 2005*)

In the domain of IS, we can also find the term "*gamification affordance*" used equivalently to the term game design element (Hamari et al., 2014). Originating from the field of psychology and perception, the term "*affordance*" was established by Gibson (1977) as actionable properties between an object and an actor and popularized by Norman (2013) as *"... the perceived and actual properties of the thing, primarily those fundamental properties that determine just how the thing could possibly be used."* (Norman 2013, p.9). In contrast to the neutral term "element," affordance refers explicitly to an inherent indication of its usage intentions and possibilities. Thus, while closely related, the term affordance is a specific type of element that is actionable as well as recognizable in terms of their intended use. For our purposes, a more open definition that includes passive, as well as actionable properties (e. g. a game-feel like the anticipation of incoming horror), is desirable.

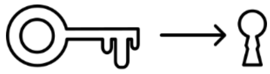


Figure 8 – Game Design Patterns
(*Björk and Holopainen 2005*)

Looking at the analyzed synonymic terms for game design elements, we find the words "game" and "element" synonymously represented, but the term design is mostly absent. However, with the purpose of our ontology to provide game design elements to be applied within a design process, the term design represents a central component to us. Thus, in a final step, we look at definitions for "game design" to gain additional insights into this facet of the term.

A short but concise definition for game design, *"...the act of deciding what a game should be"*, is provided by Schell (2014). While this definition highlights the decision-making facet of the design process, implying a choice process from preexisting elements, Adams uses more accessible terms: "*At its most elementary level, game design consists of inventing and documenting the elements of a game*." He expands this by stating: "*However, games don't exist in a vacuum; people create them to serve a purpose*." (Adams and Dormans 2012, p.18). This focus on the user can also be found in Salen and Zimmerman's (2004) Rules of

Play: *"Design is the process by which a designer creates a context to be encountered by a participant, from which meaning emerges."* (Salen and Zimmerman, 2004, p.2)

Taken together, these definitions highlight the broader context in which game design elements are embedded: the designing entity, the encountering (playing) entity, and the system that incorporates and connects the game design elements in between.

Building on our insights from analyzing prevalent definitions, we derive the following definition for the term game design element as we use it in the context of our ontology:

A component derived from contexts inherent to or associated with games that affords a designing entity to achieve desired outcomes within a receiving entity through interaction with a medium changed by the same component.

Figure 9 – Game Design Element (*Hoffmann 2022*)

2.1.2. Playing Motivations

As the desired outcomes of the process of gameful design are supposed to happen within the users, it is elementary to understand the intrinsic motivations for play that underly different game design elements. Playing motivations are useful labels if the underlying needs of the target audience are known. As stated by (Koster, 2005): "Different games appeal to different personality types, and not just because particular problems appeal to certain brain types. It's also because particular solutions appeal to particular brain types, and when we've got a good thing going, we're not likely to change it." Researchers have taken different approaches to cluster audiences and identify commonalities between preferences in various types of people. One sub-field in game studies concerned with such player segmentation is player-type research, where users are clustered according to their preferences of what they want from or enjoy in their playing experience. While no standard definition for a player type has been established ("*Player types are not a defined concept and any categorization of players or users needs to occur within the context of a particular application or domain*." Dixon (2011), p. 1,2), the consensus is that the clusters of playing behaviors that are referred to as player types emerge from the design decisions within the game with how they are being received and used by the players. Dixon further elaborates: *"The idea of Player Types assumes that there are distinct player-related phenomena that can be categorized, for example: motivations, play styles, behaviors, genre preferences and pleasures."* Dixon (2011), p. 1,2

As we are less interested in specific player clusters but in extrapolating the intrinsic motivations for play that game design elements hold, we focus on the adjacent term *playing motivations*. By extending a nominal definition of the term *incentive:* "something that encourages a person to do something," to game design elements, for our research, we define playing motivations as:

those qualities of a game design element that encourage a person to interact with it and the broader contexts it is embedded in.

2.1.3. Gamification vs. Gameful Design

While we established definitions of game design elements and playing motivations in the former chapters, this chapter focuses on the contexts in which they are applied. In IS, the understanding of the term game design element is influenced by the context in which it was first formalized: the foundational building blocks for gamification processes (Deterding et al., 2011). Looking at definitions of the term gamification, however, some limitations can be identified regarding the use and application of game design elements. Also, while mostly used synonymously, the term gameful design has since emerged as the more accepted term within the associated communities. This is partly due to some controversial application contexts in which the term gamification arrived at a negative connotation (Hung, 2017), but also because more emphasis is put on the design aspect of the application process. The following chapter expands on a rationale for using the term gameful design instead of gamification.

While coined as early as 2008 by Bret Terril, the term gamification (at that time, "gameification") was first formally defined and successfully established in IS within the well-cited publication of Deterding et al. (2011) as

"Gamification" refers to the use (rather than the extension) of design (rather than game-based technology or other gamerelated practices) elements (rather than full-fledged games) characteristic for games (rather than play or playfulness) in non-game contexts (regardless of specific usage intentions, contexts, or media of implementation).

Following definitions within IS add a focus on gamification as a value-enhancing service, as can be seen in the definition offered by Blohm & Leimeister (2013) as *"gamification encompasses the design of "gamified" service bundles*[3]*"* and the definition by Huotari (2012): *"a process of enhancing a service with affordances for gameful experiences in order to support user's overall value creation."*

This outlook on desirable outcomes can also be found in earlier definitions that, while not using the term gamification, refer to the same underlying processes under a different terminology:

"...we view game design as one aspect in the company's marketing process that aims to create demand for virtual goods that can be sold for real money." (Hamari & Lehdonvirta, 2010, p.15)

This focus on usage outcome highlights an essential facet of the gamification process, and thus the internal structure game design elements are made of – the *goals* to which they are applied. However, the presented definitions also carry an implied or explicated focus on (short-term) economic value creation. For the context of our research, we want to distance ourselves from strongly associating with specific research disciplines and application domains, as that would be too limiting for the purposes and the overall benefits a game design element ontology can serve. Second, after the initial hype of the term following the publication of Deterding et al. (2011), usage of the term declined. Looking at google search queries for gamification (Google Trends, 2022), we find that from 2010 to 2012, interest in the subject of gamification first increased steeply; however, it has since then been on a consistent, albeit slow decline. Also, while appearing in the Hype Cycle of the Gartner Institute in 2011 (Fenn & LeHong, 2011), in 2014, it was already placed into the Trough

[3] Ein Leistungsbündel ist „eine Leistung, die aus mehreren Teilen besteht, welche nicht mehr ohne Weiteres einzeln erkennbar sind, deren unterschiedliche Eigenschaften aber das hybride Produkt prägen." (Leimeister & Glauner 2008, S. 248)

of Disillusionment (Lowendahl, 2014) and has since then been absent from the Cycle. Different factors have contributed to this decline. Also, the focus on the outcome of user behavior intrinsic to some of the definitions was questioned in terms of ethicality (Marczewski, 2017; Sicart, 2015). Negative perception was further amplified by the Chinese government's implementation of a gamified point system for rating and rewarding their citizens to incentivize party loyalty and disincentivize association with critical elements of society (dos Reis & Press, 2019). Finally, a backlash against the term came from the game design community that criticized the perceived simplification and banalization of a complex design process for quick benefits (Bogost, 2011b, 2011a; M. Robertson, 2010). Most studies in the context of gamification have only assessed beneficial outcomes in the form of short-term outcomes (Nacke & Deterding, 2017); thus, gamification (specifically the application of points, badges, and leaderboards) has come to be perceived as a "quick fix" to user motivation rather than a sustainable solution (Chou, 2016; Toda et al., 2018). To summarize, while the term gamification was effective in transporting the idea of building on strategies and mechanics from game design practice into different business domains, implied promises of high gains for little input did not yet live up to the hype.

To distance themselves from the connotations mentioned above, researchers and practitioners have started to refer to the process underlying gamification under the more neutral term ***gameful design***. In an article on Make-Belief, Deterding (2016) gives the following definition: "Gameful design is defined by the end of affording ludic qualities or gamefulness (the experiential qualities characteristic for gameplay) in nongame contexts." (Deterding, 2016, p.105). The relevance of word-based framing is further highlighted in the same manuscript as they report that several studies have found that verbally or visually framing an activity as "game," "play," or "fun" (vs. "work" or "obligation") has a positive effect on motivation and performance (Birk et al., 2015; Laran & Janiszewski, 2011; Lieberoth, 2015; Littleton et al., 1999; J. Webster & Martocchio, 1993). Also, participants were likelier to have optimal or flow experiences when work was approached as play (Csikszentmihalyi, 2000). These findings indicate that the mindset created within the user is used towards achieving desirable outcomes (e. g. in terms of learning, motivation, and attitude change). In our work, we want to relate to game design and the positive influence it can have overall, rather than limiting ourselves to a process of achieving desirable outcomes in economic contexts. As such, we will refer to the context of our research as gameful design rather than gamification.

2.2. Theoretic Background and Core Theories

This work aims to provide game design practitioners and scholars of novice or expert level with a digitally useable toolset to better afford their audiences the desired experience. In the following chapter, we introduce our foundational theories and design principles relevant to the development and refinement process of an interconnected ontology of game design elements as well as a digital tool for database navigation. With regards to the ontologies' content, our research is founded on game design theory (see Chapter 2.2.1). For the development of the database navigation tool, we follow a human-centered design process, building on best practices informed by human-computer interaction (see Chapter 2.2.2) and theories from cognitive science (see Chapter 2.2.3).

Given the differences in the methodologies we apply within our different studies, we placed specific methodological elaborations within the respective chapters (e. g., card sorts for labeling are presented in Chapter 3.3.1, and keyword matching for ontological linking is presented in Chapter 3.4.1). For

2.2.1. Game Design Foundations

The goal of game design, used towards a full game product (game) or in a fragmented way (gamification), is to afford play and its accompanying benefits. A fundamental concept of play is the establishment of a separation between the pretended reality from real life. This abstract border is commonly referred to as the *magic circle* (as established by Huizinga in his foundational work on games and play "Homo Ludens" (1956)). The circle marks the cognitive and emotional separation from the surrounding real-life situation, a state that in media theory is referred to as the *willing suspension of disbelief* (Ferri, 2007). If a person plays by themselves, the magic circle is established as soon as the person decides to start playing. However, as soon as more than one person is involved, conventions need to be established in which the pretend play happens between all participants at the same time. These conventions are based on rules, which turn Paidia (the free play) into Ludus (rule-based play). One of the main tasks of game design is to establish the necessary number of rules to afford smooth play between all parties while allowing for enough freedom to let play happen. Another component of stimulating play is the assignment of *artificial significance* toward situations that outside of the game can be seen as trivial or irrelevant (like the location of a ball in relation to a goalpost). Through the same mechanism, significance can also be pushed away from a situation (like discounting a loss: "it was just a game"). This assignment and rejection of significance is part of the pretending component of play and, with it, the freedom and ease associated with it (Adams, 2014).

While these fundamental concepts are inherent to the act of playing and happen between the playing entity and the system, the role of the designer is to *predefine the realm within the magic circle* in terms of *time and space* (Huizinga, 1956). The act of designing this realm is structured by game design literature and practice through a set of foundational components and facets. The overall structure of the playing process is shaped by *rules*. The rules shape the outline of the game (the borders of the magic circle) in terms of time e. g. via the *start* and *termination condition* (the condition that ends the game) and in terms of space via the definition of a *game space* (like the chessboard, a football field, or the space allocated on the hard drive for a particular computer game). They also shape time and space within the game, e. g., in terms of time through *duration conditions* and in terms of space e. g. through objects that players interact with within the context of the game (like the ball and goalposts in football or the dice in a game of chance).

The rules further define the *semiotics* of the game and are understood as the meanings and relationships of the symbols that the game employs (Adams, 2014). Most importantly, the rules shape the player interactions within the game - the *gameplay* (the defining interactions players have with and through the game, also defined as "interaction that entertains" (Dini, 2012, p. 31). A driving force for gameplay is generated through *goals* set for the players to be achieved. Adams (2014) separates rules related to the shaping of gameplay into *challenges*: "a challenge is any task set for the player that is nontrivial to accomplish." (Adams, 2014, p.10) and *actions*: "the rules specify what actions the players may take to overcome the challenges and achieve the goal of the game." (Adams, 2014, p.11). While challenges are used as incentives to shape the directions player actions can take, actions are used to limit the pathways players can use to succeed within the challenges and achieve the goal.

Apart from the rules shaping the outlines and insides of the magic circle, there exists a set of rules outside of that: *metarules* – the rules that shape the rules themselves. Among other things, they shape the circumstances under which the existing rules can be changed and which kind of exceptions are allowed. This particular type of rule is indicative of an essential part of the design process: *balancing*. Balancing is the act of adjusting the parameters that are inherent to any element and rule within the game and is central to affording the internal *fairness* of the game. While difficult to pinpoint, as it is subject to societal, cultural, and even personal perception, fairness is one of the most important global parameters of a game – particularly of those that focus on competition as a central element. If not provided satisfactorily, it can happen that "players sometimes spontaneously decide to change the rules of a game during play if they perceive that the rules are unfair or that the rules are permitting unfair behavior." (Adams, 2014, p.12). Balancing happens on all levels of the design process – be it in the global decision of how to determine the starting entity in an asymmetric game (like tic tac toe) or the damage value a specific weapon causes on different entities depending on their armor value (as in roleplaying games like D&D). It is essential for achieving and maintaining the state of *immersion* ("the feeling of being submerged in a form of entertainment, or rather, being unaware that you are experiencing an artificial world." (Adams, 2014, p.25), a state that most players seek from or during the process of playing. Ensuring the maintenance of this state is particularly important as *immersion breakers,* and the involuntary expulsion of the players out of the magic circle is perceived as jarring and disappointing (Brian Moriarty, 1997). This coherence has become especially relevant in research on virtual environments (Peukert et al., 2019).

Apart from providing solidly balanced gameplay, immersion can be further strengthened through the presentation of a *game world*. While not strictly necessary for play, most games provide such a metaphorical context to ease the entry of the magic circle (like chess using the metaphor of two competing courts). According to Adams (2014), the reception of the game world (as well as the game itself) is linked to two central design factors: *aesthetics* ("[the game being] designed with a sense of style and created with artistic skill" (Adams, 2014, p.21)) and *harmony* ("the feeling that all parts of the game belong to a single, coherent whole" (Adams, 2014, p.21) – a quality first formalized by the game designer Brian Moriarty, (1997)). Additional components include storytelling, risks and rewards, novelty, mastery, creative and expressive play, and socializing.

The aforementioned factors are summarized by Adams (2014) as general elements of game design. Looking at the game system in relation to its interaction with the player, Adams separates two application/implementation layers of the game system: the *core mechanics,* "a symbolic and mathematical model that can be implemented algorithmically" (Adams, 2014, p.35), and the *user interface* that "mediates between the core mechanics of the game and the player" (Adams, 2014, p.37). A similar structure is presented in the MDA (mechanics, dynamics, aesthetics) framework of Hunicke et al. (2004), where mechanics refers to what is happening within the game, dynamics refers to the interaction between the game, and the player and aesthetics refers to the representation of the game with the player. Within the interplay of the core mechanics and the user interface emerges a concept related to the player's interaction - the *gameplay mode*: "the particular subset of a game's total gameplay that is available at any one time in the game, plus the user interface that presents that subset of the gameplay to the player" (Adams, 2014, p.40). This structure operates on the same structural level as the game design patterns established by (Björk & Holopainen, 2005). For most games, the *core* or *primary gameplay mode* (also referred to by other authors as *core gameplay* (Fabricatore, 2007) or *gameplay loop* (Guardiola, 2016)) is typically distinguished from the others as: "the mode in which the player spends the majority of his time" (Adams 2014, p.48). Outside of the context of games, it is this activity that needs to be identified and then treated as if it was the core gameplay for gameful design to be successfully applied. For this it is important that the core activity that is to be gamified is understood and presented as a desirable interaction while ensuring that its intended effect is maintained.

Finally, in terms of the overall design and development process, Adams (2014) distinguishes between three phases: the *concept stage*, which is performed first and whose results do not change; the *elaboration stage*, in which the primary gameplay mode is defined and potentially supplemented by additional modes, most of the design details are added, and decisions are redefined through prototyping and playtesting and the *tuning stage*, at which point no new features may be added, small adjustments can be made to polish the game. Adams particularly notes that "polishing is a subtractive process, not an additive one."[4] While the polishing of aesthetic factors mostly happens on the user interface layer and the balancing process is mostly conducted within the core mechanics, the core mechanics can also be aesthetically improved e. g. through ensuring a satisfying symmetry relating to the underlying metrics and the user interface can and should be balanced, e. g. in terms of cognitive overload. For achieving an overall harmony of the system, choices made on both levels (the model as well as the surface) need to be compared and synchronized. If these three factors are polished to their highest level, the final artifact will achieve an overall quality best described as *elegant* ("elegance is the sign of craftsmanship of the highest order" p.30). It is the highest goal of the game(ful) design process to achieve this level of quality as it is linked to sustainable and reproducible fun.

2.2.2. Human-Computer Interaction Foundations

The domain of human-computer interaction is concerned with design choices built on computational conditions and human habits. The ISO Norm definition for HCI is: "The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (ISO, 1998). Its close connection to design practice is reflected in the perspective of Hassenzahl, who describes the discipline as primarily "concerned with the making of things" (Hassenzahl, 2010, p.6). The HCI process includes the initial design as well as the following adjustments and iterations of all single elements that influence the intended interaction. In terms of the human perspective, a focus is set on the experiences that emerge from the interaction with the "made thing" (the software, tool, product). Human-focused design builds on the intrinsic needs of users as they provide foundational categories of positive experiences that can later be used to describe and classify experiences with interactive products. Invertedly, through the identification of need deprecation and frustration barriers, products can be tailored to create or shape a desirable experience (Canossa et al., 2011). While functionality and usability are necessary preconditions for need fulfillment (i.e., pleasure), they are meaningless without the users' needs: all aspects must be aligned to create a positive experience: "Needs imply instrumental actions, which in turn imply appropriate operations; and only if the action is indeed instrumental, that is, it fulfills a particular need, an experience emerges." (Hassenzahl, 2010, p.57). The differentiation between usability and experience is reflected in the model of *hedonic qualities* of a product: "a "motivator," capturing a product's perceived ability to create positive experiences through need fulfillment" (Hassenzahl, 2010, p.52) versus *pragmatic qualities*: "a "hygiene factor," enabling the fulfillment of needs through removing barriers. As Tractinsky and colleagues (2000) found in a study on the relationship between beauty and usability in automated teller

[4] This highlights an interesting addendum to the understanding of gameful design as the application of game design elements as the process should be understood to include sensible substractions inspired by the polishing process of the tuning stage.

machines (ATM): “however, most surprising is the fact that post-experimental perceptions of system usability were affected by the interface’s aesthetics and not by the actual [objective] usability of the system” (Hassenzahl, 2010, p. 140), highlighting the importance of design elements serving pleasure as well as utility.

The necessity of identifying useful constraints on the human and the product side is underlined by Norman's (2013) definition of good design: “Design is the successful application of constraints until only a unique product is left.”. This means that in designing a product, the designer must work within the limits of what the product is supposed and what it can do. Design research differentiates between *affordances*: “*relationship between the properties of an object and the capabilities of the agent that determine just how the object could possibly be used*” (Norman, 2013, p.11), and *signifiers*: “*perceivable indicators that communicate appropriate behavior to a person*” (Norman, 2013, p.14). This distinction is relevant in that these terms relate to different thought processes in the design of a product. While affordances directly relate to the requirements of the tool, the signifiers relate to the user’s ability to discover the affordances of the tool. Thus, affordances relate to the architecture and engineering of the product, while signifiers relate to facets of user interface and user experience. One of the most important design practices in HCI is the concept of *feedback:* “communicating the results of an action” (Norman, 2013, p.23). The functions of feedback are to manage expectations, provide reassurance, afford quick learning, and the development of skilled behavior. If feedback isn’t immediate or informative, users can quickly give up in frustration, resulting in a waste of resources on both ends (time in case of the human, electricity/bandwidth in case of the system). Another key concept is *mapping:* “the relationship between the elements of two sets of things” (Norman, 2013, p.28). Mapping is the process that connects the affordances of the product to the signifiers. Using spatial correspondence through spatial analogies (like “to move an object up, move the control up” (Norman, 2013, p.22) is highlighted as a common example of *natural mapping*. Other natural mappings (that are based on principles from Gestalt psychology) are based on principles of perception, like natural grouping or patterning. The necessity and effectivity of this concept are highlighted in the findings of cognitive fit theory (see chapter 2.2.3). Mapping processes build on *conceptual models:* “an explanation, usually highly simplified, of how something works” (Norman, 2013, p.25). By combining their preexisting knowledge with the information that is in front of them, users create mental models of themselves, others, the environment, and the things with which they interact. These conceptual models are formed through experience, training, and instruction and thus guide users in achieving their goals and understanding the world. Such explanatory devices don’t have to be accurate “as long as they lead to the correct behavior in the desired situation” (Norman, 2013, p.103); however, simplified models are more prone to misinterpretation. Finally, a foundational method of HCI that should preface any design process is the “Five Whys” developed by Sakichi Toyoda (Serrat, 2017). This method suggests asking the question “*why [this is the case?]*” when confronted with a problem as often as necessary to identify the given problem as a symptom or a cause and thus be able to solve this or the respective underlying problem(s) on a deeper, more sustainable level.

The presented processes and methods serve as a foundation, particularly for the design-related parts of this work (see Chapters 3.3.1 and 4.4).

2.2.3. Kernel Theories in Cognitive Science

We present three theories from cognitive science that have proven to be essential foundations for the establishment of design principles for the design and development of the digital tool as well as the game design of the artifact analyzed in the in-depth studies.

Cognitive Fit Theory – (Vessey & Galletta, 1991)

Cognitive fit theory is a central concept for the representation of information. It is located at the interception of human-based perception, translation processes, and computer-based calculational types of representation. Cognitive Fit Theory was developed in 1991 by Iris Vessey, whose background is in computer science and business administration, with the focus of her work centering around human-computer interaction.

The theory is defined as follows: "Cognitive fit is a cost-benefit characteristic that suggests that, for most effective and efficient problem solving to occur, the problem representation and any too/s or aids employed should all support the strategies (methods or processes) required to perform that task" (Vessey and Galletta 1991, p.64), "Cognitive fit results when the problem representation and the task both emphasize the same type of information." (Vessey and Galletta 1991, p.67). The central focus of cognitive fit theory is to identify mismatches between task- and information presentation and mental representation in the mind of the users to amend emerging usability issues, especially in terms of performance. The theory builds on the information processing theory of Newell and Simon (1972), according to which human problem solvers are limited information processors that will seek ways to reduce their problem-solving effort. Vessey and Galletta build on this by stating that "one of the ways to reduce processing effort is to facilitate the problem-solving processes that human problem solvers use in completing the task. This can be achieved by matching the problem representation to the task, an approach that is known as cognitive fit" (Vessey and Galletta 1991, p.65). When there is a difference in problem representation and task, additional mental processes must be conducted as problem solvers form their initial mental representations from the materials presented to them (Perrig & Kintsch, 1985). Thus, efficient problem solving occurs when the processes that are used to communicate the problem match the type of task to be accomplished.

In their work, Vessey and Galletta focus on graphs, diagrams, and tables as representation types for experimental consolidation of their theory because of their ability to quickly and intuitively allow problem-solvers to see a meaningful distinction between them (graphs are "spatial problem representations"(Vessey and Galletta 1991, p.67) that emphasize information about relationships, diagrams are specific graph representations that preserve explicit information about the topological and geometric relations (Larkin & Simon, 1987), and tables are symbolic problem representations that emphasize information on discrete data values). They conclude their analysis with the understanding that "spatial tasks are best supported by spatial representations, while symbolic tasks are best supported by symbolic representations." Through comparison and analysis of these different mathematical presentation formats, they found correlated performance differences between input representation and expected output (Umanath & Vessey, 1994; Vessey & Galletta, 1991). While rooted in experiments within the domain of mathematics, follow-up studies building on cognitive fit theory in the domain of geographic information systems found fit-related performance differences among users of map- and table-based geographic information systems with regards to adjacency, proximity, and containment tasks (Dennis & Carte, 1998; Smelcer & Carmel, 1997), showing that the findings of the theory can be transferred to different domains. The emphasized relationship between the visual representations of data and the mental models of users serves as a core foundation for the design approach of the data visualization of the database navigation tool (see Chapter 4).

Cognitive Load Theory – (Sweller et al., 2011)

Cognitive load theory, developed by John Sweller and Paul Chandler (1991) and refined in 2011, is a psychological theory based in the field of instructional design. John Sweller (born 1946) is an Australian educational psychologist whose research is focused on cognitive factors in instructional design, with specific emphasis on the instructional implications of working memory limitations. The theories' focus is to highlight the different dimensions and resulting limitations of the working memory to serve as a foundation for the development of better teaching materials adapted with these limitations in mind. Cognitive load theory is an instructional theory based on current knowledge of evolutionary educational psychology and its relation to human cognitive architecture.

The theory differentiates between primary and secondary knowledge (Geary, 2008; Geary & Berch, 2016). While the first type is acquired subconsciously (e. g. the ability to solve problems, self-regulate our thoughts, and learn to listen to and speak our native language), only the second type is subjugated to active teaching processes and is commonly domain-specific (Tricot & Sweller, 2014). In the development of their theory, Sweller et al. devised a set of principles that are based on a comparative model to the evolutionary theory, building on the argument that "both human cognition and biological evolution are sophisticated natural information processing systems that create, disseminate, use and remember information" (Sweller, Ayres, and Kalyuga 2011, p.16). The principles most relevant to our research are the

1) *Information Store Principle*: This principle is concerned with the prevalence of information stored in long-term memory (where both primary and secondary information is stored), its requirements in terms of size "A natural information store must be sufficiently large to enable it to respond flexibly and appropriately to a very large range of conditions." (Sweller, Ayres, and Kalyuga, 2011,p.25) and its function as a "complex, variable environment in which a natural information processing system must function" (Sweller, Ayres, and Kalyuga, 2011, p.17), building on those information pieces that make it possible to treat the respective environment as familiar and predictable.

2) *Borrowing and Reorganizing Principle*: This principle elaborates on a specific type of learning, by which information is borrowed from the long-term memories of others by imitation and then assimilated into the own long-term memory. This transmission of information via imitation is never exact and is prone to noise.

3) *Randomness as Genesis Principle*: This principle suggests a different type of learning, where instead of imitating others (or the environment), random approaches to learning or problem solving can be applied as a source of novel information acquirement. Within this principle, a foundational rule of thumb regarding the number of processable items (five plus-minus two) is introduced, as only a very strictly limited number of items can simultaneously be held and processed in working memory: "Faced with novelty, we lack a central executive in working memory indicating to us how to organize new information. Therefore, the absolute number of elements that we must organize becomes a critical factor" (Sweller, Ayres, and Kalyuga, 2011, p.45). Information is retained in long-term memory if it proves useful or jettisoned if it does not.

5) *Environmental Organizing and Linking Principle*: This principle elaborates how the information that is processed by the working memory and later stored in the long-term memory is linked with the context of its environment. It thus links back to principle 1) on how a natural information processing system is dependent on a context representation within long-term memory to function appropriately in its environment.

As elaborated in principles 2) and 3), the intrinsic properties of a piece of information impose a cognitive load in itself. Additionally, depending on the instructional process and presentation, extraneous cognitive load is added through the additional number of elements that must simultaneously be processed. These principles, combined with the findings of cognitive fit theory, serve as the underlying foundation for the design decisions we present throughout this work. Whereas cognitive fit theory elaborates on the "how" of visual representation, cognitive load theory gives directions on the "how many" and "how much" in terms of the number of different design elements shown to users at each point in time.

Flow Theory – (Csikszentmihalyi et al., 2014)

With the emergence of user experience as an essential component in the domain of human-computer interaction, flow theory, developed by Mihaly Csikszentmihalyi, who is a Hungarian-American psychologist, has become one of the foundational theories for experiential design. Csikszentmihalyi, Professor of Psychology and Management at Claremont Graduate University, is known for his research on the concept of flow (a highly focused mental state conducive to productivity) and positive psychology.

In their latest revisit of the theory, Csikszentmihalyi et al., (2014) describe flow as: "a subjective state that people report when they are completely involved in something to the point of forgetting time, fatigue, and everything else but the activity itself." (Csikszentmihalyi et al., 2014, p.230). The authors state that in order to achieve the intrinsically rewarding experiential involvement that defines flow, clear goals, optimal challenges, and coherent, immediate feedback are the necessary foundational features: "goals serve to add direction and purpose to behavior. Their value lies in their capacity to structure experience by channeling attention rather than being ends in themselves" (Csikszentmihalyi et al., 2014, p.231). Optimal challenges are achieved through a balance between perceived challenges and perceived skills as described in the concept of optimal arousal by Berlyne (1960) and Hunt (1965). Finally, immediate feedback informs the individual on how well they are progressing in the activity and dictating (or suggesting) whether to adjust or maintain the present course of action. One of the fundamental concepts flow theory builds upon is the concept of *Funktionslust* ("activity pleasure") elaborated by Groos (1901) and Bühler, Greenberg, and Ripin (1930) as the emission of positive feelings while performing well-trained movements. Another related concept flow theory is built upon was introduced by Piaget and Cook in 1952 as "the pleasure of being a cause" that drives infants to experiment in the earliest stages of sensorsimotor development. Further foundations for flow theory are the intrinsic psychological needs of competence and autonomy (and relatedness – but this is not touched upon in flow theory) by Deci and Ryan (1985). Research on task involvement (as conducted by Greenwald (1982); Harackiewicz, Barron, and Elliot (1998) and Harackiewicz, Manderlink, and Sansone (1984)), suggests that predictions can be made on the individual's involvement in an activity by measuring how important it is to him or her to do well in the activity. Examples of emergent types of intrinsically motivating activities are, for example, when a person is at first indifferent towards or bored by an activity (like using a computer or playing a video game), but then, as opportunities for action become clearer or the individual's skills improve, the activity begins to be interesting and, finally, enjoyable (Csikszentmihalyi et al., 2014). Flow theory is strongly related to game design theory in terms of the desired outcome and the means to get there, as it offers the foundations for improving user motivation and continuous interaction.

2.2.4. Classification Outcomes

Our overall goal is to derive an ontology that offers experts as well as novices an entry point into the broad design options offered by game design practice. The most foundational method for transforming large amounts of information into workable chunks is the process of classification. As stated by Bailey (1994), classification is the foundation for conceptualization, language, and speech, as well as mathematics, statistics, and data analysis. It is defined as "the ordering of entities into groups or classes based on their similarity" (Bailey, 1994, p.1). Among others, the benefits listed by Bailey that specifically appeal to our needs are *reduction of complexity* (particularly for the topic-novices within our users), *identification of similarities and differences* (this is of interest for both scholars as well as practitioners in terms of identifying clusters of relevance as well as gaps), *inventory and management of type*s (through this lens, the foundational elements of a generic dataset entry can be better understood) and *versatility* ("a good classification can not only represent the [entities] studied, but also locate them within a property space formed by combining the variables utilized in the analysis." (Bailey, 1994, p.12-14)). Defending their use from the argument that classifications are "merely descriptive" or "pre-theoretical," Baileys highlights their core value as the foundations for explanations.

There are different types of outcomes of classification processes. Bailey differentiates between two main types of classification: the *typology*, which is derived conceptually by creating a matrix of existing, conceptual terms and designating a resulting type-concept per cell (taxon), and the *taxonomy,* which is similar to the typology in terms of outcome but derived empirically. Building on Bailey's work as a foundation as well as a literature review of 73 IS papers on the development of taxonomies, Nickerson et al. (2013) name three different categories of methodological approaches that they derived: *inductive* – observing and analyzing empirical cases to arrive at a final result equal to the taxonomy named by Bailey and similar to the phenetic approach in Biology (clustering according to visible similarity traits), *deductive* – building on a previously derived conceptual or theoretical foundation equal to the *typology* by Bailey and similar to the cladistic approach in Biology (clustering according to traits derived by a common ancestor) and *intuitive* – an ad hoc approach building on "the researcher's perceptions of what makes sense" (Bailey, 1994, p.340).

Van Rees (2003) further differentiates the terms taxonomy, classification, and ontology by building on definitions given by the Merriam-Webster dictionary, where *classification* is defined as the "systematic arrangement in groups or categories according to established criteria." (Merriam-Webster Incorporated, 2022b) and *taxonomy* as the "orderly classification of plants and animals according to their presumed natural relationships." (Merriam-Webster Incorporated, 2022c). The definition of the latter highlights the cladistic (deriving from evolutionary relationships (Eldredge & Cracraft, 1980)) nature of taxonomies in contrast to the more simplistic phenetic (similarity-based) grouping that is associated with the term classification (see Figures 10-12). While Baileys states that classes must be mutually exclusive (Bailey, 1994, p.3), new application forms (particularly in the domain of machine learning (Tsoumakas & Katakis, 2007)) also afford multi-label classifications.

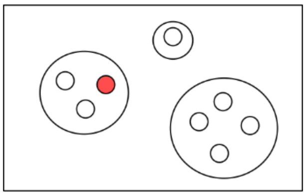


Figure 10 – Single Label Classification

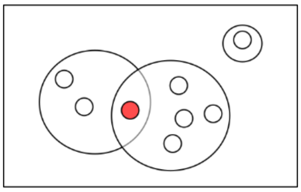


Figure 11 – Multi Label Classification

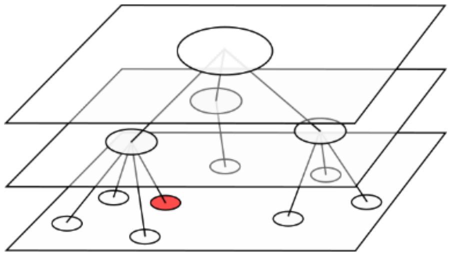


Figure 12 – Taxonomy

The term ontology, while originally referring to a branch of metaphysics, has been reinterpreted by information science to refer to a more complex, multi-dimensional classification structure during the emergence of the semantic web (Maedche & Staab, 2001). It is referred to by the WebOntology working group at W3C as a "machine-readable set of definitions that create a taxonomy of classes and subclasses and relationships between them" (McGuinness et al., 2002, p.2). It includes elements of classifications and taxonomies but expands on these constructs by further affording links (references to other datasets) and semantic data (see Figure 13).

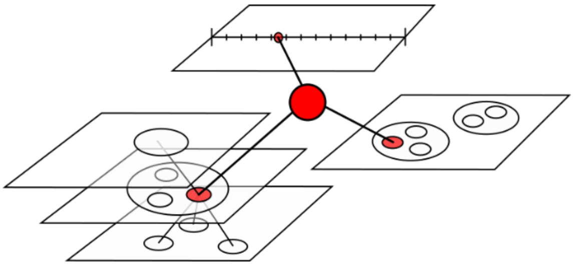


Figure 13 – Ontology

The outcome we are aiming for targets several fundamentally different target audiences: game design experts and novices, scholars, and practitioners. While their questions and goals build on the same datasets, the approaches of these user groups to achieving these means will be fundamentally different: in terms of navigation, experts can use their preexisting knowledge to orient themselves deep within the hierarchical branches of the ontology while novices will need to start their exploration from a more generalized overarching perspective. In terms of application, scholars might be more interested in abstract factors like different distributions of the depicted status quo, while practitioners will want to arrive at specific recommendable solutions to their current problem. The final product could thus incorporate taxonomic, hierarchical structures that lead from general and abstract clusters to specific and detailed entries, allowing for a drill-down-based selection process. It could also benefit from multidimensional classifications for each dimension that is relevant to the users. Given that an ontology affords both, we choose the ontology as the informational structure along which to prepare our data processing and build our design artifact.

Part II

Development of the Digital Ontology

3. Chapter 3
Data Structure Development and Aggregation

"We are much too much inclined in these days to divide people into permanent categories, forgetting that a category only exists for its special purpose and must be forgotten as soon as that purpose is served."

Dorothy L. Sayers, Are Women Human? Astute and Witty Essays on the Role of Women in Society

3.1. Introduction

The overarching intention of our research is to provide the means with which *informed* and *inspired* decisions can be made during any process of enhancing a system or service through gameful design. To make an informed decision, users need to be able to gain information on the selection pool (quantitative data) to understand the available options as well as detailed meta-information on items of interest (qualitative data) to select the most suitable option (Edwards, 1954). On the other hand, to be able to make an inspired decision, users should also be provided with the possibility to arrive at serendipitous findings that are outside their typical search algorithms. In the following chapter, we present the methodology and results of our efforts toward an ontology that serves as the foundation to achieve this goal. For this, we first aggregated two datasets, one listing game design elements and one listing playing motivations, including their names, descriptions, and metadata on their originating publication. Following that, we tested two different methods of enriching these datasets with user-relevant metadata: first, we conducted user-generated label-based clusterings through two types of card sorting (open and closed) (Chapter 3.3), and second, we explored an algorithmic approach based on keyword-mapping and -matching (Chapter 3.4). The closed card sorts, while insightful, dominantly highlighted the variance in different users' understanding of potential underlying categories. Through the following open card sorts, we found user-relevant overarching viewpoints for categorization that offer different logical entry points into the ontology. However, due to its drawbacks in terms of efficiency and reliability, we decided against further studies using this qualitative, user-generated labeling methodology and looked into an alternative, automated methodology for enriching the ontology with relevant and reliable metadata. The results of the first internal keyword matching studies allowed us to identify overlapping nodes within the emerging graph and evaluate the internal consistency of our dataset. Through an explorative clustering of the underlying keywords, we further identified a set of dominant clusters prevalent in current player type/playing motivation literature (particularly focused on the themes of social interaction, achievement, and exploration). In the second study, a keyword-based matching of two separate but related datasets (playing motivations with human needs), we were able to demonstrate a strong connection between these two topics, enriching our dataset of playing motivations with linked underlying needs. By conducting another explorative keyword clustering, we were able to identify a set of themes that are not yet prevalent in the literature underlying our playing motivations dataset but existent in game design practice (e. g., clusters around the topics of deference and submission, luxury and resistance). As such, our preliminary results provided us with leads for new, potentially not yet fully targeted audience clusters. The results of both keyword-matching studies afforded our ontology with rich, linked metadata and served to improve the ontologies schemata. In the following chapters, we document the process of building the data foundation of our ontology, starting with a pre-study on the distribution of currently researched gamification elements (Chapter 3.2.1), followed by the aggregation process of our underlying datasets of game design elements and playing motivations (Chapter 3.2). We then present the results of our explorative, card-sort-based labeling endeavors (Chapter 3.3) and, finally, the keyword-based matching studies towards the conflation and enrichment of our data (Chapter 3.4).

3.2. Dataset Aggregation

3.2.1. Literature Review on Gamification Elements

Introduction

We started our process by conducting a literature review to gain an overview of the current research landscape concerning the most researched and implemented game design/gamification elements. Studies claim a dominant focus in gamification research on points, badges, and leaderboards (Chou, 2016; Seaborn & Fels, 2015; Toda et al., 2018); however, other social-oriented elements (e. g. effects of competition (Sepehr & Head, 2013; Witt et al., 2011) and collaboration (Lounis et al., 2014; Meske et al., 2017)) have also gained traction in recent years. The first focus of our review was to gain quantitative insights into the current distributions:

RQ1: *Which game design elements emerge most prominently in gamification research?*

Apart from the elements themselves, we were also interested in the outcomes they are implemented towards. As gamification is inherently defined by its appliance in domains outside of the context of games, it is relevant to understand how the inherent qualities of different game design elements are used to achieve desired outcomes by affecting specific parameters. Thus, the second focus of our review was to gauge the parameters analyzed in these studies:

RQ2: *Which parameters are influenced by the implementation of game design elements, and to which degree?*

As noted by Nacke and Deterding (2017), it is important to investigate the changes of effects over varying situational circumstances. Given that research on gamification and gameful design is spread over a diverse range of research domains (Zhang et al., 2021), we were interested in whether domain-specific factors influenced the distribution of different emerging game design elements:

RQ3: *Are there domain-specific saliences regarding specific game design elements?*

Based on these research questions, we conducted a literature review followed by a meta-analysis on the game design elements and analyzed parameters over different domains. In terms of analyzed parameters, motivation was the most studied parameter by a large margin. However, overall, we were able to aggregate a list of 23 unique design elements excerpted from the analyzed studies, showing an emerging awareness of the multitude of possible design elements outside of PBL (a conglomerative term for the implementation of points, badges, and leaderboards (Toda et al., 2018)). We conclude that aggregating a game design element ontology can support and strengthen this desirable tendency.

Methodology

We started our research endeavor by conducting a literature review on studies of game design elements in gamification contexts. This process was based on the approaches recommended by Webster and Watson (2002). We used the KIT-Katalog Plus (Karlsruhe Institute of Technology, 2012) as our primary source and the platform "Google Scholar" (Google, 2004) as the secondary source for this forward-backward search. We first searched for gamification studies without specifying application domains[5]. To cast a wider net, we next added keywords linked to more specific domains and uses to our search process[6]. We then continued with a forward and backward search on the most relevant findings from the previous search. Of the 492 findings from queries entered into the KIT-Katalog Plus, we found 57 studies to pass the criteria relevant to our research: studies that presented a practical study measuring and evaluating the effect of one or several game design elements on specific parameters (see Table 26 in Appendix A.1.1). We proceeded to create a concept matrix in which we accumulated and structured the findings of the literature review according to the game design elements, metainformation on the studies they were extracted from, the measured effects, and the application domain. The matrix was blindly counterchecked by two separate members of the research team.

Results

The final matrix consists of 23 unique game design elements, 16 categories of the monitored effects, their directions (positive effect, negative effect, no findings), and ten different application domains. The three game design elements that were investigated most often across the analyzed gamification studies were *badges* (33, 57.9%), *points* (31, 54.4%), and *leaderboards* (32, 56.2%), as can be seen in Figure 14. In terms of the measured parameters, *motivation* emerged as the most dominantly tested parameter across all studies (34, 59.6%) (see Figure 15). Other prominent parameters include engagement (14, 24.6%), activity (10, 17.5%), performance (8, 14,0%) and enjoyment (6, 10.5%). The final correlation matrix where we map the topic domains and game design elements clustered by overarching category show *points*, *badges,* and *leaderboards* to appear in papers across most domains with appearance rates over 50% over all studies and between 80%-100% appearance rates across all domains (see Table 27 in Appendix A.1.2).

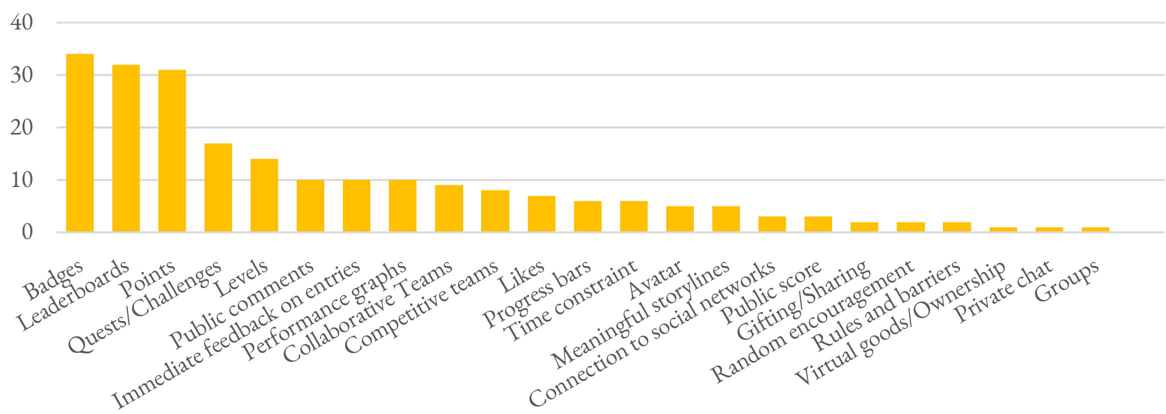


Figure 14 – Distribution of Game Design Elements from Gamification Literature

[5] ("gamif*" AND "study" AND "education") OR "health" OR "sustainability" OR "online participation"
[6] "teams" OR "feedback" OR "crossfit" OR "runtastic" OR "freeletics" OR "biology" OR "trial"

In terms of the distribution of studies across topic domains, *education* emerges as the domain with the highest number of studies, featuring 23 (40.3%) of the 57 studies with great distance from the second domain, *online participation*, which featured 8 (14.0%) studies, followed by health (7, 12.3%) and fitness (6, 10.5%).

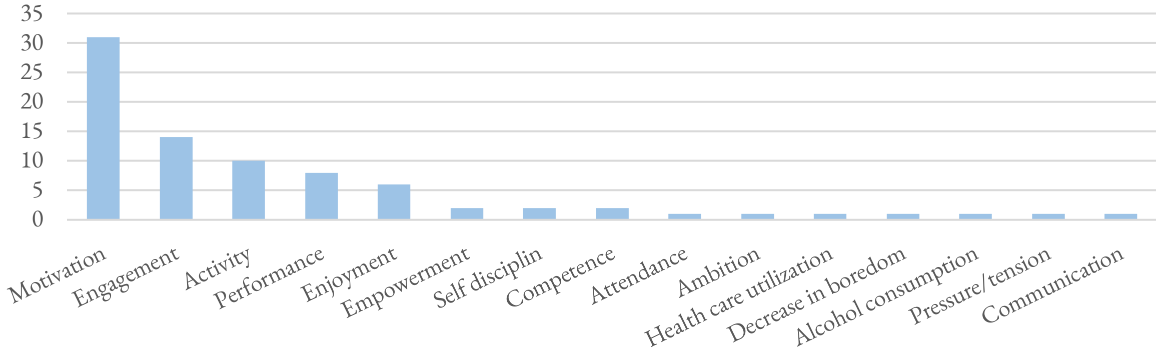


Figure 15 – Distribution of Topic Domains from Gamification Literature

Discussion

Our results concur with the literature in that points, badges, and leaderboards emerged as the most studied game design elements overall and across all or almost all analyzed content domains. We conclude that the dominance of these design elements seems to have persisted since the study conducted by Seaborn and Fels (2014). Their study concludes that existing reviews about gamification are limited in terms of the variety of game design elements and contexts and that more empirical research employing statistical analysis and reports on effect sizes must be conducted. In this regard, however, it seems that some progress has been made regarding the total of 23 unique game design elements we found over the 57 studies we analyzed. Also, each content domain featured at least five different game design elements, showing that research seems to evolve in a more diverse direction. Also, apart from the mainly analyzed outcome parameter "motivation," 14 different parameters were analyzed in these studies, showing the broad application possibilities for game design elements. From our findings, we conclude that while there is an emerging interest in different design elements across domains, there is still a prevalent research gap in terms of the diversity of analyzed game design elements.

3.2.2. Dataset Aggregation of Game Design Elements

Through the evaluation of the findings of the pre-study outlined in the previous chapter, we identified an imbalance in gamification research in terms of a few elements being the focus of a lot of the studies we analyzed (points, badges, leaderboards, and progress indications) in contrast to a broad number of design elements only looked at by one or two studies. Evaluating commonalities of these dominant elements, we see most prominently easily implementable interface-related game mechanics while at the same time noting an absence of facets that relate to more ubiquitous concepts relating to aesthetics or fantasy.

Given this skewed distribution, we decided to compile a new dataset with elements extracted from the literature on game design. For this endeavor, we altered our review and extraction process for several reasons. First, the gamification studies we had aggregated in the previous study generally focused on one or a small set of elements and tested them for specific effects. In contrast, the focus of game design literature is set on teaching a spectrum of components and techniques, thus offering a huge range of different elements that are often already pre-sorted by the authors into context-dependent categories. These aggregations are oftentimes based on experience rather than scientific methods, as many authors in this domain originate from the applied side of the game design process. Through this approach, we build on the game design expert perspective that we found lacking in our previous aggregation. We started our selection process with a sample of books by well-acknowledged experts from the domain and then included further literature from a forward-backward search, following the process recommended by Webster & Watson (2002).

We started our literature review with the works of Schell (2014), Salen and Zimmerman (2004), and Koster (2005). Each of these books follows a different angle: empirical, theoretical, and psychological. For the first starting point, we chose Schells' "The Art of Game Design" (2014). Due to his first-hand experience as a game developer and designer (and later entrepreneur and manager), the author of the first book is considered a mentor within the game design community. With its multi-faceted perspective (referred to as lenses in the book) based on design principles founded in psychological theories, his work serves as an important foundation for the game design element dataset as well as the subsequent forward-backward search. The backward search from this book led to the articles "Engaging by Design: How Engagement Strategies in Popular Computer and Video Games can inform Instructional Design" (Dickey, 2005) and "Chris Crawford on Game Design" by Crawford (2003). While the first publication is more theoretically oriented, focused on game design strategies, the role of narrative, and methods of interactive design, the second is narrated from a more personal and practical perspective and is focused on describing the foundational skills behind the design and architecture of games. We found these two perspectives interesting facets to complement our dataset. The subsequent forward search yielded "Game Design Workshop: A Playcentric Approach to Creating Innovative Games" by Fullerton (2008), a book that is valuable to our research as it looks at the essential structure of games, distinguishing between formal, dramatic, and dynamic elements.

The second book we chose was "Rules of Play- Game Design Fundamentals" by Salen and Zimmerman (2004). This work introduces basic concepts of game design that are referred to as "fundamentals." Due to the wide range of these fundamentals - from abstract concepts (interactivity, player choice, action, and complexity) to concrete elements (backstory, character, real-time video), this publication serves as another valuable foundation for our dataset.

From the backward search, we included Egenfeldt-Nielsen, Smith, and Tosca's "Understanding Video Games: The Essential Introduction" (2010) for its focus on player culture – inside and outside the boundaries of games. In terms of a forward search, we included an article from the publication-website "Game Developer" (formerly "Gamasutra" (Informa, 1997)). The site was founded in 1997 and t is a well-established source within the developer community, with established professionals from the industry participating through interviews as well as publishing blogs and project post-mortems. In the article "The Aesthetics of Game Art and Game Design," Chris Solarski (2015) analyzes the psychology of shapes and dynamic composition, lending an additional yet unaddressed perspective on aesthetical factors of game design and was thus included. Another publication we found through our forward-search was the book "Game Architecture and Design" by Rollings and Morris (2004). This too is a practitioner's report on game design, interestingly however in this book game design and game architecture are discussed as separate processes. With its deep and multifaceted categorization of different elements, this work serves as a valuable source for our dataset.

Lastly, we used Raph Koster's "A Theory of Fun," published in Koster (2005), as a final starting point. Its perspective on the fun aspect of games, as well as its focus on practical implementation, made it a valuable addition to the foundations of our dataset. We found two works through our backward search that we chose to include in our dataset: "Game Design: Theory & Practice" by Rouse (2004) and "Triadic Game Design: Balancing Reality, Meaning and Play" by Harteveld (2011).

While we included the first of these two publications for the breadth of the topics and facets it covers, the second publication was of interest due to its special focus on serious games. As serious games are rooted thematically stronger in problems and topics of real life, they are considered the links between entertainment-oriented games and gamification (Deterding, 2016), which might close a potential gap between the elements procured from our gamification studies and this practitioners-based dataset.

After aggregating the foundational literature, we extracted the game design elements by hand, copying each new game design element emerging from the respective text into our dataset, including the according description as well as the meta-data on the source (title, author, year, ISSN/ISBN).

The definition we chose as a baseline for what consists of a game design element is comparatively open (see Chapter 2.2.1); thus, elements were included independent of their hierarchical level given throughout these texts. As each author operates on their framework or ontology (explicit or implicit), we included all metadata on these hierarchies through a label called "author category." In preparation for the keyword matching studies presented in Chapter 3.4, we later also conducted a keyword extraction from the authors' descriptions. The process was done by hand, where we analyzed each sentence of the description and extracted the defining keyword(s). The list was re-checked twice by two different members of the research team. In these iterations, we set a special focus on the keywords not losing their meaning from being out of context.

The final dataset includes a total of 595 game design elements by name, enriched with metadata on author category, a long and short description, key characteristics, a representative picture link, and metadata of the originating literature: research id, scientific field, author(s), title, abstract and year (the full dataset is uploaded at https://gonku.de/sup-mat-phd-gho/Dataset-GDE.xlsx).

3.2.3. Dataset Aggregation of Player Types and Playing Motivations

Based on best practices in game design, it is essential to have an understanding of the target audiences that the game or gameful measure is supposed to address. In this regard, research on player types and playing motivations has come into focus with the growing distribution of digital games. Understanding the varying needs of a target audience is the central key to marketing a product; thus, many researchers of different game studies domains (philosophy, history, social studies,) as well as information systems and economics studies, have been interested in separating the different facets of the medium that resonate with needs and personality characteristics (Tuunanen & Hamari, 2012).

To prepare the affordance of target audience-related metadata, we conducted another literature review looking for playing motivations within the subdomain of player-type research. Once more, we conducted a search with an open search on Google Scholar (Google, 2004) as well as the KIT-Katalog Plus (Karlsruhe Institute of Technology, 2012), looking for combinations of words relating to player types and incentives[7]. We followed this up with a systematic forward-backward search as proposed by Webster & Watson (2002). After aggregating a list of papers relevant to our research, we focused on three well-cited and established publications as the starting point for the forward and backward search. Our selection criterion for the inclusion of papers to our backlog was for publications to inherently provide lists and aggregations on player types or motivations. The first publication was the "Meta-synthesis of player typologies" by Tuunanen and Hamari (2012), which consists of an aggregation of studies that conducted player type segmentation based on behavioral and psychographic criteria. The backward search from this publication yielded 12 different studies with between two and eight different items per framework. Furthermore, we chose Dixon's (2011) „Player Types und Gamification" and Stewart's (2011) Model of Personality and Play Styles as two further publications from which we conducted a further forward and backward search. This search yielded a sum of 46 publications, of which we selected 34 to be included in our final list. Finally, we conducted another literature search in Google Scholar (scholar.google.de) on the keyword combinations ("gamification" or "game design" and "player type") and ("gamification" or "game design" and "playing incentive" or "playing motivation"). Of the 7 publications we additionally found and took into consideration, 5 more publications were added to the list, resulting in 39 publications in total. The final list includes publications from the domains of philosophy, history, social studies, information systems, and economics that span a timeframe of over 50 years.

As with the dataset for the game design elements, we extracted the playing motivations and player types we found by hand, adding similar metadata as the other list (name, description, source data (title, author, year, ISSN/ISBN)). Using the same method for keyword extraction we applied to the game design element dataset, we extended the dataset with relevant keywords. This column was once again re-checked by two different members of the research team. As the motivations operate on the same hierarchical levels, no author categories were found. The final dataset includes a total of 234 playing incentives by name, including a short description, key characteristics, a representative picture link, and metadata of the originating literature: research id, scientific field, author(s), title, abstract, and year (the full dataset is uploaded at https://gonku.de/sup-mat-phd-gho/Dataset-PM.xlsx).

Apart from serving as the basis for our follow-up research, based on their extent, scope, and meticulousness, these datasets contribute to research and practice as extensive foundational data for meta-studies, in-depth research as well as inspirational tools for gaining an overview of existing design elements and playing motivations.

[7] Search terms in Google: play* AND (type OR motiva* OR incent*)

3.3. Dataset Labeling

After having gathered our foundational datasets, our next concern was how to enrich them with user-relevant metadata. The final ontology is meant to provide for user needs as different as offering recommendations for specific problems as well as offering a comprehensive overview. For our dataset to be developed into a fully functional ontology, the data needs to be further labeled and classified. The following chapters outline our research process toward achieving this goal. We first report the results of a closed card sort that we designed to test the compatibility of our dataset with preexisting classifications. Therein where we evaluate four different classifications of games, game design elements, and playing motivations in terms of their suitability with exemplary items from our dataset (Chapter 3.3.2). Following that, we conducted an open card sort where game design experts empirically developed their own category structure based on a preset of cards from our dataset (Chapter 3.3.3). The following chapter outlines our rationale for the selection of the card sort methodology and the tool we developed for conducting the experiments.

3.3.1. Methodology Selection and Tool Development

According to Bailey, successful classification is achieved by ascertaining the key or fundamental characteristics on which the classification is to be based, as the final result is directly shaped by the selection of variables (Bailey, 1994). This suggests a single usage purpose for the outcome of the classification based on the single perspective taken by the conductor of the process. In the approaches listed by Nickerson et al. (2013), one entity conducts the sorting and classification, typically the researcher or their team. However, other domains like user experience design use methodologies that integrate and conflate user perspectives (Fincher & Tenenberg, 2005; Rugg & McGeorge, 1997).

Kelly's (1955) Personal Construct Theory (PCT) acknowledges the uniquely constructed understanding of the world by an individual; thus, the resulting ontology should be able to serve a diverse group of users' needs to assess and incorporate different perspectives to arrive at a useable common ground. As we anticipate different entry points into our dataset based on different levels of expertise (novice vs. expert) and usage intention (scholar vs. practitioner), we chose to use the methodology of card sorts as outlined by (Rugg & McGeorge, 1997) for gaining user-based insights into expectable classification dimensions.

As a method, card sorting offers several benefits to our needs: first, when conducted as an open sort, it reflects the user's needs and mental structure (inductive/empirical result), and second, when conducted as a closed sort, it can assess existing structures with regard to their compatibility with a dataset (deductive/constructed result). Finally, it is compatible with our dataset structure as it is typically conducted via cards that are typically structured by the name of the item, a short description, and/or a picture (Fincher & Tenenberg, 2005) suitable to the way items are structured within our dataset. This minimalistic layout naturally prevents extraneous cognitive overload and eases focus on the task – thus, intuitive classifications can naturally emerge (Rugg & McGeorge, 2005).

While card sorting can be conducted analogously, which is beneficial for card sorts that are conducted in group sessions, for single-person sorts, digital card sorting affords higher levels of convenience, particularly concerning data management. While digital card sorting tools exist, we decided to design and develop our own tool as we were concerned about matters of data protection (data hosted and provided outside of our control) and wanted to be able to include functionalities specific to our needs. The design rationale behind the development is documented in Appendix A2.1.

3.3.2. Closed Sorts

To arrive at a user-needs-oriented classification structure that can intuitively be used and understood, we started by looking into preexisting structures into which to sort our game design elements. Given the already extensive prevalence of well-established frameworks around games and playing motivations in the field of game theory, we wanted to first assess existing, established frameworks from game design, gamification design, and playing motivations in terms of their suitability as classification labels for a dataset of game design elements.

We started by collating a set of existing playing motivations frameworks from game theory and research to test their suitability for the study through preliminary sorts. Since the surge of interest around the topic of game research, different types of frameworks underlying game design and gamification research have been developed under varying perspectives. For example, in his work, Crawford (2003) offers a classification according to *medium* - differentiating between card, board, athletic, and computer games. He further offers an additional classification structure relating to *foundational elements* of gameplay: representation, interaction, conflict, and safety. Salen and Zimmerman (2004), p.5) base their foundational classification based on *fundamental philosophical perspectives* of game derived from Huizinga (1956) and Caillois's (1961) works, differentiating between rules, play, and culture. Focusing on yet another perspective, the MDA framework by Hunicke, LeBlanc & Zubek (2004) builds on the *relationship between player and developer* by differentiating a game's mechanics, dynamics, and aesthetics. Most frameworks in the domain of game design are derived empirically from observing player data, building on the prevalent technologies at the time of their development, and are influenced by the authors' perspective and field of competence.

We started the process by selecting 13 frameworks with a wide variety of viewpoints: four from the field of game studies, two from the field of gamification, two from player type and two from playing motivation research, and a set of frameworks from psychology relating to personality and human motivation. To narrow our selection down for the actual experiment, we conducted a series of quick sorts on all 13 frameworks by only sorting items into the top-level categories of the authors' categories (Brucker, 2010; Wood & Wood, 2008). We used a preset of items of our dataset featuring 98 game design elements from two sources: "Triadic game design: Balancing reality, meaning and play" (Harteveld, 2011) and "Game design workshop: a playcentric approach to creating innovative games" (Fullerton, 2008). For each framework, we created a structure in the card sorting tool consisting of a set of immutable top layers drawn from the framework and an additional category called "no category" for items that did not suit any of the framework's categories. An overview of the outcomes of these preliminary sorts is presented in Table 28 in Appendix A.2.2.

Of the four frameworks we sorted from the field of game studies, we found the "MDA"-framework by Hunicke et al. (2004) to be most suitable for further consideration as it produced the lowest number of items that could not be sorted into one of the categories (in contrast to Dignan's (2011) and Poels et al.'s., (2007) framework, where respectively 30% and 15%of items could not be sorted) as well as a relatively even distribution across categories (in contrast to Schell's (2014) framework, where 67% of all items had been sorted into one category). Of the two frameworks originating from gamification, we chose Robinson and Bellotti's (2013) framework[8] (VAC), as more than half the categories offered by the framework of Deterding et al. (2011) did not apply to our card set. With the aim in mind to link our game design element dataset to our playing motivations dataset, we included the framework[9] (MfP) by Yee (2006) as well as the "Octalysis"-framework by Chou (2019) in our structure as they outperformed the player type frameworks (Kahn et al., 2015; Tondello et al., 2016). While the psychological models and frameworks (King et al. 2009, Deci and Ryan 1985; Reiss 2001; Murray 2008; Myers and Myers 2010) performed overall well, we decided to push the inclusion of psychological studies to be studied in a future experiment to keep the experimental design concise. In summary, the four frameworks we selected for the closed card sort are:

1) the "MDA"-framework by Hunicke et al. (2004)
2) the "VAC"-framework by Robinson and Bellotti (2013)
3) the "MfP"-framework by Yee (2006)
4) the "Octalysis"-framework by Chou (2019).

For in-depth information on the chosen frameworks, see Appendix A2.3.

Experimental Design

We designed the experiment as an online experiment, particularly due to advantages in terms of flexibility and convenience (Evans & Mathur, 2018), as we were asking participants to conduct two individual sorts at two different points in time. The experiment was presented and organized through a survey structure consisting of two parts: the first part consists of the assessment of demographic factors, level of expertise, and player type, and the second part leads to a link where the card sort is performed. Participants received the links for the two sorts separately. We designed the full setup in English to stay consistent with the used scales, frameworks, and dataset items. The practical part of the experiment featured a closed card sort, specifically a repeated single-criterion sort (Rugg & McGeorge, 1997), where the same elements are sorted exclusively into an existing structure by different participants. While we had used only one preset for the first round of framework selections with 98 items from the dataset, we added another preset of 87 items from the same dataset but in a different segment. The second preset featured randomized items from our original dataset to compare the outcomes between the two sets and thus be able to better assess the frameworks in terms of compatibility for the full dataset. In summary, our experiment consisted of two presets of items that were to be sorted into the four frameworks. Participants were randomly assigned based on a structure where they either had to sort the same preset for two different frameworks or the two different presets into the same framework (see Table 29 in Appendix A.2.4).

Experimental Conduct

We recruited participants via a mailing list targeting students at the university (a sample group that, according to Druckman and Kam (2011), "do not intrinsically pose a problem for a study's external validity," p.1). Instead of aiming for a specialized group of experts, we were interested in the perceived usefulness of the

[8] We abbreviated the name of this framework: "Taxonomy of Gamification Elements for Varying Anticipated Commitment" to VAC for easier readability

[9] We abbreviated the name of this framework: "Motivations for Play" to MfP for easier readability

selected framework for interested laymen. Participation was incentivized through a lottery of two 25€ gift cards for Amazon.com (1994). After signing up for the experiment, participants received an email with the link to the survey and experimental instructions. They were briefed to contact the experiment administration if any questions or problems arose.

In the first part of the survey, participants answered questions on their demographics (age, gender, education, occupation) as well as their gaming experience and game design expertise (number of hours played per week, how long participants have been playing video games (Karle et al., 2010) and their preferred playing devices and modes. This section was followed by an assessment of their player type through the Hexad user type test (Tondello et al., 2016), consisting of 24 statements that are assessed on a 6-point Likert scale (strongly disagree (-3) to strongly agree (3)). The test assesses the percentage of participants to fall into each of six different player type categories: (Socializers, Free Spirits, Achievers, Philanthropists, Players, and Disruptors). Once finished, participants were given a link that opened the card sorting tool in a new browser tab with a unique URL for each sort. On the welcome screen, the participants were given information on the contents of the framework underlying their sorting process as well as general instructions on tool usage and the next steps and were then asked to commence with the sorting process. They were further instructed to send an email back to the experimenter on finishing the sort. After two days, they then received the link to the second sort as well as given the option to send feedback on the sorting experience inquiring in terms of card content (best fitting items, most difficult to match) as well as game aspects that weren't covered by the framework and game aspects that weren't covered by the items. They received a final email with the debriefing, as well as a notification of their status in terms of winning the reward lottery. For the survey, we used the online-tool LimeSurvey (LimeSurvey GmbH, 2003), and for the card sort, we used the tool we had developed (Hoffmann & Martin, 2018), see Appendix A.2.1.

Analysis

In terms of evaluation, we chose to evaluate the sort outcomes according to the following parameters suggested by Rugg and McGeorge's (2005) card sort analysis:

1) The *fitness* of the predetermined structure towards our dataset. We measure this via the number of items participants did not sort into the "no category" folder (see Figure 16).

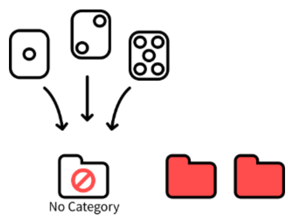


Figure 16 – Fitness

2) The *commonality* of each framework's categories. We measure this via the number of times the same item was sorted into the same category folder by different participants (see Figure 17).

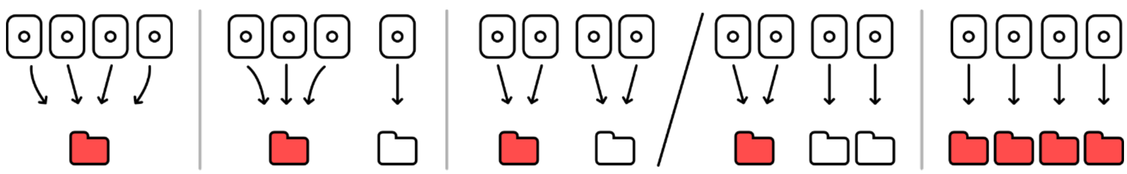


Figure 17 – Commonality

3) The *consistency* of the frameworks' performance with regard to different extracts of our dataset. We measure this by evaluating the standard deviation of the commonality average between the two different presets (see Figure 18).

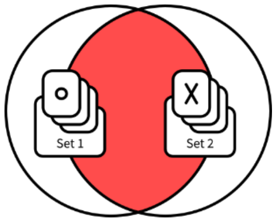


Figure 18 – Consistency

We added the last measure as an evaluation parameter specific to the design of our study to test the frameworks' ability to accommodate the full set of elements of our dataset. Thereby we wanted to gain better insights if low or high outcomes in the commonality measure were stemming from the framework compatibility with the whole dataset or specific to a flaw within one of the presets.

Results

In total, 16 people participated in the experiment over a period of 6 weeks. Three participants only conducted one of the two card sorts, resulting in a total of 29 closed sorts we could use for the analysis. The missing sorts affect 3 of the four frameworks (Yee, 2006; Hunicke et al., 2004; Chou, 2016). All sorts were complete in that all items were sorted into the framework's structure or into "no category." The age range of our participants spanned from the minimum age of 22 to the maximum age of 30 and averaged 25 years. All but one of the participants came from an academic background and had already obtained an academic degree (3 | 18.75%) or were still studying (12 | 75%). Most of the participants stated to have experience playing video games (13 | 81.25%), and two participants stated to have additional experience in either developing games or conducting game research. Regarding their playing experience, the group was very heterogenous; participants reported playing time from none to over 16 hours per week, with an average of 5.93 hours of playing per week (for a full overview, see Appendix A.2.5).

Descriptive Analysis

On average, it took participants 28,06 min (min: 17 min., max: 50 min., median 25.5 min.) to finish the survey and the first sort and 28,62 min for the second sort (min: 9 min., max: 109 min., median: 20 min). When comparing sorting times divided by set of items, it took participants on average 27,75 min. (min: 19 min., max. 50 min., median: 24 min.) to sort the first set of items and 28,93 min. (min: 9 min., max: 109 min., median: 22 min.) to sort the second set of items. Comparing the presets of items, in sorts based on preset 1 (98 items), on average, 16.26 (18.5%, min: 0, max: 68, median: 9) of the elements were assigned to no category while preset 2 (87 items) on average 9.78 (10.7%; min: 0, max: 49, median: 4) were assigned to no category (see Table 32 in Appendix A.2.6).

Looking at the framework's overall *fitness*, the MDA framework performed best out of the three, with an average of 3.6% of items being labeled as not fitting into the framework's structure, while the Octalysis framework performed worst in that regard, with almost every fourth item (22.8%) labeled as not fitting. In terms of *commonality*, the VAC framework performed best with 8.4% of elements sorted into the same category by 3 of the 4 participants of that group and 64.5% of elements sorted into different categories by all participants, while the MfP framework performed worst with a commonality of 3.7% elements sorted similarly by 3 of 4 participants and the highest number of elements (80.2%) sorted into different categories by all participants. With regard to *consistency*, the MDA framework performed best with an average delta of 8.94% of commonality between sets, and the VAC framework of Robinson et al. was second best with an average delta of 10.83%. The full overview for all assessed parameters is given in Table 1.

Table 1 – Aggregated Sorting Outcomes for Both Card Sets Over the Four Frameworks

Evaluation Parameters	MDA	VAC	MfP	Octalysis
Fitness				
Set 1	94.8%	75.3%	72.3%	83.6%
Set 2	98.0%	90.8%	97.6%	70.7%
Average	**96.4%**	83.1%	85.0%	*77.2%*
Commonality				
4 – Set1	-	1.6%	0.3%	1.8%
4 – Set2	3.9%	4.4%	-	-
Average	-	3.0%	-	-
3 – Set1	4.0%	7.4%	4.5%	6.8%
3 – Set2	7.7%	9.4%	2.8%	6.0%
Average	5.9%	**8.4%**	*3.7%*	6.4%
2 – Set1	16.2%	26.3%	17.7%	28.1%
2 – Set2	15.2%	22.0%	14.3%	19.3%
Average	*15.7%*	**24.1%**	16.0%	23.7%
1 – Set1	79.8%	64.7%	77.6%	63.5%
1 – Set2	73.4%	64.3%	82.8%	74.6%
Average	76.7%	64.5%	80.2%	69.1%
Consistency				
Standard deviation between Set1 and Set 2 for **Commonality 3**	**3.41%**	6.36%	3.73%	4.95%
Standard deviation between Set1 and Set 2 for **Commonality 2**	7.64%	**6.67%**	9.48%	9.1 %
Standard deviation between Set1 and Set 2 for **Commonality 1**	**8.77%**	9.39%	11.88%	11.23%
Average standard deviation between Set1 and Set 2	**6.48%**	7.58%	8.23%	8.43%

A full overview of the distributions for each framework, as well as the commonalities for 4, 3, 2, and single participants, is given in Tables 33-40 in Appendix A.2.6.

Looking at the categories on an individual level, Table 2 presents the categories with the highest and the lowest commonality values:

Table 2 – Comparative Overview of Framework Categories with Highest and Lowest Commonality Values

Category	Framework	Average Commonality 3/4		Category	Framework	Average Commonality 1/1/1/1	Not used
Mechanics	MfP	11%		Escapism	MfP	96%	0
Discovery	MfP	14%		Relationship	MfP	100%	2
Challenge	MDA	23%		Socializing	MfP	88%	0
Fellowship	MDA	8%		Teamwork	MfP	90%	2

Development & accomplishment	Octalysis	12%		Discovery	MDA	87%	0
Social influence & relatedness	Octalysis	23%		Expression	MDA	85%	0
General Framing	VAC	10%		Narrative	MDA	84%	0
Resources and Constraints	VAC	16%		Sensation	MDA	93%	0
Social Features	VAC	13%		Submission	MDA	100%	3
				Empowerment of Creativity	Octalysis	82%	0
				Loss & Avoidance	Octalysis	80%	1
				Scarcity & Impatience	Octalysis	82%	0
				Feedback and Status Information	VAC	85%	1
				Extrinsic Incentives	VAC	82%	1

Of the low commonality categories, eight categories emerged that were not used in at least one sort, with "Submission" standing out as not having been used by participants in 3 different sorts.

Discussion & Conclusion

In this experiment, we wanted to gain insights into the viability of using existing frameworks from game design research as labels for our dataset. Given the overall low levels of commonality, with the best performing framework achieving an overall commonality of 25% for 2 of 4 items, we conclude that none of the four tested frameworks provide suitable foundational categories for further labelings.

In evaluating the frameworks in terms of their suitability to our dataset, we find that two frameworks emerge as the most suitable to our cause, however, individually succeeding in different categories. In terms of *fitness*, the MDA framework performed best with the lowest number of items sorted into the "no category" folder. However, in terms of *commonality*, the MDA framework performed worst, featuring the highest number of categories with low commonality. Here, the VAC framework performed best, with the highest overall average commonality as well as the highest number of categories with high commonality and the lowest number of categories with low commonality. In terms of *consistency*, the MDA framework performed best with the VAC framework second. This result was not surprising, as these frameworks were the closest to the original dataset, stemming from game design and gamification research. While the other two frameworks, originating from player type and incentive research, did not perform as well, they still performed well enough to enrich our dataset with fitting connections to this type of research. This is encouraging, as it indicates that linking playing motivations to game design elements can be successfully achieved, which will result in a richer overall ontological structure if fully executed.

In terms of individual categories, overall commonalities range between 2,3% and a maximum of 20.8% per category. In the MfP framework, there appeared three categories that showed a higher item density than the others, while in the other three frameworks (MDA, Octalysis, VAC), the same thing happened with two categories. All other categories showed an even distribution (except for the "Submission" category within the MDA framework). When we look at single categories independent of their framework, we see that each category was used in at least four of the seven or eight sorts that were performed on it. One in four categories was not used by at least one of the participants; however, this overall occurrence only happened in 3.75% of all cases. Three categories stood out as performing significantly worse than others. First, the "Submission" category was not used in three of the seven sorts of the MDA framework. Furthermore, only 2,7% of items had been sorted there, making it also the category with the lowest distribution. While we can't conclusively tell if the dataset lacked elements fitting this category or if participants failed to understand its meaning, the general audience's connotation of the word "submission" might differ from the authors' understanding in relation to games. The category that performed best in terms of commonality within the MfP framework was mechanics. It is also the subcategory with the highest distribution. Interestingly, the category of the same name within the MDA framework only performed at an 8% commonality - despite having a similar distribution level of items. Through the comparison of item commonality of categories between frameworks, we thus discovered a challenge for the ontology. This can be highlighted through the example of the "Challenge" category (MDA framework). This category showed the second-highest commonality within the framework (14%), which indicates that there was a common understanding among the participants of the elements that suit this category. Furthermore, the most similar category of the Octalysis framework, "Development & Accomplishment," also had the second-highest item commonality (with 12%) within the framework. However, when looking at the two contentually, most related categories of the MfP framework: "Advancement" and "Competition," item commonality was much lower (7% and 2%). While these categories can be seen as facets of the "Challenge" category, their specificity seemed to open the space for more disagreement. These examples highlight the problem of fine-tuning the hierarchical levels for the labeling structure in affording users a high level of utility through commonly understood categories as well as their expected content. From the overall number of items that were sorted into the "no category" folder, as well as those elements that participants commonly named as difficult to place, we can see that even taken together, these frameworks do not yet cover the full range of categories that would be needed to accommodate all elements of our aggregated dataset.

To gain deeper insights into the low overall suitability, we further evaluated the two other factors influencing our experimental setup, the participants and the underlying method. In terms of the expertise of the topic domain, the participants of the experiment were representative of a target audience that has adjacent but not core domain knowledge of the contents of our ontology (a full overview of player data is given in Table 30 in Appendix A.2.5). Through the player type test, we found that almost half of our participants fell into the "philanthropist" group (7/16) and almost a third into the "socializer" group (5/16), placing three-quarters of our participants into a social-related frame when it comes to game interests. This topical preference is reflected within the sorts of the experiment – the social-related categories from each framework (except for the "socializing" subgroup in the MfP framework) all showed a high item commonality over all their sorts. This indicates that these types of categories and their related elements were best and most similarly understood by this group of experts. Furthermore, when dividing participants according to their video game playing experience, sorts conducted by respondents without video game experience showed a lower usage of "No category" with 3.85% of all elements assigned, compared to 20.51 % on the side of respondents with video game experience. This indicates that experts were less satisfied with the given categories compared to laymen. As we conducted the sorting with a mixed audience in terms of experts and laymen, further sorts with different parts of the peer group will be necessary to confirm this conclusion as well as solidify certain categories as more useful than others.

In terms of method, we found the closed card sort process suitable for evaluating and comparing the frameworks' fitness to our dataset. On average, each sort took around 28 minutes; it was thus possible to conduct two sorts per participant while staying within an acceptable overall timeframe. Most participants spent less time on the second sort (a median difference of 5 minutes), and on average, six more items were labeled as "no category." This could indicate that less focus or interest was given to the second sort – however, similarly, participants could have saved time due to being more familiar with the content matter through the first sort. When analyzing the distribution of elements over the categories within the frameworks, we found them to be relatively evenly distributed, despite the differences in categories between the frameworks. As the number of elements per category is permanently visible in the tool during the sorting, looking at the relatively even distributions, there might have been an implicit bias of participants aiming to distribute the items evenly.

In summary, none of the tested frameworks showed enough promise for us to continue to pursue this avenue of labeling our data. As the biggest concern related to the low commonality values, we realize that we underestimated the underlying multidimensionality of our data, given that so many of the participants seemed to either have a different understanding of the meaning of the items themselves or the meanings of the categories they had to sort the items into. We conclude that it would be essential to continue our process by gaining deeper insights into the varying viewpoints that underly potential classifications for our dataset.

3.3.3. Open Sorts

As the closed card sorts had not produced a useable classification foundation for our ontology, we decided to follow up with another card sort study focused on finding a suitable foundational structure by developing it from the ground up. The previous study showed that the categories of the tested frameworks were not yet fully sufficient to incorporate all the game design elements of our dataset in a consistent manner. Given the multifaceted nature of our data, we decided to gain a better understanding of its possible classification dimensions by identifying further potential viewpoints. For this, we designed the follow-up experiment as an open card sort, thus shifting our approach from a deductive to an inductive approach. We conducted this study under the following research question:

RQ2: *What underlying classification viewpoints emerge from an empiric classification of a dataset of diversely aggregated game design elements?*

Experimental Design

We prepared this experiment to consist of two sessions: the main experiment and a debriefing session to gain qualitative feedback through an open discussion with the participants, as suggested by Soranzo and Cooksey (2015). The experimental session was sectioned into three parts: I) a short survey assessing demographic factors, expertise, preferred playing devices, and modes II) the open card sort experiment, and III) a second survey assessing player type.

The first part (I) consisted of a survey assessing participants' demographics (age, gender, education, occupation) as well as their gaming experience and game design expertise (self-assessed (hobby, research, development)), the number of hours played per week, and for how long participants have been playing video games (Karle et al., 2010). In the second part (II), participants were given the link to their card sorting session. As in the previous experiment, participants were given a short overview of the upcoming steps of the experiment. To ensure comparability to the previous experiment, we chose the same preset of items as for the closed card sorting (Set1 – 98 items). Inherent to an open card sort is the lack of a predetermined structure; thus, participants would only see the "no category" folder and create the folder structure through the sorting process. We configured the tool to arrange the items in a random order for each time the link is opened to prevent sorting biases from a given order of items. Once finished with the sorting process, in part three (III), participants were asked to return to the survey to complete the experiment, where their player type was assessed via a personality test. As in the previous experiment, we used the Hexad user type by Tondello et al. (2016). This was of interest to our research as it allowed us to understand potential links between sorting viewpoints and the participants' personalities – related to a game-specific scenario, as well as to compare this group of experts to the mixed group from the previous study. We set the personality test and the device-specific questions after the sorting to prevent framing effects. The debriefing session was scheduled to be held three days after the experiment as an open discussion with a focus on our main research interest: to gain qualitative insights into the underlying structure of participants' categorizations.

As we were looking for game design experts, we connected with the GameLab Karlsruhe at the Karlsruhe University of Arts and Design (2009) and invited designers and practitioners from the lab to participate in the open card sort experiment. In total, we found eight experts that agreed to participate in the experiment. The experts were not fiscally compensated and participated based on intrinsic interest. Further information on the participant selection can be found in Appendix A.3.1.

Experimental conduct

At the beginning of the experiment, all participants were briefed on the purpose of the experiment as well as the procedure and received an instructional sheet on the basics of the card sorting tool, following the recommendations by Wood and Wood (2008). If something was unclear to them, they were instructed to raise their hand for the experimental conductor to come and help them. We gave participants a choice to conduct the card sorting individually or as a group. Two people decided to work as a group and were asked to go into a separate room for the duration of the sorting where they could have an open discussion without disturbing the others. The remaining six participants were briefed to stay in the same room. All participants were then placed in front of a computer with a digital survey already opened for them. They had been given papers and pens to note problems and ideas regarding the tool and the questionnaire during the experiment. Finally, all participants were asked to turn off their phones and during individual sorts not to communicate with each other throughout the procedure. They were then asked to start with the experiment. The first session of the experiment was conducted in one sitting with no set time limit.

Analysis

In contrast to the closed card sort, where we could evaluate the preexisting structures mostly through quantitative measures, this open card sort was meant to help us gain qualitative insights into underlying structures that originate from the dataset and suitable viewpoints that can be derived from this. While we use the factors of *hierarchical depth*, the *number of top-level categories* and the *total number of categories* to compare and cluster the structures created by the participants, our analysis is mostly focused on the rationale the participants gave during the discussion session and the resulting qualitative meta-analysis of the individual classifications to identify their underlying viewpoints.

Results

General/Descriptive

The overall experiment duration of the first session ranged from 85 to 185 minutes. The group sort was aborted by the participants after 160 minutes. At this time, the two participants from the group sort had completed their category structure but were not finished with sorting the items. They volunteered to finish their sorts individually at home. On average, participants spent 15 minutes (min: 8, max: 21) on the survey at the beginning of the experiment and 6 minutes on the player type evaluation at the end (min: 4, max:10).

Participants

The average playtime of all participants amounted from at least four hours per week up to more than 16 hours. All participants stated having played video games for more than ten years, which indicates a substantial video game experience (Karle, Watter, and Shedden 2010). All participants stated that they play at least once a week (5 participants reported playing daily), and except for one participant, all participants have experience playing all game modes (single player, multi-player offline, and online). The average age of participants was 25.75, ranging from 23 to 31; gender was evenly distributed with three persons identifying as female, four as male, and one as else. (A full overview of all assessed user data can be found in Table 41 in Appendix A.3.2) With regards to the personality test assessing player profiles, the results were unevenly distributed as 5 of our participants all fell under the Free Spirit category, two were identified as an Achiever and a Philanthropist, and one Participant equally fell under the Philanthropist and Player category. However, in several cases, the general distribution of values was relatively even, resulting in low peaks for their type-indication (specifically for participants 2, 3, 4, and 8 – see Table 42 in Appendix A.3.2). Data on the player type average of the experiment compared with the global average given on the website of the Hexad framework (Gamified UK, 2018) is presented in Table 43 in Appendix A.3.2.

Card Sorts

In terms of structure depth, participants created between two to four levels of hierarchy (depth), averaging a depth of 3.1. The number of top-level categories varied from 2 to 14, with an average of 5 top-level categories. When clustering the sorts according to the chosen approach, it emerged that participants who built their structure based on an abstract top-level chose either two or three top-level categories, while participants who based their structure on practice-oriented clusters featured between 6 to 14 top-level categories. This division between approach and outcome is further visible within the layer depth as the group that included abstract clusters had the highest number of depth levels (three participants with four layers and one with three), while the practical group displayed a low number of levels (two participants with two layers of depth and one with three). On average, 16.87 categories were developed per sort, with a minimum of 9 categories in one case and a maximum of 21 in three cases setting the median at 18.5. For an overview, see Table 3 (Visualizations of the emerging hierarchical structures of each sort are uploaded at https://gonku.de/sup-mat-phd-gho/OCS-Individual-Sorts.pdf).

Table 3 – Overview of Clustered Category Structures Emerging From Open Card Sort

ID	Viewpoint/ Process	Sort	Layer Depth	Total Layers	Ab stract	Con crete	Player Type
1	Bottom-Up	i	2	20 (14|6)	n	y	Achiever
2	Bottom-Up	i	2	12 (6|6)	n	y	Free Spirit

8	Top-Down	g	3	14 (2| 4 |8)	y	n	Philanthropist/ Player
6	Top-Down	g	4	17 (2| 4 |8|3)	y	n	Free Spirit
7	SAM/MDA	i	4	21 (3|11|5|2)	y	y	Free Spirit
4	SAM/MDA	i	4	21 (3|10|7|1)	y	y	Free Spirit
3	SAM/MDA	i	3	9 (3| 4 |2)	y	y	Free Spirit
5	Design Guide	i	3	21 (7|10|4)	n	y	Philanthropist

We clustered participants' frameworks according to similarities regarding their sorting approach as well as the resulting underlying structure based on the discussion session where participants gave rationales on their process and intention.

The *Bottom-Up* viewpoint emerged from the participants' straightforward approach to clustering similar items. Both participants whose structures fit into this cluster stated that they started their process by scanning the items to gain an overview of the dataset and then going through the preset, adding new categories whenever they came across an item that did not fit the categories they had established at that point. The sorts labeled *Top-Down* stem from those participants that started as a group but then finished individually. Both participants stated that their process consisted of developing their structure without considering the items first. Only after agreeing on the final structure did they commence with the actual sorting. The sorting structure is characterized by its symmetrical structure, where elements are divided into game "internal" and "external" elements at the top level. They are then further grouped into categories named "physical" and "psychological" (these subcategories are used for internal and external). On the lowest level, the elements are further separated into categories named "input" and "output" (these subcategories thus exist four times within the derived structure). During the debriefing, these participants further stated that they intentionally aimed for a categorization that would be significantly different from the other sorts. We named the viewpoints *SAM* (story, aesthetics, mechanics)/*MDA* (mechanics, dynamics, aesthetics) due to their similarity to two top-level classification frameworks from game theory that participants were familiar with (story, aesthetics, and mechanics by Schell (2014) and mechanics, dynamics, aesthetics by Hunicke et al. (2004)). One of the participants explicitly stated that their top layers were built on these frameworks, while the others stated that they might have been influenced by their recent lectures. While their top layer categories operate on an abstract level, they are subdivided into more pragmatic clusters deeper within the tree. This indicates that the participants of the SAM/MDA dimension used a top-down approach for their top-level categories but then followed a bottom-up approach similar to participants 1 and 2. Interestingly, while they were only explicitly used as top-level categories in three of the sortings (3, 4, 7), the framework-based categories were also represented in one of the bottom-up sortings (1), displaying a similar set of elements within. Finally, the *design guide* viewpoint was named after the efforts this participant made to structure their classification according to future usages of the game design elements. In this approach, the top layer categories were framed as questions to the game developer (e. g. "How will the gameplay and the portrayed information be structured?", "What should the gameplay be like?", "What world needs to be built for that?"). The sublayers were then structured similar to the other individual sorts, following a pragmatical clustering structure.

Apart from the top-level-based clusters, we further found meta-clusters of categories within the trees: *Areas of competencies*: ("Music," "World Building," "Storytelling"), *player perspective* ("Act of Play," "Player Behavior," "Mediated Player Interaction," "Player Internals," "Experience"), *Design Process* ("Parameters," "Objectives," "Actions," "Pacing"), *Design Choices* ("Incentives," "Framing/Presentation") and *Abstract* ("Physical"/ "Mental," "Input"/ "Output").

Certain elements were always grouped similarly – one such emerging group was the five economy principles (*simple bartering, complex bartering, simple market, complex market, meta economy*) by Fullerton (2008) and labeled "economy" by all participants (except for one member of the group sorting that did not deviate from their original structure). Similar grouping also happened around the multiplayer modes (*team competition, cooperative play, multilateral competition, unilateral competition, player versus player* (Fullerton, 2008); however, we observed a lot of variation regarding the labeling: "Competition," "Game structure," "Multiplayer," "Two and more players," "Player interaction patterns," or "Multiplayer modes." Only one participant used the same label for these elements as Fullerton ("player interaction patterns"). Furthermore, the participant that had used the label "Multiplayer" deviated from grouping these elements by exempting the element "Single-player versus game" from their category. Finally, elements that had been sorted by the SAM/MDA group into the "aesthetics"-related category were found in similar categories of the other participants, labeled "Content," "Setting," "Presentation," or "Visualization."

Discussion and Conclusions

In terms of our research question, we gained several relevant viewpoints to consider for future use within the ontology, as well as valuable feedback on the perceived usefulness of a singular, unified taxonomy. We were able to identify four distinctly different viewpoints emerging from the different approaches that participants took towards their categorization process (bottom-up/emergent, top-down/abstracted, SAM/MDA, design guidance) as well as three additional viewpoints through a qualitative meta-clustering of the final categories (designers' perspectives (areas of competences, design process), player perspectives and incentive perspectives).

Overall, we were surprised at the diversity of approaches taken by our participants. The most intuitive approach taken by two participants was to sort items into a naturally emerging structure. Interestingly, all other participants chose to structure at least part of their process by first undergoing a rational process devising initial categories. While three participants built on preexisting structures from game research (SAM/MDA), one participant approached the task through a preexisting premise: to build a structure that can serve as a support tool for design processes. Finally, the two participants working as a group went for the most abstract approach, building a structure that was solely made from dichotomous values (internal vs. external, physical vs. psychological). In comparison to the other viewpoints, the top-down structure developed in the group sort did not work as well with the contents of the dataset. This could be seen in significantly higher use of the duplication function, the higher density within certain categories (participant 8 sorted 66 elements into a single category- making it the category with the highest density), as well as the time it took to complete the sort(s). However, we think that the idea behind the developed structure, to break down each value of interest into dichotomies and thus afford clear separation of the items, is interesting and should be considered for future incorporation, albeit outside of a hierarchical structure.

Through comparison of element clusters within the structures, we saw an inherent structure emerging from the given elements, where even the groups with more abstract top layers built more pragmatic structures in the lower levels of their classifications. An overall hierarchy could thus likely be achieved through a dedicated merging process, where for example, either the abstract layer is dismissed from the SAM/MDA classifications or the bottom-up clusters are sorted into a unified, overarching abstract layer. However, during the debriefing session, participants worded strong doubts that browsing the ontology in the form of a folder-based categorization of game design elements as offered through our tool would help them in game design and development. The suggestion emerged that instead of using the categories hierarchically, a tag-based approach combined with a filter function would be considered useful, as this would allow for the different viewpoints to exist in parallel. Furthermore, when the discussion turned towards the direction of discussion of different options for presenting the final structure, participants stated that they would deem it helpful to see how the elements are connected in contrast to the folder-based separation they were using during the card sorting.

In terms of methodology, we were satisfied with the results produced by the open sort and the tools' features to facilitate the process (a summary of the tools' usability evaluation and suggestions derived through the post-experimental session can be found in Appendix A.3.3). Particularly the discussion and evaluation session allowed us to better understand the thought processes of our peers and gain insights into the varying approaches they took to arrive at their final classifications. It needs to be noted, however, that we underestimated the qualitative difference in terms of sorting times between closed and open card sorts. Only one of the participants managed to sort all items in a matter of an hour (this specific sort had only a two-layer-depth with 12 categories created overall). Given the overall high duration of all sorts, we learned that for high-quality sorts to be conducted in a reasonable amount of time, 98 items are too high a number (at least in terms of our content matter) and should be systematically reduced in future experiments. We conclude that, while effective, the process of open card sorts is time-consuming and effortful for the participants and not easily scalable without being able to automate the attachments of categories to elements while preventing redundancies. Thus, the development of a crowd-based aggregation process would be an important next step.

Outlook

In summary, with our overall goal to generate one satisfactory game design element ontology, we conclude that the multifacetedness of each element makes it too difficult for us to generate one satisfactory structure. Through the outcomes of our studies and the discussion session concluding the open card sort, we realized that the hierarchical outcome structure resulting from the card sort process is detrimental to our goal of providing a tool with which different user groups can make informed and inspired decisions. Given the overall number of different structure outcomes over these two sorts, most users would likely fail to find suitable game design elements if the final ontology offered only one of these viewpoints.

By consolidating the input gained from the qualitative analysis as well as the group discussion, we conclude that the focus of our research efforts should be shifted from aiming for the creation of an overall comprehensive categorization toward a multidimensional interactive structure that would afford navigating elements in terms the different viewpoints that might be attached to them. If we think about viewpoints as sets of label clusters that are attached to each element, we can use smaller sets of viewpoints that, instead of evaluating their fit regarding the overall structure, could be evaluated regarding their benefit for a certain use-case or group of people. By attaching viewpoints only to those elements they fit, every element of the dataset would reflect the different needs and perspectives of the users in a more precise and diverse manner.

3.4. Dataset Conflation & Enrichment

While the card sorting methodologies' strength is its close orientation to the human viewpoint, its weakness lies in scaling the process to larger amounts of data. Given that this limitation is inherent to human-based classifications and with our datasets amounting to almost 600 entries, we wanted to explore automated methods for clustering and linking datasets. Due to its semantic structure, we designed the follow-up studies around an algorithmic methodology, matching the extracted keywords of our dataset based on identical keywords to arrive at an enriched ontology through linked data. The following chapters report on the results of two studies where we applied this method. In the first study, we compare the connections of entries within the dataset of playing motivations with each other to evaluate the dataset in terms of consistency, outliers, and potential gaps. Through an explorative keyword clustering, we further explore overarching thematical clusters around the topics of social factors (relatedness), achievement, exploration, and emotions with high consistencies between linked nodes. In the second study, we connect a second dataset of human needs to the playing motivations dataset to evaluate their compatibility with each other and use another explorative clustering to identify categories yet missing from playing motivations literature.

3.4.1. Keyword-based Clustering of Playing Motivations

Our main goal is to enrich our initial dataset of game design elements with relevant data for theory and practice; thus, we wanted to explore possibilities to connect findings from player type and playing motivation research back to related game design elements that might inspire the àdequate target audience to the desired action. For this, we considered different automated methodologies that would allow us to create meaningful connections between these datasets. Given the text-dominant logical structure of our dataset (a title and description plus author metadata), we were looking for semantic-based algorithms. In the domain of computer science, keyword matching algorithms have proven to be efficient tools for linking semantic information (Devanur & Hayes, 2009; Uthayan & Anandha Mala, 2015). Its most prominent use-case is in search engine optimization (Cahill & Chalut, 2009) and user-relevant advertisement. In the case of Google's Advertisement algorithm, users' search queries are matched with potentially relevant results ads through keyword match types that allow for varying degrees of precision (the "broad match" category reaches more but less focused user groups, building on a loose semantic connection, while the "exact match" category links query based on the same meaning or same intent of the specific search term (Google Ads, n.d.). This type of algorithm affords identification of connections between items as well as evaluation of their respective strength. Furthermore, it allows the evaluation of datasets qualitatively and quantitatively in terms of their underlying graph structure. As this method perfectly suits our aims toward an interconnected ontology, we chose to build on this methodology. Current development tendencies are leaning towards complex underlying graph structures that embed keywords' meanings on a vector space to identify semantic closeness between elements (Mikolov et al., 2013). This is particularly necessary where user input is at play – given the vast differences in how users formulate queries and generate input (Lucas & Topi, 2004).

In our case, however, we are building on datasets derived from theory building on more similar input structures. Building on our learnings from the card-sorting process, where participants suggested the use of labels to comprehensively choose suitable elements, we wanted the algorithm to remain transparent to the end-user in terms of its process and results. We thus chose to build an algorithm based on a simplified variant of keyword matching where we directly link items based on their identical matching keywords. To explore this approach, we first applied it to compare entries on the player motivations dataset itself. The dataset's entries stem from different sources with slightly varying perspectives (playing motivations, playing incentives, player types). They also share a certain degree of overlap as some of the sources build on others, extending as well as omitting aspects. As such, we conclude this to be a good foundation for testing the potential of the algorithm we developed. Our main goal for this first study was to test the viability of the developed algorithm in terms of its ability to connect items in terms of their similarity and relevance to each other and to detect overlaps as well as outliers. We built our analysis around the following research questions:

RQ1: *How well does semantic matching of keyword-based metadata identify meaningful connections between similar playing motivations?*

RQ2*: What overarching thematic motives be identified?*

Building on the methodology of network analysis (Borgatti et al., 2009; Freeman, 2004), we analyzed the resulting overall graph in terms of its size, density, degree centrality, isolated nodes, and outliers (low degree nodes). Our results showed that the algorithm worked very well in linking similar and related items – despite its fundamental limitations regarding its keyword dependence (some items only offered very little descriptive text to draw keywords from, resulting in high variability between matchable keywords and thus skewed degree centrality towards items with more expansive descriptions to draw from). Our evaluation found several sensible thematic clusters that reflected the underlying theories many player-type frameworks are built on as well as outliers that highlighted tendencies for certain topical omissions in playing motivation research.

Preparation & Analysis

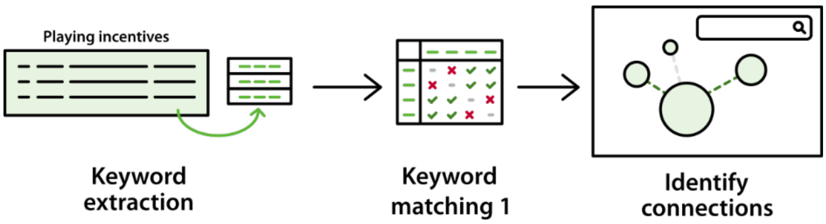


Figure 19 – Process Overview Keyword Matching 1

We started the process by extracting keywords for each element in the dataset based on their title and descriptions. The process was done by hand by analyzing the title as well as each sentence of the description and extracting the defining keyword(s). The resulting list was then re-checked twice by two members of the research team. Next, we first performed a two-step normalization process on the keywords. First, using the Levenshtein-Distance (Levenshtein, 1966), we calculated a matrix of pair-wise edit distances. We analyzed all pairs with low, albeit non-zero, edit distances to identify misspellings and differences in conjugation. After investigating the edit distances, we decided to employ a transformer-based machine learning model from the spaCy toolkit (Honnibal, 2015) to further normalize the keywords. To do this, on each keyword, we performed the lemmatization transformation[10] provided by the toolkit. This enabled us to then match

[10] Lemmatisation in computational linguistics is the algorithmic process of determining the lemma of a word based on its intended meaning (e. g. lemma for the word "better" is "good" while the lemma for the word "walking" is "walk") is the base form for the word "walking", and hence this is matched in both stemming and lemmatisation.

keywords by testing for simple string equality. Depending on the length of the description, the number of extracted keywords varied from two to twenty keywords per element.

The keyword matching algorithm compares each item (node[11]) to each other node in the dataset and reports *edges* based on common keywords. The process results in an undirected graph built on nominal data. We analyze the overall graph by identifying its *order* (number of nodes), *size* (number of edges), and *density* (ratio between the edges present in a graph and the maximum number of edges that the graph can contain) as well as its relevant actors by assessing each node's *degree* (number of other nodes it connected to, see Figure 20), their *closeness centrality* (average farness (inverse distance) to all other nodes (1 / sum(distance from u to all other nodes)), see Figure 21) and the *connection strength* between two matching nodes (average number of keywords that matched between two nodes, see Figure 22).

Figure 20 – Degree

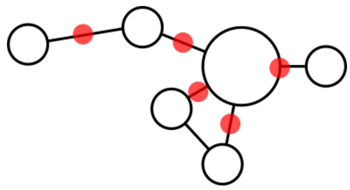


Figure 21 – Closeness Centrality

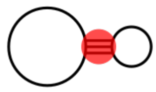


Figure 22 – Connection Strength

We further conducted a qualitative analysis to identify thematic clusters as well as outliers by scrutinizing the extrema of the dataset: nodes with the highest and lowest degrees and dyads (node-pairs) with the strongest connections. We also analyzed and clustered the underlying keywords in terms of their frequency of appearance within the dataset to gain a more detailed perspective of the underlying motives. To ensure the consistency of the emerging connections, we further conducted in-depth analyses of a set of exemplary nodes to assess their semantic consistency with their connected nodes. Finally, we conducted an explorative clustering with algorithms employed by the graph-tool Cytoscape (Institute for Systems Biology, 2002). An overview of the process can be seen in Figure 19.

Results

In terms of descriptive evaluation of the resulting graph, its *order* (number of nodes) is 234, of which 230 are connected to at least one other node in the same list. Its *size* (number of edges) is 2808 of 27261 maximum possible number of edges. We find an *average degree* of 24 (min: 0, max: 78, median: 21). The *density* of the graph is 10.30%. In terms of relevant actors, when looking at the *degree,* we find eighteen nodes connected to 50 or more other nodes within the dataset and four isolated nodes (the nodes with the highest and lowest degree values are depicted in Table 4, the full table can be found in the Supplementary Materials https://gonku.de/sup-mat-phd-gho/PvP-Degree.xlsx).

Table 4 – Overview of Player Type/Playing Incentive Nodes with the Highest and Lowest Degrees

Author	Playing Motivation/ Player Type	Degree		Closeness Centrality	Avg. Rel. Con. Str.*
		Total	Average		
Marczewski (2015)	Achiever (intrinsic)	78	33.91%	56.13%	24.1%
Tseng (2011)	Social gamer	66	28.33%	54.89%	19.0%
Lazzaro (2004)	The People Factor	63	27.04%	54.63%	23.4%

[11] To stay consistent with graph analysis nomenclature, we will refer to the items of our dataset as nodes within the graph for the purposes of this study.

Bateman, Boon (2005)	Participant	62	26.61%	54.36%	19.7%
Lazzaro (2004)	Altered States	62	26.61%	54.10%	22.1%
Taylor (2003)	Team Sport and Combat	62	26.61%	53.59%	24.6%
Lazzaro (2004)	Easy Fun	60	25.75%	53.97%	25.6%
Marczewski (2013)	Relatedness	58	24.89%	53.84%	21.9%
Yee (2006)	Role Playing	57	24.46%	54.23%	24.4%
Tseng (2011)	Aggressive gamer	57	24.46%	51.62%	22.0%
Jacobs, Ip (2003)	Hardcore gamer	56	24.03%	53.21%	19.3%
Sherry et al. (2006)	Social Interaction	54	23.18%	52.46%	26.6%
Marczewski (2013)	Mastery	53	22.75%	52.46%	29.5%
Kahn et al. (2015)	Socializer	52	22.32%	52.22%	21.1%
Yee (2006)	Relationship	51	21.89%	52.46%	22.2%
Yee (2002)	Immersion	51	21.89%	52.10%	25.3%
Marczewski (2015)	Socializer (intrinsic)	51	21.89%	51.86%	22.5%
Whang, Chang (2009)	Community-oriented player	50	21.46%	52.10%	15.8%
Lazzaro (2004)	Disgust	0	0.00%	0.00%	0.00%
Griffiths (1995)	Good sound effects	0	0.00%	0.00%	0.00%
Utz (2000)	Skeptics	0	0.00%	0.00%	0.00%
Voiskounsky et al. (2005)	Recreational refreshment	0	0.00%	0.00%	0.00%
Marczewski (2015)	Disruptor	1	0.43%	23.67%	25.00%
Jansz, Tanis (2007)	Enjoyment	1	0.43%	23.87%	100.00%
Whang, Chang (2009)	Discriminative	1	0.43%	31.61%	25.00%
Griffiths (1995)	Beeing Good at Playing	1	0.43%	37.89%	33.30%
Griffiths (1995)	Nothing Else to Do	2	0.86%	29.27%	50.00%
Callois (1961)	Alea (Chance)	2	0.86%	31.13%	29.20%
Griffiths (1995)	Can't Stop Playing	2	0.86%	32.34%	33.30%
Fullerton (2008)	Director	2	0.86%	32.38%	20.00%
Drachen et al. (2009)	Pacifists	2	0.86%	33.74%	25.00%
Stewart (2011)	Externals	2	0.86%	33.74%	33.30%
Griffiths (1995)	Violence	2	0.86%	35.44%	100.00%
Griffiths (1995)	Favourite Sporting Activity	2	0.86%	35.06%	50.00%
Voiskounsky et al. (2005)	Cognitive Stimulation	2	0.86%	37.02%	33.30%

* Relative Average Connection Strength

Dyadic Analysis

The distribution in terms of the number of matching keywords is shown in Table 5:

Table 5 – Distribution of Node Pairs by Number of Matching Keywords

Number of Node-Pairs	Number of Matching Keywords	Average Relative Connection strength	Percentage of total Pair- possibilities
1	11	92.30%	0.0037%
1	6	55.00%	0.0037%
4	5	45.79%	0.0147%
33	4	40.38%	0.1211%
106	3	35.41%	0.3852%
360	2	25.95%	1.3206%
2304	1	17.00%	8.4516%

The dyads with the strongest connections (5 or more matching keywords) are shown in Table 6:

Table 6 – Overview of Node Pairs (Dyads) With the Strongest Connections

Name A	Source A	Name B	Source B	Common Keywords		Avg. Rel. Con. Str.
				strings	Amt.	
Manager	Bateman (2005)	Achiever	Bartle (1996)	achiever, efficiency, achievement, point, hierarchy, set, mastery, gather, action, status, goal	11	92.30%
Immersion	Yee (2002)	Role Playing	Yee (2006)	story, play, role-playing, immersion, fantasy, role	6	55.00%
Community and Socialization	Taylor (2003)	Socializer	Kahn et al. (2015)	friend, play, socialize, relationship, partner	5	33.15%
Relatedness	Marczewski (2013)	Socializer (intrinsic)	Marczewski (2015)	interaction, relatedness, socialize, other, connection	5	41.65%
Immersion	Yee (2002)	Fantasy	Jansz, Tanis (2007)	believe, play, role-playing, fantasy, role	5	62.50%
Networker	Marczewski (2015)	Socializer (intrinsic)	Marczewski (2015)	relatedness, socialize, other, network, connection	5	45.85%

* Average Relative Connection Strength | Amount

The full tables of all dyads can be found in the Supplementary Materials https://gonku.de/sup-mat-phd-gho/PvP-Dyads.xlsx).

Keyword Analysis

The dataset contains 1626 keywords in total, chosen from 754 distinct keywords. Of these distinct keywords, 283 occurred with between 2 (124x) and 29 (1x) nodes of the dataset (see Figure 23).

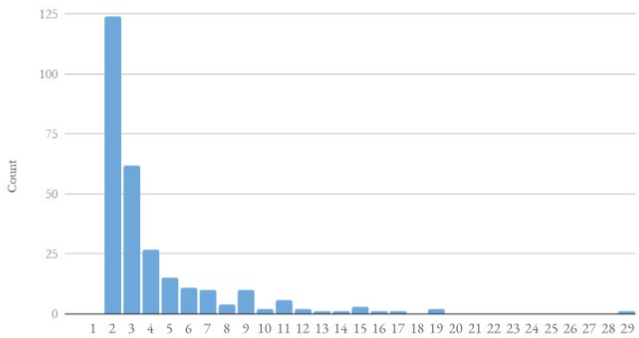


Figure 23 – Most Frequently Occurring Keywords (>9)

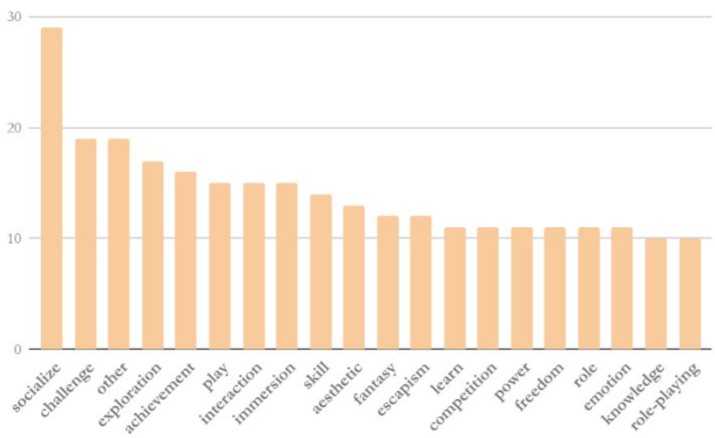


Figure 24 – Distribution of Keywords Occurring More Than Once in the Dataset

The keywords that occurred most frequently in the dataset with ten or more nodes containing them are listed in Figure 24, with "socialize" emerging as the most frequent.

Exemplary Node Analysis

To gain detailed insights into the graph's structure, we conducted a deep-dive analysis on a sample of exemplary nodes. For this, we chose the four nodes containing the player types by Bartle (1996) (Killer, Achiever, Socializer, Explorer), as they represent a very well-known and well-researched set of player types with clear descriptions. For this, we first extracted nodes from the overall dataset that were connected to each of the four player types via one or more keywords. Next, we assessed if the four player types themselves were pairwise connected. This was not the case. However, when looking at intermediate nodes (nodes that two or more of the player types connect to), all four types were interconnected by at least one or more common nodes (intermediate connectivity) (see Table 7):

Table 7 – Pairwise Connections Between Bartle's Player Types in the Dataset

Type: Absolute Connected Nodes (avg intermediate connectivity to the other types)	**Killer** Absolute Number of Intermediate Nodes (Average Intermediate Connectivity)	**Achiever** Absolute Number of Intermediate Nodes (Average Intermediate Connectivity)	**Socializer** Absolute Number of Intermediate Nodes (Average Intermediate Connectivity)	**Explorer** Absolute Number of Intermediate Nodes (Average Intermediate Connectivity)
Killer: 36 (33.3%)	0	7 (19.4%)	3 (8.3%)	2 (5.6%)
Achiever: 39 (41.0%)	7 (18.0%)	0	5 (12.8%)	4 (10.3%)
Socializer: 47 (27.7%)	3 (6.4%)	5 (10.6%)	0	5 (10.6%)
Explorer: 42 (26.2%)	2 (4.8%)	4 (9.5%)	5 (11.9%)	0

Finally, we evaluated the names and meanings of the connected nodes to assess the algorithm's effectiveness in producing meaningful results. An overview of the full lists of connected nodes in comparison to the item's descriptions is given in Tables 44-47 in Appendix A.4.1, as well as additional explorative analysis of meaning generation through the connected nodes.

Explorative Clustering

Finally, in an effort to identify overarching motifs, we conducted an explorative preliminary semantic clustering process. For this, we extracted all keywords with high frequencies (appearing in 5 nodes or more). We chose this threshold as they approximately represent the upper 10% (70/754) of all keywords in the dataset and are still manageable in terms of manual sorting.

We first conducted a visual, map-based sort, where the keywords were laid out on a 2D map and then clustered according to perceived sematic similarity relating to overarching concepts of motivations. We started with the keywords with the highest frequency and then rearranged the words and clusters according to each new keyword introduced to the structure (see Figure 25).

Next, we transferred the results into a table-based structure (see Table 8), where we refined the resulting overarching themes in terms of logical substructures. The process was conducted hand, and the results were co-checked and refined by two independent members of the research team.

The three biggest groups that emerged (in terms of number and frequency) relate to *social* factors (with two subgroups divided into collaborative and competitive social interaction motives), *achievement-*based factors (with two subgroups differentiating player and system-related factors), and *exploration-based* factors (with three subgroups divided into system-, story- and self-exploration). Two smaller groups emerged around *escapism* and *creative* factors of playing motivations. We further identified a separate cluster around the theme of *emotional* factors and finally created one group ("*else*") to contain overarching keywords (play, personal, gaming) that did not directly fit any of the other themes.

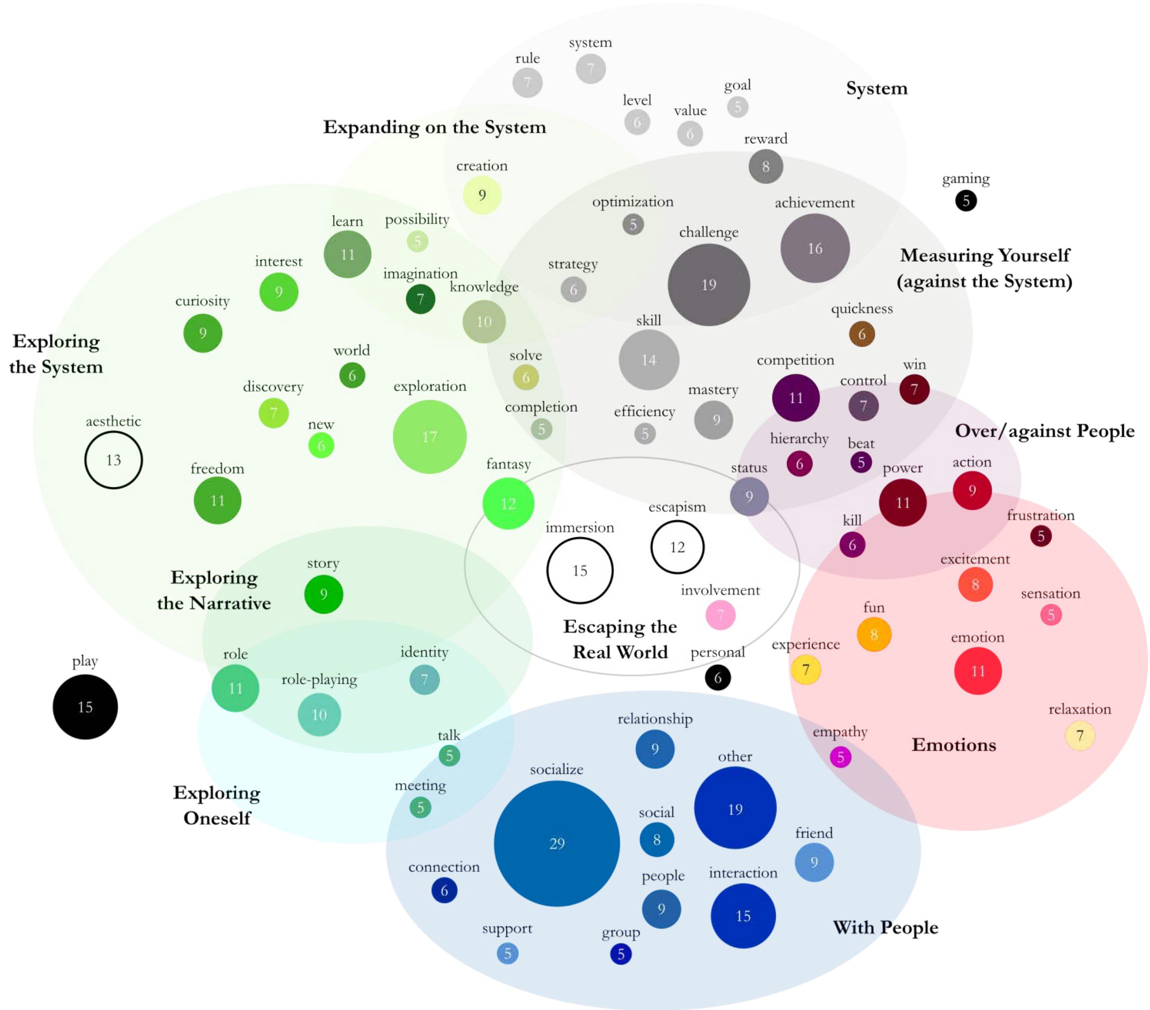


Figure 25 – Explorative Clustering of Frequent Keywords Based on Emerging Themes

Table 8 – Explorative Preliminary Semantic Cluster Structure of Most Frequent Keywords in the Dataset

Social Factors (Interacting)		Explorative Factors (Exploring)			Achievement Factors (Achieving)		Escapism Factors (Immersing)	Creative Factors (Creating)	Emotional Factors (Feeling)	Else
With People	Against People	The System	The Narrative	Oneself	Player-related	System-related				
socialize (29)	competition (11)	exploration (17)	fantasy (12)	*role* (11)	challenge (19)	reward (8)	immersion (15)	creation (9)	emotion (11)	*play* (15)
other (19)	power (11)	*aesthetic* (13)	story (9)	role-playing (10)	achievement (16)	rule (7)	escapism (12)	imagination (7)	action (9)	*personal* (6)
interaction (15)	status (9)	learn (11)		identity (7)	skill (14)	system (7)	involvement (7)		fun (8)	*gaming* (5)
relationship (9)	kill (6)	freedom (11)			mastery (9)	level (6)			excitement (8)	
friend (9)	hierarchy (6)	knowledge (10)			control (7)	value (6)			experience (7)	
people (9)	beat (5)	curiosity (9)			*win* (7)	goal (5)			relaxation (7)	
social (8)		interest (9)			strategy (6)				sensation (5)	
connection (6)		discovery (7)			solve (6)				frustration (5)	

talk (5)		new (6)			quickness (6)					
empathy (5)		world (6)			completion (5)					
meeting (5)		*possibility* (5)			optimization (5)					
group (5)					efficiency (5)					
support (5)										
129	48	95	21	28	105	39	34	16	60	
177		144			144		34	16	60	

**Cursive keywords* mark words that emerged as ambiguous and could be sorted into different categories depending on the context

Assessing the overall frequency of keywords relating to the different themes, we find that the distribution is skewed towards social-related motivations /player types, with a total of 177 of the most frequent keywords falling under that category (13 unique collaborative keywords ("with people"), 129 in total; and six competitive keywords ("against people"), 48 in total) (see Table 8). The other two dominant clusters, reflecting explorations and achievement-based factors, are evenly distributed with a total of 144 keywords each. However, the exploration cluster is divided into three subclusters, while the achievement-based cluster only features two.

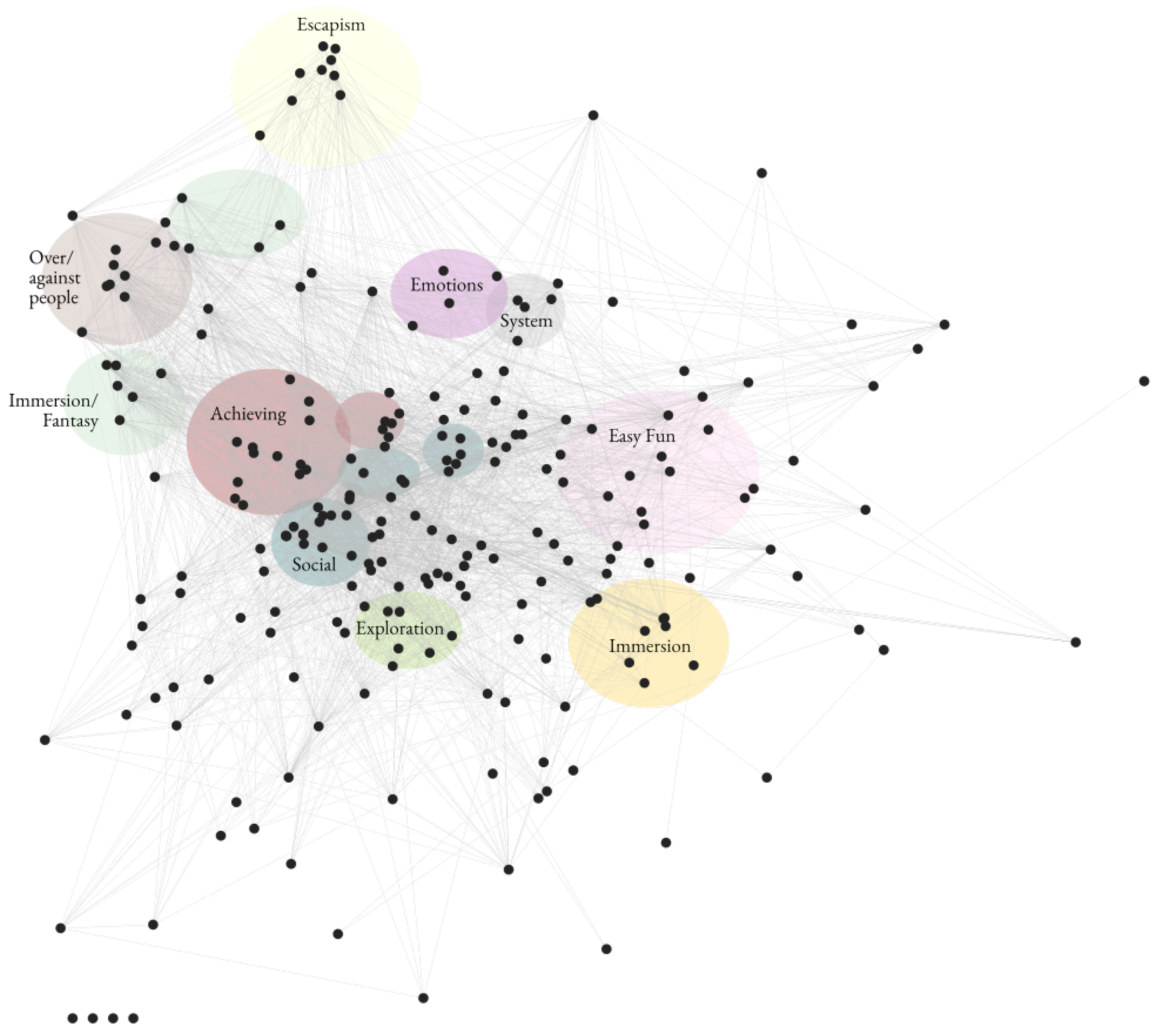


Figure 26 – Inverted Self-organizing Map of the Dataset With Marked Clusters

In a final explorative clustering approach, we calculated the full graph in the graph-analysis tool Cytoscape (Institute for Systems Biology, 2002) and analyzed the result of an Inverted Self-Organizing Map-Layout Algorithm. The results of this automated approach produced clusters around similar themes. After plotting the graph, we used the tool Inkscape (Gitlab.com/inkscape/inkscape, 2003) to mark emerging Clusters according to meaningful motifs (see Figure 26). Detailed images of the marked regions are listed in Appendix A.4.2 (Figures 79-86).

Discussion and Conclusions

With regard to our first research question, on how well semantic matching of keyword-based metadata can be applied to identify meaningful connections between similar playing motivations, our results show that the methodology of connecting nodes via their matching keywords proved to be able to produce meaningful connections. This is particularly visible when looking at the node pairs with the highest number of matching keywords, which are very consistent, often even matching by name. Even when looking at nodes with few connections, we find equally meaningful matches (e. g. “Violence” connects to “Hardcore Gamer” and “Fighter”). Also, exemplary analysis of representative nodes showed high thematic consistency, as could be seen in our analysis of Bartle’s player types, but also with other sample nodes (e. g. “Just Enjoyment” connecting to: “Enjoyment,” “Game being fun,” Enjoyment of different lifestyle” and “Enjoyment”[12]).

Concerning our second research question on what overarching thematic motives can be identified through this process, we were able to identify overarching, recurring clusters through various methods that represent current perspectives of player type and playing motivation research. This highlights the homogeneity and strong relationship between these two adjacent fields. However, it is important to mention that the results and any conclusions drawn from this first analysis (particularly the clustering) underly several biases that relate to the dataset and its collation process. First, the entries of the dataset consist of studies that, in part, build on each other. For example, Zackariasson et al. (2010) build on the motivations of (Yee, 2006), who himself builds his framework on the player types of Bartle (1996) (as explicitly and implicitly many of the other studies do – given the prominence of his framework). We decided to list all player types and playing motivations separately in the original dataset, as each of the authors brings a different perspective and dimension to their framework while also adding new facets, motivations, and types. However, this choice leads to a distortion and potential inflation of certain clusters; thus, the resulting cluster sizes should be considered as tendencies, not representative values. Also, our clustering process was likely influenced by our own prevalent structures emergent from a previously formed understanding of playing motivations, in large part built from the same frameworks. Finally, most of the frameworks we used in our dataset were extracted from games and behaviors mostly prevalent in western cultures. This is due to the nature of our collation process, building on citation numbers, forward- and backward search, and the focus on English publications. While we included literature that specifically aimed to include concepts from other cultures[13], it is unlikely that our dataset reflects internationally representative cluster sizes. However, such underlying biases must be addressed at the dataset collation level and do not take away from the effectiveness of the methodology to analyze and relate the existing data. Also, the main clusters we found (socializing, exploring, achieving,

[12] Several nodes are named similarly but originate from different authors

[13] e. g. "Kvell": the feeling of expressing pride in one’s child or mentee to others, "Fiero": the joy of personal triumph over adversity, "Schadenfreude": the gloating over the misfortune of a rival (Lazzaro, 2004)

creating, escaping, feeling) can be related to the most foundational instincts of human beings (Seeking, Rage, Fear, Panic (Sorrow/Distress), Lust, Care, Play) as listed by the neuroscientist and psychobiologist Panksepp (2004). Thus, while the emerging clusters likely are not entirely representative in terms of real-life prevalence concerning their sizes and distribution, they still represent essential foundational incentives for play. We conclude that the methodology is sufficient towards our goal of efficiently enriching our datasets as it affords meaningful connections between related items with little error and the automated creation of meaningful hierarchical metadata.

Through our analysis, we were also able to evaluate the underlying dataset in terms of its diversity and completeness. We find that while our list contains a large number of items, it not only holds many redundancies (which is unavoidable given the underlying methodology) but also implies gaps in player type/playing motivations research. This becomes most visible when we analyze the isolated nodes:

First of the four, "Disgust" (keywords: feces, blood, urine, disgust, mucus, vomit, rejection) is one of a list of emotions (Lazzaro, 2004) extracted from analyzing player's facial expressions and body language while playing. While this emotion is not as intuitively classifiable as a motivation for play, it is an emotion emerging during involved play and one that does not deter players from repeated participation in games that evoke it. The change of physical state is a strong motivational factor for many entertainment media (e. g. fear and shock in horror movies or haunted houses or panic-like adrenaline rush in rollercoasters or gambling) (Niles, 1977). A node that reflects this motivation and thus logically should have been connected is the "Altered States" incentive. The lack of a common keyword between these two nodes thus highlights the limitations of the keyword methodology. In summary, the lack of connections for this node is not an indicator of this motivation not belonging to the dataset, but more a reflection of the limitations of our methodology as well as an overall scarceness of visceral and negatively connotated playing motivations within our dataset.

The second unconnected playing incentive, "Good sound effects" (keywords: sound, sound-effect, music, good sound effects), as listed and analyzed by (Griffiths & Hunt, 1995), represents a specific, sensation-related playing incentive – which is also underrepresented within the current dataset. While this specific incentive could be subsumed under the more general "Aesthetics" motivation, its lone emergence still points toward a phenomenon where most motivations relate to cognitive and emotional incentives in contrast to visceral experiences.

The third item that did not match with any other item was "Skeptics" (keywords: disinterest, skeptics, skepticism). This specific player type stems from a publication focusing on the development of friendships in MUDs (Utz, 2000). This item is an indication of completely different research potential regarding player types: looking at the audiences that are related to games but stand outside of the magical circle –users with aversions and reserve towards game and play. Thus, while belonging to the overall cosmos of player types, it does not directly fit the overall dataset of explicit motivations and thus is a comprehensible outlier.

The last item, "Recreational refreshment" (keywords: refreshment, recreationality), relates to a general emotional and sensational incentive for playing. It was surprising that this item was not matched to any of the clusters around "Escapism" or "Easy Fun." Especially since the term itself consists of two incentives that (independent of each other) represent strong incentives for what is generally sought after within any entertainment or leisure activity. This case most strongly reveals the limitations of our keyword-based algorithm and shows the necessity to further improve our process by using newer, vector-based natural language understanding algorithms that are focused on semantic meaning rather than specific words.

Looking at the other items with few connections, we most dominantly find three potentially underlying reasons. The first directly relates to their low numbers of keywords (e. g. "Enjoyment" connects to "Just Enjoyment" through its one existing keyword: "enjoy," highlighting the methodological limitations with regards to our keyword extraction process). The second theme relates to the respective items indicating more unspecific, undirected, or involuntary reasons for playing ("being good at playing" (1), "Nothing else to do," "Can't stop playing"). The third theme seems to center around the items being negatively loaded (e. g. "Discriminative" (1), "Disruptor" (1), "Violence" (2)), showcasing a high number of negatively connotated keywords: (*boredom, violence, separation, discrimination, uncertainty*). This last theme, in particular, seems indicative of a tendency in playing motivation research to lean towards more optimistic terms and types and focus less on the negative player type-related counterparts or reasons *not* to play.

In summary, despite limitations relating to the methodological process (the number of extractable keywords depending on the quality and extent of the item's description and the matching algorithm's need for identical keywords) and relating to the biases that emerged from the underlying dataset (cultural and systemic limitations and tendencies towards positively framed items), our different analyses not only show the general methodology to work sufficiently well towards our goals of identifying meaningful connections between thematically related datasets but also allow for qualitative identification and evaluation of gaps in the current state of the art.

3.4.2. Ontology Expansion through Human Needs

While successful in terms of methodology, the findings of our previous study highlighted certain limitations concerning the underlying dataset of playing motivations. Particularly the isolated nodes showed thematical components that seemed underrepresented in the current dataset, indicating that the dataset might be unevenly distributed and even incomplete.

We start our inquiry into the origins of this problem by looking at the underlying methodologies by which the respective authors originally constructed their player type and playing motivation frameworks. One of the most common methodologies for taxonomy development in this field is built on the analysis of empirical in-game data, which is extracted by observing and clustering playing behavior within a specific game or game genre (e. g. the player types by Bartle (1996), who derived them from the analysis of player's behaviors in MMORPGs, or Jansz and Tanis' (2007) survey-based evaluation of players of online first-person shooters or Drachen et al. 's (2009) player automated analysis of playing behaviors in "Tomb Raider: Underworld"). This empirical data is then processed via different follow-up methodologies. Bartle (1996) expands on his empiric methodology by developing a typological model building on two underlying diametric parameters (behavior: "acting on" vs. "interacting" and surrounding: "players" vs "world") resulting in his typology of four distinct categories of player types while e. g. Drachen et al. (2009) expand on their empirical data using an automated clustering algorithm (emergent self-organizing maps) from which four dominant clusters emerge. Independent of follow-up framework extraction, this approach is always limited by the underlying games' or genre's preset design configurations; thus, player types and playing motivations derived through this method will be limited in terms of achieving or approximating completeness. This can be seen in the high variability of the number and type of the identified player dimensions that emerge from the aforementioned studies. As this type of framework is the most prevalent in our dataset, its limitations are most influential.

Another methodological approach that we find in the studies underlying our dataset is to start from preexisting research on player types. Several of the frameworks underlying our dataset build on the Bartle player-type model as a starting point for their models, as can be seen with the approach by Yee (2006), who used Bartle’s research to develop a 40 questions survey taken by 3000 MMORPG players, on which he conducted a factor analysis of survey data, resulting in a model that incorporates factors of age, gender, usage patterns, and playing behaviors. Another model incorporating the Bartle framework is the Hexad model identified by Tondello et al. (2016). While building on player-type research, their framework is rooted in the literature on human motivation (specifically the three core intrinsic motivations by Ryan and Deci (2000), complimented by Pink’s Drive theory (2011) and the implicit inclusion of Sheldon's (2011) model on ten candidate psychological needs) and then expands on Bartle’s player types resulting in a model with three diametrical axes (extrinsic vs. intrinsic rewards, socially extrinsic and intrinsic rewards, and the exertion of creativity within the boundaries of the system vs. the exertion of creativity towards a change the system itself). Another approach using psychometric models is taken by (Nacke et al., 2014), who developed their BrainHex model based on neurobiological findings, then assessed and adapted it through a large-scale player-based survey. Their research specifically focuses on potential relationships between personality types and their model. In contrast to the game-based taxonomies, these models are not limited by preexisting design configurations. In summary, while some of the authors did not expand onto other fields, limiting their theoretic foundation to their domain (e. g. Sherry et al. 2006; Yee 2006), we find two of our underlying frameworks that are built on theory from psychology, specifically motivational and personality-based theories (Nacke et al., 2014; Tondello et al., 2016).

This methodological analysis gives some indications of a skewness within our dataset towards dominantly occurring motifs as well as potential topical omissions - given the small number of studies venturing outside of their core research domain. A relevant facet of consideration is that the central goal of these studies is to conflate complex and unique behaviors into comprehensive models that reflect the biggest, most dominantly emergent behavior clusters. In contrast, our goal is to arrive at a dataset with exhaustive and nuanced motivations to enrich our game design elements dataset with relevant and actionable meta-data. As such, the currently collated dataset is not yet sufficient for achieving this goal. Given the successful application of our keyword-matching algorithm on our mixed dataset of player types and playing motivations in terms of identifying clusters as well as relevant outliers, repeating this approach with another related dataset can help us towards a more complete picture. Using the same methodology as the previous study, we decided to conduct a second study focused on supplementing the original dataset through meaningfully related frameworks to arrive at a more complete dataset of playing motivations.

Building on the approaches of Nacke et al. (2011) and Tondello et al. (2016) et al. to explore the domains of psychological research on motivation and human needs for additional foundational theory and regarding their potential as suitable datasets for our follow-up study. A relationship between deprivation and motivation was observed as early as 1938 when Murray, in his “Explorations in Personality,” considered the duration of deprivation as an important factor in motivation (Veltkamp et al., 2008). Later, this model was refined through the incorporation of findings from animal labs, showing that the deprived organism has to have an understanding of behavior that leads to a reduction in the deprecation, only then leading to a higher motivation towards that particular behavior (Geen, 1995; Hull, 1930). Building on this, we reason that human needs, deprecations humans regularly have to face and satisfy, are implicitly linked to motivations for play – given that the respective game offers a behavioral path towards satisfaction of this need. In psychotherapy, models such as insight approaches, cognitive-behavioral approaches, and experiential approaches (e. g., Gestalt therapy (Perls et al., 1951)) are already in use - connecting the rational to the experiential system via fantasy (Epstein, 1993; Epstein & Brodsky, 1993). We thus argue that it is possible to satisfy certain human needs through gameplay – building on the human capabilities of using fantasy (implicit theories of reality) towards achieving an emotionally satisfying life (Epstein, 1993). As such, while we expect that a certain subset of human needs cannot be translated into playing motivations (foundational survival-based needs that require consumptive interaction with a world outside of the human mind), a matching of the playing motivations dataset with a dataset of human needs seems feasible and could potentially lead to insights in terms of yet missing playing motivations.

We explore these assumptions under the following research questions:

RQ1: *How strong is the overlap between playing motivations and human needs?*

RQ2: *Can yet unidentified playing motivations be identified through an inverted analysis of human needs?*

The following study explores these two research questions through a keyword matching between a dataset of human needs and the already existing playing motivations dataset. Our results show a strong and consistent overlap between these two datasets. Preliminary explorative analyses further indicate thematical clusters that can be translated into new playing motivations. Through the evaluation of needs and keywords that did not connect to the items and keywords of the dataset of playing motivations but are manifested in existing games, we find their prevalence in practice but not in theory. Our results show the close relationship between the two fields, their overlaps and differences, as well as the overall suitability of our methodology as a dataset connection tool.

Preparation & Analysis

We started the preparation of this study by conducting a structured literature review on human needs (Webster & Watson, 2002). We started with an open search within online catalogs (the Association for Informational Systems electronic Library (AISeL), Elsevier, and the KIT library), including the search terms: "needs," "incentive," "motivation," and "personality trait*" and "desire*." As we were predominantly focused on theories that deal with motivation, personality, and works that aim to explain human behavior, we then narrowed down the results to contain the most representative frameworks from the domains of general psychology, personality psychology, and I&O psychology. Of the 34 publications we evaluated for inclusion, the following 11 were chosen for inclusion: from general psychology, we chose Maslow's: the hierarchy of needs (Maslow, 1943), the self-determination theory (Ryan & Deci, 2000), the three needs theory (McClelland, 1967), as well as the ERG theory by Alderfer (1972) and the fundamental human needs theory (Max-Neef et al., 1990). From the field of personality psychology, we included the five-factor model (Big5 by Gosling et al., 2003), the 16 basic desires by Reiss (2001), the Murray system of needs (Murray, 1938), and the Meyers-Briggs type indicator (Myers & Myers, 2010). While the latter four publications are viewed more controversially in the field, they are well cited and fitting for our purposes for their manifold contextual cross-connections to the field of gamification. Especially the five-factor model constitutes the base of many personality-related studies and theories in the field of gamification. In cases where sequential scientific work on the same theory had been conducted as a result of the proceeding scientific research, the latest version was chosen for the inclusion, as far as it was validated and acknowledged by the corresponding scientific domains. Finally, with regards to the economic applications of gamification, we included two studies from I&O psychology, the Job Characteristics Theory (Hackman & Oldham, 1975) and the Motivation to work (Herzberg, 1974) (the full dataset is uploaded at https://gonku.de/sup-mat-phd-gho/Dataset-HN.xlsx).

We extracted each framework's item by name and description and included the publication metadata. Next, we extracted the keywords by hand and cross-validated them with a member of the team. Once more, we used a lemmatization algorithm (Explosion AI, 2019) on the keywords to raise the final matching accuracy. Depending on the length of the description, the number of extracted keywords varied from three to thirteen keywords per element. For the comparative matching, we adapted the algorithm to use two input lists (playing motivations/player types and needs/personality traits) and create the following outputs:

I) An undirected, bipartite graph containing only those nodes from both datasets that matched with a node from the other dataset. This was achieved by comparing the keywords of each node of one dataset to the keywords of each node of the other dataset (but not within the same dataset), generating edges between them based on common keywords, and discarding all unconnected nodes.

II) An undirected graph containing all nodes of both datasets. In this graph, edges are generated between all nodes that match another via one or more common keywords indiscriminate of their original dataset.

To assess the overall compatibility of the two datasets, we evaluate the *overlap* of the two graphs (combined average of the percentages of playing motivation nodes and human needs nodes subsumed in the shared graph). As in the previous study, we further analyze graphs I) and II) in terms of their *order*, *size,* and *density* and assess each node's *degree*, *closeness centrality,* and the *connection strength* between two matching nodes. We further assess the fitness of the two datasets through the qualitative measures: Looking at the bipartite (I) and the full graph's (II) nodes with the highest and lowest degrees and dyads (node-pairs) with the strongest connections. To identify human needs that could be included despite not connecting to the preexisting playing motivations dataset, we conduct an explorative clustering of those keywords from the human needs dataset that were not represented in the playing motivations dataset. Finally, we compare these clusters with the in-game mechanics of published games to assess their validity as playing motivations and create a resulting list of suggestions to be included in the overall combined dataset. An overview of the process can be seen in Figure 27.

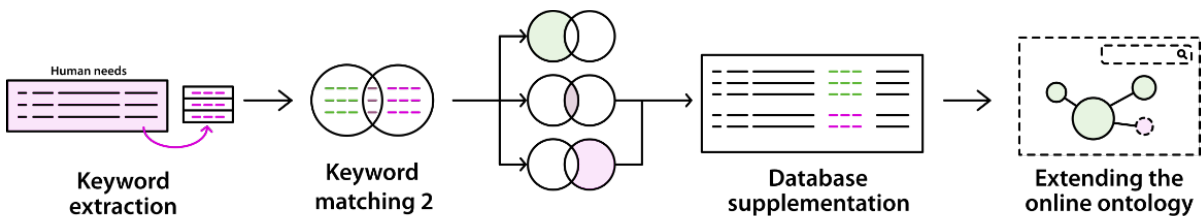


Figure 27 – Process Overview Keyword Matching 2

Results

When assessing the general fitness of the two datasets, we measure the average *overlap* of the two graphs at 89.59% (with 94.57% of all human needs nodes connecting to playing motivation nodes and 84.62% of all playing motivation nodes connecting to human needs nodes via at least one keyword). Looking at the bipartite graph (I), its *order* is 285 (connecting 87 of 92 human needs with 198 of 234 playing motivation). Its *size* is 1864 of 10764 maximum possible edges, resulting in an *edge density* of 17.32% with an average degree of 6.5 (min: 1, max: 36, median: 5). In terms of *degree*, the 20 nodes of highest degrees (36-15) all originate from the human needs dataset except for one playing motivation ("relatedness," 17 edges). Looking at the full graph (II), its *order* is 326 (with four isolated nodes, three originating from the playing motivations dataset and one from the human needs dataset). Its size is 7920 of 52975 possible edges, resulting in an edge density of 14.95% with an average degree of 24.15 (min: 0, max: 92, median: 20).

The nodes of the highest degrees of the partite graph (I) are depicted in Table 9 (the full Table can be found in the Supplementary Materials https://gonku.de/sup-mat-phd-gho/PvN-Degree.xlsx):

Table 9 – Overview of Nodes With the Highest Degrees Within the Partite Graph

Author	**Dataset Origin**	**Playing Motivation/ Player Type**	**Degree**		**Closeness Centrality**	**Avg. rel. Connection Strength**
			Edges	Average degree		
Herzberg (1959)	HN	Motivator	36	11.08%	31.17%	27.80%
Murray (1938)	HN	Achievement	33	10.15%	31.09%	23.70%
McClelland (1967)	HN	Power	33	10.15%	32.24%	21.20%
Reiss (2001)	HN	Curiosity	27	8.31%	31.09%	24.20%
Myers & Myers (1980)	HN	Extroverted	26	8.00%	30.10%	17.40%
Murray (1938)	HN	Play	26	8.00%	30.25%	24.10%
McClelland (1967)	HN	Affiliation	25	7.69%	29.52%	20.80%
Myers & Myers (1980)	HN	Social Contact	24	7.38%	28.16%	19.60%
Reiss (2001)	HN	Intuitive	24	7.38%	29.73%	19.50%
Allport & Odbert (1936)	HN	Agreeable/ Disagreeable	21	6.46%	28.89%	20.80%
Murray (1938)	HN	Recognition	21	6.46%	28.29%	20.30%
McClelland (1967)	HN	Achievement	21	6.46%	27.84%	25.50%
Reiss (2001)	HN	Vengeance	20	6.15%	29.16%	23.30%
Deci (1975)	HN	Relatedness	19	5.85%	29.30%	29.80%
Allport & Odbert (1936)	HN	Openness to experience	18	5.54%	29.88%	16.00%
Murray (1938)	HN	Cognizance	18	5.54%	28.75%	23.60%
Hackmann et al. (1975)	HN	Relatedness	17	5.23%	28.96%	22.60%
Marczewski (2013)	PM	Autonomy	17	5.23%	30.13%	21.20%
Max-Neef et al. (1987)	HN	Freedom	16	4.92%	29.30%	18.40%
Reiss (2001)	HN	Power	15	4.62%	27.96%	23.90%

Looking at the nodes of each separate graph that did not connect with a node of the other graph, the five isolated nodes from the needs graph were: "Eating," "Assertive," "Subsistence," "Abasement," and "Instrumentality." The list of the 36 isolated nodes from the playing motivations graph can be found in Table 48 in Appendix A.4.3.

Looking at the isolated nodes of the full graph (II), the three nodes from the playing motivation dataset that did not connect were: "Good Sound Effects," "Skeptics," and "Recreational Refreshment," and the one node from the human needs dataset that did not connect to any other was "Abasement." The table containing the full graph (II) can be found in the Supplementary Materials: https://gonku.de/sup-mat-phd-gho/PNvPN-Degree.xlsx).

Dyadic Analysis

The distribution of matching keywords in the bipartite graph (I) is shown in Table 10:

Table 10 – Distribution of Node Pairs of the Bipartite Graph by Number of Matching Keywords

Number of Node-Pairs	Number of Matching Keywords	Average Relative Connection strength	Percentage of total Pair- possibilities
13	3	39.63%	0.1208%
110	2	29.25%	1.0219%
809	1	15.77%	7.5158%

The distribution of matching keywords in the full graph (II) is shown in Table 11:

Table 11 – Distribution of Node Pairs of the Full Graph by Number of Matching Keywords

Number of Node-Pairs	Number of Matching Keywords	Average Relative Connection strength	Percentage of total Pair- possibilities
1	11	92.30%	0.0019%
1	6	55.00%	0.0019%
4	5	45.79%	0.0 76%
33	4	40.38%	0.0623%
122	3	36.03%	0.2303%
509	2	27.14%	0.9608%
3291	1	16.66%	6.2124%

The dyads with the strongest connections (5 or more matching keywords) in the bipartite graph (I) are listed in Table 12:

Table 12 – Overview Over Node Pairs (Dyads) With the Strongest Connections in the Bipartite Graph

Name A	Source A		Name B	Source B		Common Keywords		Avg. rel. Con. Str.*
	Src.*	Author		Src.*	Author	Strings	Amt.*	
Curiosity	HN	Reiss (2001)	Hardcore gamer	PM	Jacobs & Ip (2003)	learn, knowledge, seek	3	28.25%
Order	HN	Reiss (2001)	Guardian	PM	Keirsey & Bates (1984)	security, rule, organization	3	22.50%
Social Contact	HN	Reiss (2001)	Socializer	PM	Bartle (1996)	contact, interaction, friendship	3	30.25%
Vengeance	HN	Reiss (2001)	Negative Affect	PM	Poels et al. (2007)	frustration, revenge, anger	3	37.50%
Autonomy	HN	Deci (1975)	Free Spirit (intrinsic)	PM	Marczewski (2015)	freedom, free, autonomy	3	59.40%
Relatedness	HN	Deci (1975)	Relatedness	PM	Marczewski (2013)	relatedness, belong, connection	3	40.00%
Relatedness	HN	Deci (1975)	Socializer (intrinsic)	PM	Marczewski (2015)	social, relatedness, connection	3	45.00%
Safety	HN	Maslow (1943)	Guardian	PM	Keirsey & Bates (1984)	security, safety, protection	3	28.95%
Identity	HN	Max-Neef et al. (1987)	Competitive ness	PM	Vorderer et al. (2003)	esteem, self-esteem, value	3	35.40%

Achievement	HN	Murray (1938)	Achievement	PM	Yee (2002)	power, accomplishment, achievement	3	36.45%
Achievement	HN	Murray (1938)	Achievement oriented	PM	Whang & Chang (2009)	achievement, success, accomplishment	3	58.95%
Autonomy	HN	Murray (1938)	Autonomy	PM	Marczewski (2013)	autonomy, independence, freedom	3	52.50%
Autonomy	HN	Hackmann et al. (1975)	Autonomy	PM	Marczewski (2013)	autonomy, freedom, responsibility	3	40.00%

*Source | Amount | Average relative Connection Strength

The full tables of all dyads between both the bipartite (I) and the full graph (II) can be found in the Supplementary Materials: https://gonku.de/sup-mat-phd-gho/PvN-Dyads.xlsx | https://gonku.de/sup-mat-phd-gho/PNvPN-Dyads.xlsx.

Keyword Analysis

In total, the bipartite graph (I) contains 807 keywords, a total of which 158 are distinct (accounting for 15.51% of the overall number of distinct keywords over both datasets and 36.57% of the overall number of keywords over both datasets). The distribution of keywords occurrence in the bipartite graph (I) can be seen in Figure 28:

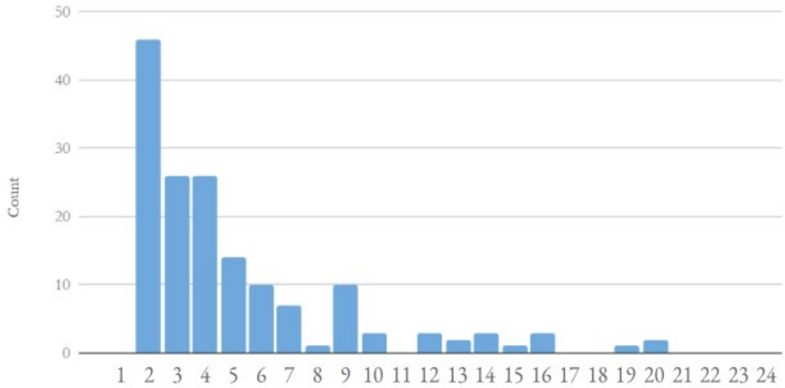


Figure 28 – Distribution of Keywords Occurring More Than Once

The distribution of keywords occurrence in the full graph (II) can be seen in Figure 29:

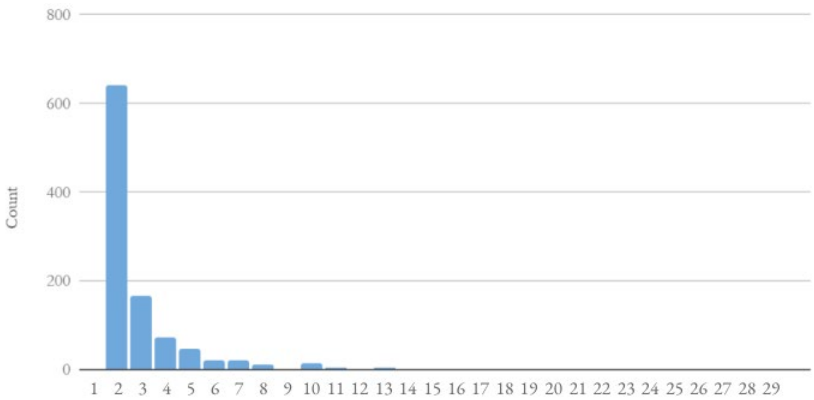


Figure 29 – Distribution of Keywords Occurring More Than Once in the Full Graph

The keywords that occurred most frequently in the bipartite graph (I) with ten or more nodes containing them are shown in Figure 30.

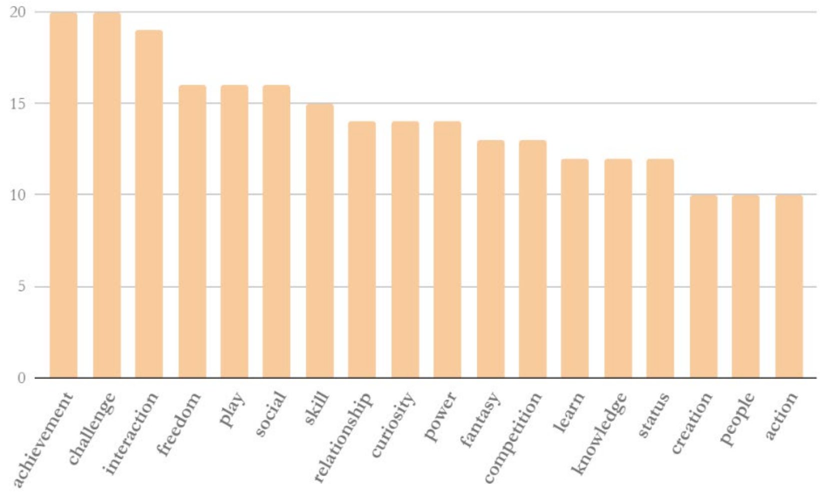


Figure 30 – Bipartite Graph (HNvPM): Most Frequently Occurring Keywords (>9)

In total, 265 keywords from the human needs dataset did not match with keywords from playing motivations (see Supplementary Materials: https://gonku.de/sup-mat-phd-gho/PvN-Keywords-onlyInNeeds.xlsx).

Explorative Analysis

Given the discrepancy between node-overlap (89.59%) and keyword-overlap (15.51% of the distinct keywords, 36.57% of the total number of keywords), we conducted another explorative semantic sorting analysis (via the same method as in the previous study) to identify keywords that could have matched with the playing motivations dataset given a matching algorithm that can match semantically close keywords instead of exact ones (like e. g. power and force. For this analysis, we built on the explorative theme map derived from sorting the most common playing motivations keywords (see Chapter 3.4.1, Figure 25). Where applicable, we sorted keywords into the preexisting theme clusters; otherwise, we expanded the map, grouping the remaining keywords under common themes (see Figure 31). In this first explorative sorting process, we placed each keyword in a unique position on the map. In a second iteration, we tagged each keyword with all suitable categories to reflect ambiguous interpretations and overlaps. Overall, 151 of the 265 unmatched human needs keywords were labeled with one category, 93 with two categories, 16 with three categories, and 4 with four categories (namely: “child,” “fairness,” “temper,” “self-dramatization”), and one, “need,” with no category. The full list can be found in the Supplementary Materials: https://gonku.de/sup-mat-phd-gho/HN-Keyword-Clusters.xlsx).

For our analysis, we differentiate between three overarching groups, i) the preexisting themes that emerged through the explorative clustering of the playing motivations keywords, ii) themes that represent categories that do not directly relate to playing motivations, and iii) new themes that have a potential of being added to the playing motivations canon (see Figure 34). Of the 266 unmatched human needs keywords, 100 matched with one or more of the preexisting themes (i) (see Figure 32). On the other hand, 73 of the keywords did not directly fit the playing motivations dataset (ii) (see Figure 33). Through another explorative keyword sort process, we identified four categories that formed around such keywords:

- *Motivation* – keywords that related to the concept of motivation itself, which would be tautological as a playing motivation ("determined," "willingness," "motivational," "motivator," "expectancy," "drive," "habit," "effortful").
- *Physical need* – keywords that directly relate to needs that must be fulfilled to subsist or directly relate to the living body (e. g. "water," "air," "subsistence," "meal," "hunger," "food," "dining," "eat," "physical," "sex," "hygiene," "cold," "shelter," "animalistic," "existence," "child").
- *Neutral parameter* – keywords that relate to parameters that do not directly relate to single themes of playing motivations but can affect the full product (game) depending on their manifestation (e. g. "primary," "secondary," "objectivity," "correlation," "potential," "satisfaction," "dissatisfaction," "portability," "high," "probability," "capacity").
- *Personality* – keywords that relate to general character and personality traits (e. g. "agreeable," "receptiveness," "even-tempered," "open-mindedness," "narrow-minded," "conservative," "realistic," "down-to-earth," "thoughtful," "conscientiousness," "shy," "insecure," "vulnerable," "neurotic," "neuroticism").

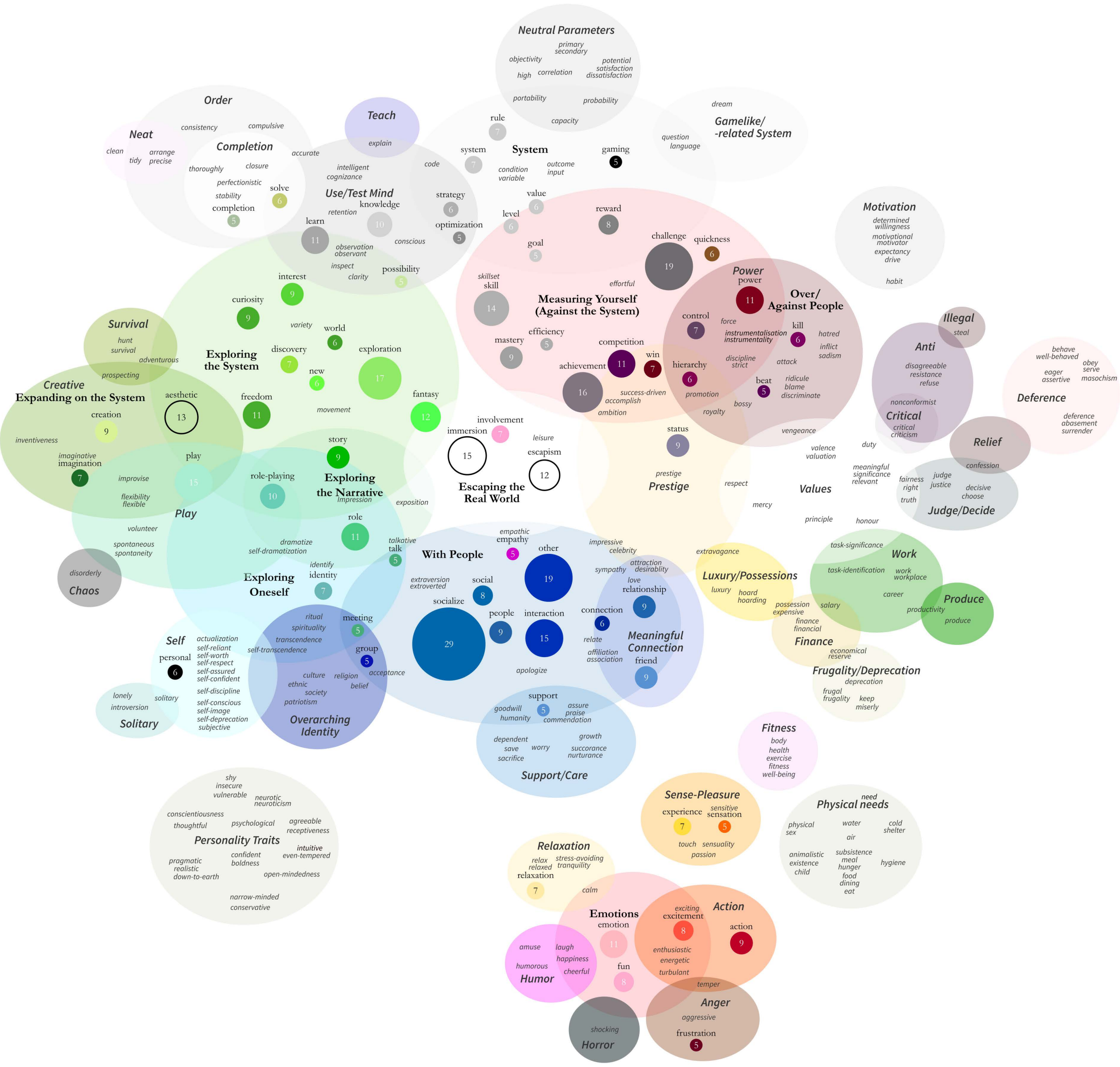


Figure 31 – Explorative Preliminary Clustering of Frequent Keywords Based on Emerging Themes

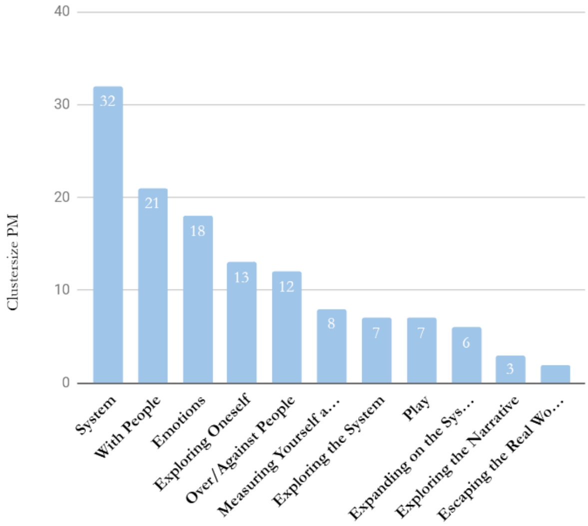


Figure 32 – Distribution of Keywords Clustered by Overarching Themes Fitting Playing Motivations

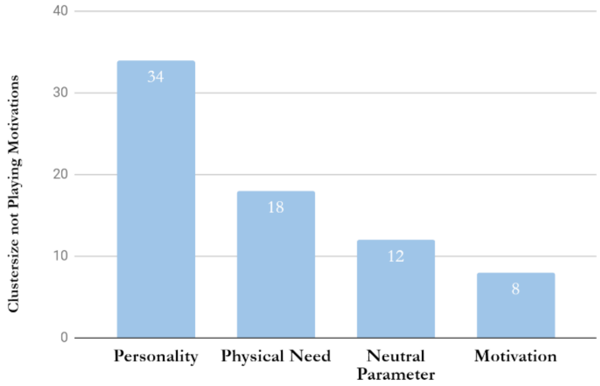


Figure 33 – Distribution of Keywords Clustered by Overarching Themes not Fitting Playing Motivations

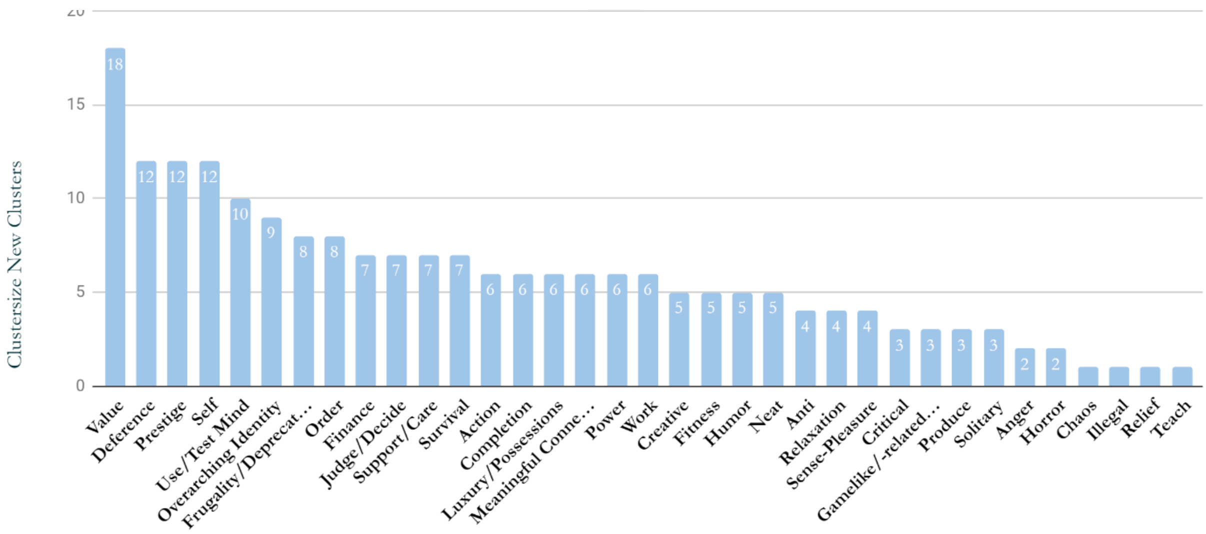


Figure 34 – Distribution Keywords Clustered by New Emergent Themes

Several keywords, however, particularly in the personality category (n=16), were matched with at least one other category. Overall, we deemed 53 keywords (19.92% of all unmatched keywords) as not fitting the overall theme of playing motivations (the list can be found in Table 49 in Appendix A.4.3). Through this preliminary explorative semantic sorting, we found 35 additional themes that relate to motifs players could seek in a game or through play.

To consolidate their eligibility, we related each of their subsumed keywords to an existing exemplary game genre, game system, or game mechanic (the full list can be found in the Supplementary Materials: https://gonku.de/sup-mat-phd-gho/HN-Keyword-to-Game.xlsx). Finally, based on our explorative clustering process, we collated a list of suggestions for new playing motivations and player types that, based on our analyses, are not yet represented in current literature (see Table 50 in Appendix A.4.4).

Discussion and Conclusions

Given the success of the first study in identifying meaningful connections between the items of our dataset, we decided to extend our experimental design to test the algorithm's potential to identify links to an adjacent dataset. For this, we collated an additional dataset of human needs from psychology literature and once more extracted relevant keywords. In this study, we were interested in gauging the relevance of human needs to playing motivations, using the algorithm to identify the degree of overlap between the two datasets. While the overlap in terms of nodes was very high, with 94.57% of the human needs nodes (87/92) connecting to at least one node of the playing motivations dataset, the overlap in terms of unique keywords was much smaller (158/423). To better assess if this relates to a weak connection between the two datasets or flaws in the algorithm, we conducted an explorative clustering, where we sorted the unmatched keywords from the human needs dataset into thematical groups of i) themes already prevalent in the playing motivations dataset, ii) themes that were unsuited for the playing motivations dataset and iii) themes that could extend the playing motivations dataset. An explorative preliminary analysis showed that most keywords could be directly or indirectly thematically subsumed, leaving only 19.92% (n=53) keywords that did not fit, highlighting that our methodology of using exact instead of semantic keyword matching accounts for a large part of this discrepancy. Also, particularly during the second round of explorative keyword sorting, we were conflicted with some of the keywords we had initially sorted into the group of unsuitable keywords. This happened particularly with the physical needs, as several game mechanics use them metaphorically as game mechanics – sometimes indirectly as a means to provide more engaging content (like offering cooking mechanics in open-world games (e. g. in The Legend of Zelda: Breath of the Wild (Nintendo EPD, 2017)), to connect to underlying animalistic desires as incentives for play (like using the candy aesthetic in games like Candy Crush Saga (King, 2012)) or to tell a story that can resonate with our fears (like the survival mechanic in Shelter (Might and Delight, 2013)). Similarly, given a broader meta-analysis, we would likely be able to connect most personality traits to suitable or often cooccurring game mechanics, as there is already a set of research connected to this topic. To arrive at a ground truth, we only sorted items into one of the playing motivations-related groups if we could find a real-life representation within an exemplary related game genre, game system, or game mechanic. By this method, we found 35 preliminary clusters that show promise for further analysis for inclusion into the field of playing motivations. Thus, in answer to our second research question, our preliminary results indicate that human needs can serve to complement playing motivations. However, given the limitations with regards to biases and ambiguities underlying this qualitative, explorative method, a future model-based quantitative study should be conducted to solidify these preliminary results.

In terms of qualitative evaluation, we find a high level of consistency between the matched nodes of the bipartite graph. When we compare the nodes connected to the node with the highest degree, we find strong thematic similarities. For example, the node "Motivation" is linked to the following nodes:

"Aggressive gamer", "Objectivist", "Conqueror", "Mastery", "Competition", "Suspense", "Challenge", "Hard fun", "Difficult to Play", "Helping & Support", "Mechanics", "Manager", "Achiever", "Teamwork", "Challenge", "Self-Seeker", "Autonomy", "Consumer", "Achiever", "Hardcore Gamer", "Challenge-based Immersion", "Challenge", "Progress & Provocation", "For Challenge", "Team-sport and Combat", "Achiever", "Being Challenging", "Achievement", "The People Factor", "Challenge", "Director", "Advancement", "Challenging", "Conqueror", "Achiever (intrinsic)", "Achievement-oriented."

And the original description of the item is:

"*As long as these factors are present in the workplace, people will be pleased, and performance will be better (individual needs will be satisfied). Missing factors don't imply unsatisfied employees. Typical examples of motivating factors are Achievement, recognition, work itself (challenging), responsibility, and advancement (promotion).*" (Herzberg, 1964).

Equally, when looking at the thirteen dyads with the strongest connections, we find high thematic congruency, with six items even matching by name. This might stem from playing motivations being in parts inspired by research on human needs; however, this further speaks for the algorithm's success in identifying meaningful connections.

On the other hand, when looking at the human needs nodes that did not connect in the bipartite graph, we find two nodes relating to physical needs ("Eating," "Subsistence") and two nodes relating to a submissive disposition ("Assertive," "Abasement"). The first two relate to basic physical needs and, as such, do not fit the playing motivations dataset. The second two, on the other hand, fit one of the new thematic categories that emerged during the explorative keyword sorting around the topic of deference and submission. The fact that these two nodes did not connect could be a reflection of a bias within game design towards a tendency to focus on the facet of empowerment through video games (being able to do and experience things a player would or could not in their normal life) and less on the opposite: the relief that comes with the release of duties and responsibilities. Together with "Deference" being one of the second largest clusters with 12 related unmatched human needs keywords, we first assumed this to be an indicator that this facet is not well enough reflected within playing motivations and player types. However, after conducting some explorative research on games building on this need in a game-related online forum (r/gaming), we find that it is not (yet) strongly reflected. While certain games like The Stanley Parable (Galactic Cafe, 2013) and Portal 1 and 2 (Valve, 2007) play with the concept through a narrative that reinforces the players' role as lab rats, no game emerged as a strong candidate building on the concept of deference.

The final node that did not connect, "Instrumentality," stems from Vroom's expectancy theory which is adjacent but not central to the topic of human needs, stemming from research into motivations to work and referring to an expected linear relationship between work put in (input) and reward (output). As such, this node points to the aspects of gaming and playing concerned with practical outcomes in contrast to pure leisure. Despite games and play generally fulfilling important functions in human lives (Pellegrini, 2009), outside of the realm of health- and serious games, they are rarely directly associated by their users with their intrinsic instrumentality (like e. g. stress-relief or wish-fulfillment) but with their intrinsic fun outside of real-life necessities (Koster, 2005).

In summary, our results showed a high degree of overlap and congruency in terms of nodes that matched between the two datasets, highlighting the thematic fitness of the two datasets. Thus, our preliminary results indicate that the keyword matching methodology has potential as a method for linking relevant items of two related but topically different datasets. Our explorative keyword analysis and clustering of the remaining keywords allowed for more detailed insights into where and why the two datasets did not suit each other and where there could be potential for each dataset to complement the other. In future studies, we plan to consolidate our preliminary findings through expert interviews and a field study.

This study contributes to research and practice in several ways. First, we showcase the potential for a new methodology that builds on meaningful links between ontological items. Second, we complement existing playing motivations with matching human needs, thus allowing for better design decisions with regard to target audiences' needs. Third, through this method, we identify a list of new types of playing motivations that show promise for further research. The resulting ontology can serve as an inspirational tool as well as a foundation for making qualitative design decisions toward specific target audiences.

3.4.3. Outlook

Based on the results of these two studies, we conclude that the keyword matching algorithm we used worked sufficiently for arriving at an initial set of automatically generated clusters for our ontology as well as a connector to produce meaningful links between the tested datasets. However, while the methodology proved sufficient for our explorative context, we want to address its limitations in future implementations, specifically its qualitative variety stemming from the number and descriptive quality of each item's keywords as well as its limitation to only matching exact keywords. Since we conducted this set of studies, methodologies as employed by word2vec (first published by Mikolov et al. in 2013) and BERT (Devlin et al., 2018) have gained in popularity and usability that can now overcome these limitations as they can match similar keywords through a semantic vector space or incorporate the aggregated meaning behind the item's descriptions, matching them by general semantic similarity. With such additions, we will be able to build bigger as well as more concise clusters. Thus, for future automated cluster processes, we will adapt our algorithm to incorporate the newest advances in natural language understanding. Also, during our research, other promising adjacent fields emerged (like personality types and game genres) that would be able to further complement our linked data cosmos towards other relevant nuances. We plan to conduct further dataset collations once the algorithm is overhauled and a platform exists in which each dataset can be linked and browsed in a user-friendly way.

Having built this preliminary foundation of connected datasets, including relevant metadata, our next step concerns the development of a tool that allows users in theory and practice to efficiently gain access to the respectively desired information.

4. Chapter 4
Tool Development for Database Navigation[14]

"This could be the new Google."

Nikita Singareddynm (Investor at RRE Ventures),
Commentary on the online field study for Kubun

4.1. Introduction

Navigating vast amounts of knowledge fast and intuitively is a competence that is increasing in importance proportional to the fast pace at which information is produced and stored today (Mohanty, 2015). The size of databases and their collective count in diverse fields and topics of interest are rising alongside the growing storage capabilities of personal computers, cloud-based services, and possibilities regarding bandwidth. Considering the academic context alone, the amount of publications constantly increases, with more than 2.5 million articles published each year (Plume & van Weijen, 2014). As stated by Scott and O'Sullivan (2005), in the future, students will not be expected to know long lists of countries, formulas, or any other knowledge by heart but to be able to find relevant and valid information on the internet or other corpora of structured or unstructured data. However, to consistently ensure successful data retrieval, the related software tools must offer easy and intuitive user interactions for navigating these large amounts of information and obtaining a sensible overview of the spectrum of relevant items (Scott & O'Sullivan, 2005).

One of the largest datasets frequently requested and accessed by any participant of the internet are websites that fulfill a specific search criterion – establishing search engines as the most used database navigation tools and thus setting the standards for best practices in information navigation and retrieval. However, while radical changes have happened in matters of visual design and user experience when it comes to the websites themselves (especially since the establishment of the term web 2.0 in 1999 (DiNucci, 1999), the most commonly used search engines (Google, Bing, Baidu (Clement, 2020)), have not significantly changed. In terms of Google's user interaction, the only relevant change that departed from the original design (consisting of a search bar for entering search queries connected by operator-based input logic and yielding

[14] This chapter comprises a working paper in the works to be submitted to the IEEE Transactions on Affective Computing. Full reference: Hoffmann, G., Martin, R., Weinhardt C. (2022). "Search by Example – Interface Design for Preference-based Visual Browsing of Semantic Datasets". Working Paper. Note: The manuscript builds on the Master Thesis of Raphael Martin: "Implementation and Evaluation of an Interface Design Affording Personalized Visual Browsing for Databases: Kubun", 2019, supervised by me, Greta Hoffmann. As a joint work in the department of computer science, the development of the demonstrator was focus of that work, while the idea, design execution and writing of the working paper were conducted by me (except for Chapter 4.4). The appendix is also based on joint work by the authors. Tables, figures, and appendices were renamed, reformatted, and newly referenced to fit the structure of the thesis. Chapter and section numbering and respective cross-references were modified. Formatting and reference style are adapted, and references updated. Opening quotation is not part of the article.

an equidistant list of text-based results) was to omit any mention of the modifiers and operators after realizing that users did not (or were not able to) use them. This interaction flaw was compensated through radical technological improvements, optimizing data analytics and machine learning strategies to approximate the users' intended queries as well as possible. With growing processing power and analytics methodology, this seems to produce satisfying results for most common purposes; however, crucial affordances for fluent, intuitive, serendipitous[15], and ubiquitous browsing (in contrast to searching) processes are still missing from database interfaces connecting the users with their contents.

In our research project, we work towards understanding and addressing these insufficiencies with the interaction design of current database navigation interfaces. Best-practice strategies from the domain of user experience design (Norman, 2013) suggest to not just solve a problem as given (e. g. by further improving the algorithms of current search engines) but to try to understand the underlying issues and to use observation techniques to inform the subsequent design of systems (Dix et al., 2004). As suggested by practices of human-centered design (Cooley, 2000), we identify problems intrinsic to the interfaces of current databases and translate the collated research and requirements into a new interaction design. In the following, we present a list of system and usability requirements extracted from literature and further informed by user tests, the design and implementation of a working demonstrator, and findings and implications from said user tests as well as a field study that we conducted. The demonstrator serves as a proof of concept for the untapped potential for efficient, user-centric information browsing processes.

The main practical contribution of our research consists of affording engineering and design practitioners with a tested and user-approved open-source product that can be deployed as a new, more intuitive way of database browsing. By applying research to design in terms of systematically understanding the content matter on an abstract layer and then retranslating it into new, more intuitive visual and interactional metaphors, we contribute to the overall effort of applied human-computer interaction. In terms of a theoretical contribution, we identify a gap in the research domain of computer/ information science in terms of real-life inspired browsing behaviors that we bridge with our research in the domain of library science. We further identify a set of user requirements extracted from literature and refined through the evaluation processes of our user evaluations.

4.2. Related Work

During the emergence of the internet as a global phenomenon, search engines emerged as the dominant tool for accessing database information (at that time, mostly website databases and library catalogs) (Cohen-Almagor, 2013). To enhance their search results, providers of search engines added functionality through advanced search queries. This direction was not popular among users, especially technology novices (Buchanan & McKay, 2011; Fields et al., 2005; Lucas & Topi, 2004). An analysis of search-engine use (an AltaVista query log containing almost 1 billion entries) found that 79.6% of all users did not use a single operator in their search queries, and for 85% of the queries, only the first result screen was viewed (Silverstein et al., 1998). Some of these usability issues continue to persist, as twenty years later, in a study on book selection behavior in physical libraries, digital search enhancers like increased number of query terms, use of

[15] Serendipity is defined as "an unexpected, accidental discovery of interesting information" (Roberts 1989) or as "finding without seeking" (Ross 1999).

search modifiers (phrase search, compulsory terms), and use of Boolean and propositional logic were still found to be used minimally and understood poorly (Buchanan & McKay, 2011).

A study that tested search behavior between two treatment groups (computer experts and novices) found that expertise in the use of technology did not solely contribute to success in the tested information retrieval tasks, but that domain knowledge was another primary contributing factor (Hölscher & Strube, 2000). The study concluded that those participants that could rely on both types of expertise were most successful in their search behavior; however, "overall web-based information seeking turned out to be rather difficult for the participants" [15, p.11]. This indicates that tools that logically base the design of their interface on the way the computational system operates in the background (Boolean operators, compulsory terms) afford good usability to more technically advanced users but are unintuitive and difficult to grasp for users with little to no background or training in computer science as these affordances don't relate to their real-life searching interactions. Best practices implemented in widespread search engines like Google and Bing now operate on a compromise-oriented solution. While still functionally intact, no indicators of the existence of search operators are provided on the main pages, thus affording experts with the functionality while protecting novices from being overwhelmed. While this approach is effective with regards to cognitive offloading, it prevents novices from discovering functions for precise research and thus successfully using the search engine to its full potential.

Researching the version history of Google (Versionmuseum, 2020), we found that a service referred to as the "browsing option" was offered with the iteration that came in the year 2000. It was removed from the main page only one year later (2001) and entirely discontinued in 2011 (Waybackmachine, 2020). This service offered a category-based directory as classifications of pages by participants. While the site is no longer available, a link to the original page as well as a description of its workings can still be found on the website "googleguide.com" (Google, 2007). On the guide page, the authors, by referring to the work of Sherman et al. (2001), state that browsing is more useful if one is familiar with a participant (especially as it affords serendipitous findings) while searching (with the aid of specialized tools) is more likely to provide satisfactory results if you are unfamiliar with the content matter. They conclude that for using the web, either process can provide good results, depending on context and prior knowledge. As the directory service was abandoned, it seems that this specific interaction design for browsing did not resonate with the audience.

To broaden our understanding of user expectations and behaviors relating to real-life browsing processes, we continued our literature review in the domain of library science as "using the web to find information has much in common with using the library" (Sherman et al. 2001, p.18). We found that since the rise of computational infrastructure, a field of library science has been formed that is concerned with the design of digital libraries, specifically focusing on the development of systems and tools for the navigation of vast catalogs. While progress has been made in digitizing analog catalog systems to improve their usability within the realm of their related real-life libraries, we found many publications that criticized the usability of digital libraries and their catalogs (Buchanan & McKay, 2011; Hinze et al., 2012; Schamber & Marchionini, 1996). By analyzing the browsing behavior of people within real libraries, Stelmaszewska and Blandford (2004) aggregated requirements for successful digital browsing processes. Their research reveals that browsing in real libraries happens in an entirely different manner from the search strategies employed in using the internet. While the use of search tools and strategies is an integral part of the information-seeking process in the library, they are only used for very specific ancillary purposes in the general browsing process – like gaining meta-information or precise location details on specific entries. In contrast, the real-life browsing process further consists of gaining a glancing overview over whole sections of entities, closing in on an area, and finally making a choice while always being able to peek into random items of interest. They further make the point that such discovery-based experiences and serendipitous findings are not afforded by scrolling textual lists of ranked results offered by a digital search (Stelmaszewska & Blandford, 2004). In their work, Hinze et al. (2012) highlight another problem of digital libraries: metadata that is afforded effortlessly on a visual level in a library or storefront (such as size, weight, font, color, or amount of dust as an indicator of recent use) is lacking in results produced by common search engines, as they only present their results in a very restricted combination of a title and a few lines of text. This implies that choices made in a real-life library are not made based on single criteria but in terms of sets of underlying parameters. Thus, another restriction of digital database navigation tools leads back to the lack of control in clarifying sets of preferences on a criteria level (filters do not afford to highlight preferences but only limit the final list of results – this is further discussed in Chapter 4.3.1). Even when using more advanced operators, users are forced to express their criteria as binary statements codifying certain properties instead of being able to communicate their preferences in combination with an associated set of perceived importance, allowing for fuzzy and compromise-based search.

Further indicators we found as to why browsing affordances potentially do not yet deliver overall satisfactory interactions or results are listed as follows. First, one of the most frequently mentioned issues users reportedly faced in the digital space was being overwhelmed with the quantity of information (e. g. in a study on cognitive strategies in web searching, users reported having trouble with remembering the content of each window when more than three windows were opened) (Navarro-Prieto et al., 1999; Stelmaszewska & Blandford, 2004). Second, the internal models that users have of their information searching strategies (the way users approach a search task in their mind) differ from the external representations that they are offered by search machines and digital catalogs (Buchanan & McKay, 2011; Navarro-Prieto et al., 1999). As is shown in research on cognitive fit theory (Vessey & Galletta, 1991), a negative relationship emerges between performance and task if the task representation does not match the conceptual task formulation. Relating to this, we identified a general lack of affordances to support "mental mapping" processes (Hinze et al., 2012). Even though digital browsing represents the traversal of the digital space, barely any navigational affordances are given to the users to help with orienting themselves within the digital area they are currently perambulating (De Ruiter, 2002; Tan & Wei, 2006). Another problem relates to the result representation. Studies show that learning and understanding are better derived from the combination of words and pictures than from words alone (Kolers & Brison, 1984; Mayer, 2009). This puts the effectiveness of presenting text-based results into question. Finally, we see a problem with search engines relying on hidden metatags. By obscuring the algorithms and their influencing parameters that determine the results, search engines bereave their users of self-determination and any transparency on what results are shown, when, and why.

4.3. Design Process

Through our design and research efforts, we aim to arrive at a digital affordance that offers enhanced browsing possibilities. Reporting feedback on the insights gained during the design and testing process informs research as well as practice (Van Aken, 2005). Efforts to close the gap between the domains of design, psychology, and technology have long been conducted by researchers as well as practitioners termed "exploratory action research" and "case study approaches" (Candy et al., 1996; Clibbon et al., 1995), the "task-artifact cycle" (Carroll, 1991) or "design science research" (Kuechler & Vaishnavi, 2008). Therein theoretical advancement is achieved by conducting a transformation from cause-effect predictions to artifact design recommendations (Voigt et al., 2012). On the other hand, as each design draws upon theory, the quality of the resulting design validates this theory when the application of design principles has been successful. Thus, according to Hanseth and Lyytinen (2004), the value of design theories for IS research and practice is twofold. First, through the effective transfer of practical knowledge to new situations (external validity) and second, through their normativity (not only aiming to explain the world but seeking to change it). As Hassenzahl (2010) writes: "each interactive product is a proposition, a new opening in the design space. It will inevitably alter expectations and ultimately the requirements users formulate concerning future technologies." (Hassenzahl, 2010, p.62).

4.3.1. Requirements Engineering

In our literature review, we identified a gap between the browsing behavior afforded in real life and the browsing affordances currently offered by the interfaces connecting users with the information catalogs of the world wide web. As our primary research goal is to integrate usability requirements for successful browsing strategies into a working demonstrator, we extracted requirements from literature (core requirements) as well as through interviews with the target audience(s) (user requirements) and translated them into boundary conditions for a technical framework (system requirements) as suggested by (Maiden, 2008). We separated the requirements into user- and system requirements and further structured them according to the predefined typesets of requirement categories (S. Robertson & Robertson, 2013; Roman, 1985). Finally, we applied three priority categories: primary, secondary, and "not used." The latter category was set up due to limitations of time and resources and served as an outlook to further iterations.

With our focus on user interactions and the corresponding interface design, we chose usability (UR), look-and-feel (LaFR), and performance (PR) as our primary requirements and reliability (RelR), maintainability (MR), training (TR), availability (AR) as secondary requirements. The other requirement categories were excluded as they did not relate to the scope and goals of the project at the current stage. We started by extracting a list of requirements from the literature (see Supplementary Materials: https://gonku.de/sup-mat-phd-gho/Kubun-Requirements.pdf). The list was later consolidated and iteratively adapted during the user tests that we conducted during the evaluation part of this study.

4.3.2. Market Analysis

As a recurring criticism of database interfaces was related to the result presentations and their lack of affordances for serendipitous findings, we conducted a market analysis of current database visualizations. During this research, we found several innovative interfaces that use graphs for visualization by displaying node-based entities connected by lines representing their relationships. While these graph-based interfaces offer intuitive ways of exploring entities and their relationships with each other, they commonly suffer in terms of high cognitive load as they consistently show huge quantities of nodes and their connections even on the default zoom-level and pan, displaying vast datasets on a single page (Advanced Research Consortium at Texas A&M U., 2015; Arcade Analytics LTD, 2019; Kumu, 2019; Mauri et al., 2017; Rhumbl, 2019).

While this might seemingly afford users an overview of the structure of complex datasets, this kind of presentation is difficult to understand and interpret (Knight & Munro, 1999).

Next, we looked at the interfaces of digital libraries. Most interfaces offered a similar set of features: a search affordance for titles, authors, publishers, and ISBN/DOI and filter functions on characteristics like subject and publishing date to manually curtail the result list (Karlsruhe Institute of Technology, 2009; Universitätsbibliothek Heidelberg, 2020). Some websites offered additional innovative features, like the World Digital Library (2019), where the interface was designed with a focus on an image-based interaction and suggestions based on similar items, and the German Digital Library (Deutsche Digitale Bibliothek, 2019), which additionally offers a compare-mode that allows for entities to be viewed side-by-side, the Virtual Library of the University of Würzburg (2019) that offers a virtual bookshelf aimed to recreate the process of real-life book browsing and the Europeana Collections (Europeana Foundation, 2019) that feature innovative, abstract criteria for sorting, like using the color-palette of images. However, apart from these divergent features all named applications otherwise adhered to the established design of searching for terms and subsequently filtering the results. Finally, while some of these tools featured recommender features based on similar results, users were not able to express their personal preferences to specify which criteria influencing the underlying search algorithm signify similarity to them. In conclusion, our market analysis showed that available tools incorporate certain features that would suit the requirements we aggregated, however, no single application we found satisfyingly affords all requirements we listed for intuitive browsing of datasets.

4.3.3. Design Development

Based on our requirements and the inspiration gained through our market analysis, we designed the core interaction features of the visual interface. The two needs we identified that are not addressed by current design standards are I) the need for a result layout that affords serendipitous findings and II) the need to communicate which personal criteria should determine the resulting output of similar entities, thus allowing for intuitive and effortless comparison of relevant results.

Name

According to Norman (2013), groupings and proximity are important principles of Gestalt psychology. Gestalt theory, a theory developed in the early 20th century, is based on the principle that a thing is comprehended in relation to its surroundings (Wertheimer, 1938). We chose the Japanese word for "classification" (**区分 [Kubun]**) as its components highlight a significant context of ontologies. In Japanese, the word is made up of two Kanji: the first, **区 [Ku]**, translates as "district," and the second, **分 [Bun]**), translates as "part." This constellation, where the single part is given context through its surrounding district, resonated with our understanding of how ontologies should be presented for easy human interpretation.

Result Visualization

Results of a search are currently almost always presented in textual list form. However, thereby relevant metadata is omitted that in real life would influence decision-making (like visual and haptic cues). Apart from problems relating to the textual representation of results, the equidistant representation of list items further omits information on how those list items relate to each other. Based on the guidelines suggested by Cognitive Fit Theory (Vessey & Galletta, 1991), to offer representations that closely resemble the mental concepts of related tasks, we designed a result representation that features a top-down view on a star map-inspired plane of nodes. By using a map-inspired feature, we want to reinforce the metaphor of guiding users traversing the digital space, inspiring exploration of the visually related content through an easy overview of related points of interest. As the visualization feature offers spatial information (by highlighting the distances objects have toward each other), different kinds of subconsciously intelligible metadata can be embedded using cardinal points, areas, and clusters (see Figure 35). Further decision parameters inaccessible by typical listed results (color, arrangement, sympathetic or interesting looking content) are thus afforded.

Criteria-based preferences

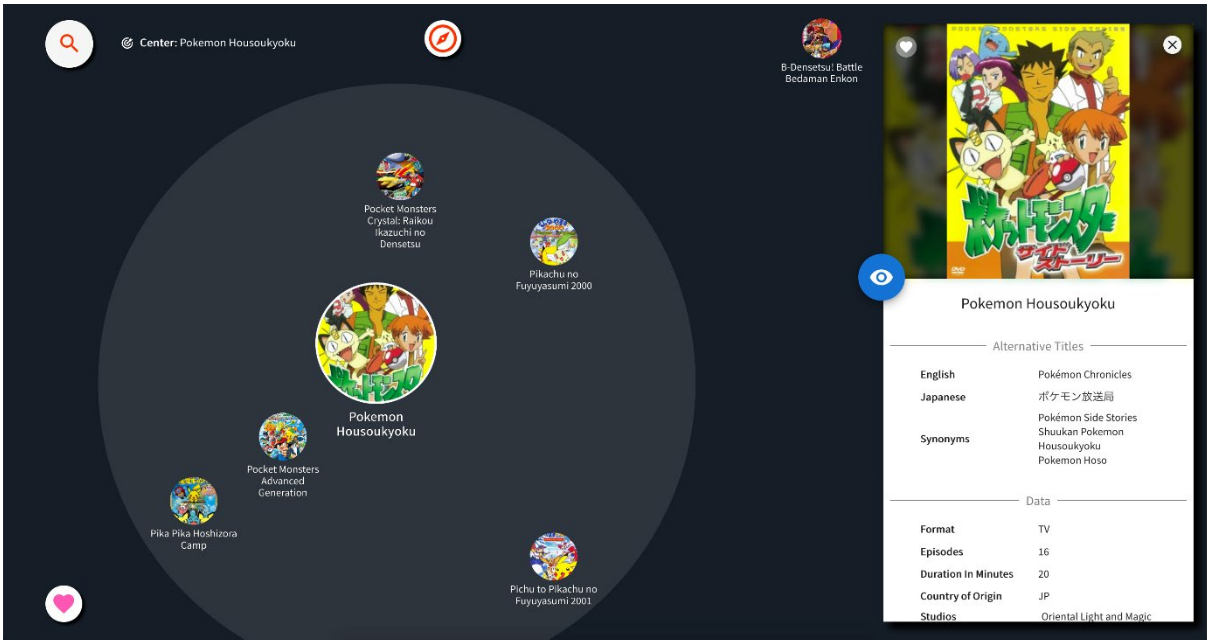


Figure 35 – Intermediate Screenshot: Map-based Visualization of Results

Currently, many database interfaces implement three interaction features to help users with finding the desired entry: recommenders, filters, and sorting the result lists by certain criteria. To understand why these features do not always lead to satisfactory results, we analyzed them in more detail. While recommenders are aimed to help with the browsing process by suggesting desirable results, oftentimes, users are dissatisfied with the quality of their suggestions (Nguyen et al., 2018). Even though efforts are made to improve them through efforts of data analytics, machine learning, and user analytics, it remains difficult to afford temporal preferences that play into the user's tastes. Apart from these factors, as recommenders are designed to make choices for the user, they inherently are not designed to help users with gaining control over the dataset. While filters do afford that control, they are used for limiting the list of results based on the set criteria. This makes them difficult to control as their effectiveness depends on the size of the database. If the database is large, more filters have to be used before the set of results is browsable; on the other hand, it easily happens that no results are found if the criteria are set too harshly. Finally, the affordance to sort the result list by certain criteria is helpful, but its utility is limited as it uses only one criterion as the determining factor for finding relevant results. Given the level of familiarity with the content, users might have preexisting preferences due to their specific tastes or a specific context or circumstance. As such, they should be afforded to use a combination of parameters to prioritize the results offered by the search.

With these factors in mind, we designed an interactive feature that combines the best of both control features by using the inherent criteria of every element of the dataset to allow users to highlight facets that are relevant to them and arrange the results accordingly. This way, a set of results will always be shown, even if a perfect match is not available. This process is called weighting, as these parameters are added to the similarity calculation with a higher weight. However, in contrast to the common understanding of the term, where weights are determined by the distributors of information (see e. g. Robertson and Jones, 1976 and Jones, 1979, we determine weights dynamically and based on user input. The process is based on the way choices are made in real life: opportunistically, parameter-based, and in contrast to similar items ("I prefer this to that," "I like this one a lot better than the other").

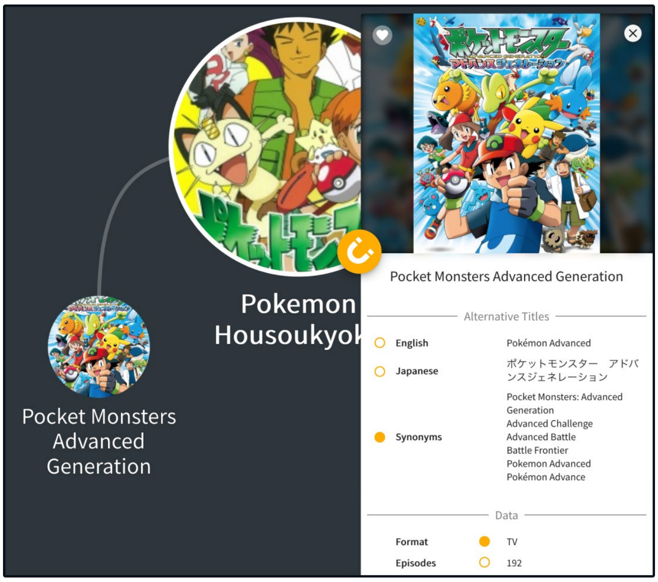


Figure 36 – Mockup: Weighting Mode as Represented in the Sidepanel

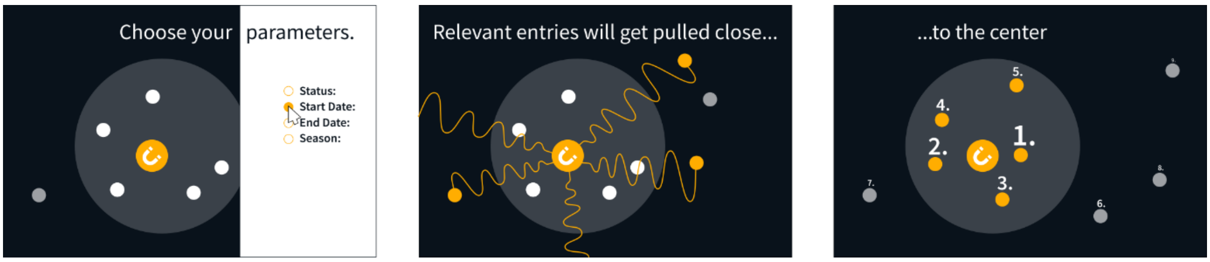


Figure 37 – Mockup: Weighting Mode Tutorial

We based the design of the interface on the metaphor of a magnet. The given parameters of the selected node are used as relative starting points. Now, a magnetic force can be applied to specific parameters that the users choose as the features most relevant to them. Based on their input, other entries in the database are pulled closer to the center (see Figures 36, 37). This process thus differs from the surgical application of a filter as it does not exclude any result but only rearranges the set of results. This is important for the affordance of browsing flow, as other factors, formally not seen as important, might organically emerge during the process. Furthermore, this interaction affords accidental discovery (serendipity), a feature considered beneficial to the browsing process. The design is tailored to reflect a human decision-making process that happens impulsively, opportunistically, and instinctively. The parameters are presented in a structured manner in a panel on the right side of the screen, where users can use the weighting feature by checking a radio button. Thus, operations on the database happen intuitively at the same location where the parameter is first discovered. As we designed the process to start with a search for an already known item of reference within the database, users are immediately placed into a realm of similar results and thus a predetermined set of likely preferences that then only need to be adjusted. As soon as another item emerges as an attractive starting point, it can be set as the new center around which the results are presented.

The described features constitute the core of our interface design for new browsing interactions. Several additional features have been included in the design of the tool (like a search feature, as well as a "chips" bar, highlighting the choices that are currently influencing the resulting layout in the node-view and a favorite function to "collect" relevant results during the browsing process). However, since these features are already well tested and were not deviated by us from their standard implementation, we limit this section to describing the design process for the innovative design features. Designs and Screenshots of the overall process are documented in Appendix B.1.1.

4.4. Development of the Demonstrator

After establishing our requirements, we began working on implementing a demonstrator for the project. From a technical perspective, the demonstrator consists of two parts: A front-end providing the user interface and a back-end that is tasked with data management and calculation of the actual results. We conducted an extensive process of selecting and researching existing technology and implementation strategies for both components: For the front-end, we decided on employing VueJS, a JavaScript component framework, in addition to D3.js, a popular data visualization library. For rendering, we decided on using SVG because of its flexibility and ease of integration with the browser's event loop.

For the backend, we started our selection process by evaluating several popular graph databases since the data we want to store and the process is inherently graph-like (as it consists of a set of entities that, in turn, are associated with a set of properties as well as relations between them). However, we did not find a database system explicitly optimized for implementing a k-nearest-neighbor (k-NN) search for high-dimensional, semantic data. Because of the high dimensionality of our data, algorithms typically used to reduce the runtime of k-NN queries cannot be employed easily. This problem is commonly referred to as the Curse of Dimensionality (Zimek et al., 2012). Also, algorithms that partition the search space proved to be problematic, as our data attributes are not necessarily ordinal in nature. We further decided against heuristic approaches, as we wanted to produce accurate, deterministic results. We concluded that we would be unable to attain runtimes that we considered adequate for a real-time system with existing graph databases. In the end, we implemented our own database system, specifically optimizing it for multi-core k-NN queries in high-dimensional scenarios.

For the field test, we selected a publicly available dataset of Anime ("A style of Japanese film and television animation, typically aimed at adults as well as children" (Oxford Dictionary, 2019)) that features about 14.000 entities, a set of relevant properties, like genres and tags, release date, number of episodes and seasons, as well as cover images for each entity. We chose this dataset as it holds several benefits pertaining to our testing needs: the dataset is of a manageable size, it has an active userbase with deep domain knowledge, and as we have expertise in this domain ourselves, we can evaluate and adjust the underlying similarity algorithm.

4.5. Evaluation

4.5.1. User Tests

Because of their high relevance to practice, as well as their ability to capture the social contexts of the interviewees, we conducted a series of user tests based on the guidelines established by Kruse & Lenger (2014). We used semi-structured guideline-based interviews, as they afford interviewers to flexibly steer the conversation along with the guideline while allowing the interviewees to speak freely. The format was chosen as a means to evaluate our core ideas as well as refine our gathered requirements and iteratively improve the demonstrator (Wessel, 2010).

We partitioned our interview guideline into three parts: (I) introductory questions on the demographics of the users, (II) a task to be performed interactively in the tool, and (III) a two-folded section inquiring about the usability of the tool (including a SUS survey for a comparable, quantifiable measure (Bangor et al., 2008)) and on the look-and-feel of the interface. These dimensions were then further divided into feature-based sections and translated into single interview questions. The main goal of the questions within the section category "usability" was to determine if the features we had designed (the map-based graph visualization and the weighting interaction) successfully afforded a better browsing interaction. We further inquired about the general acceptance and perceived utility of the general idea for the tool as well as its visual qualities in the section category "look-and-feel." Certain questions and question groups within the requirement-based categories "usability" and "look-and-feel" changed over four iterations of interviews. Each iteration featured a distinct set of 3 or 4 users. For the iterative process, we implemented a quick feedback and iteration cycle: the cycles were conducted on a 1-week to 1-week basis (1 week of implementation, one week of interviews). The interviews were protocolled via audio-recorder and in written form as well as via recording functions of the video-recording conference tool "Zoom" (however, the recordings were only used for completing the written protocols and were deleted afterward to ensure data protection).

4.5.2. Field Study

While user interviews are valuable in terms of immediate insight and feedback on usability issues as well as in gathering further design ideas, we also wanted to gain objective insights with an unbiased, neutral audience. We launched the field study after implementing the final set of feature requests from the fourth iteration of user studies. The link was publicly posted to four discussion boards related to the content (anime) of the underlying database (three German: anime-community.de (Völkl, 2000), animexx.de (Animexx e.V., 2000), animetreff.de (Anime Kultur Verein 2014)and one international: animesuki.com (vBulletin Solutions Inc., 2019)), as well as to seven subforums (subreddits) of the discussion board Reddit: r/anime, r/visualization, r/dataisbeautiful, r/usability, r/UI_Design, r/web_design, and r/webdev. In addressing the anime community, we wanted to learn the tool's capabilities of serving the potential browsing needs of an entertainment-oriented community, while the other subreddits were addressed in hopes of receiving feedback from peers within the larger HCI community. We designed the posts with an embedded link within a short, explanatory text, featuring a screenshot of the tool as an eye-catcher (see Figures 92 and 93 in Appendix B.1.2).

4.6. Results

4.6.1. User Tests

We interviewed 13 users in total. In terms of panel, we first identified potential stakeholders relevant to the core domains of the project: computer science (2 (m, w)), design (3 (w, w, m)), and humanities/library science (3 (w, w, m)). We further invited users from promising domains related to the research goals of the project: two consumer experts (w, m) and three candidates with connections to further potential use-case domains (w, w, m). The mean age of the interviewees was 33.23 years, with a median of 31, a minimum of 23, and a maximum of 65 years. The gender was distributed with slightly more female participants (61.54%, 8) than males (38.46%, 5). The interviews lasted between 45 and 60 min (except for two interviews that went up to 90 minutes). The mean of the measured SUS value was 80 (translating to a score in between the "good" and "excellent" range, with a median of 82.5, a minimum of 60 (between "ok" and "good") and a maximum of 92.5 (between "excellent" and "best imaginable") (Bangor et al., 2008). The full SuS Evaluation is presented in Figure 38).

Figure 38 – Outcome SuS Evaluation over the Four Cycles

Concerning the two innovative features (graph view and weighting), all participants perceived them as beneficial to the searching process and understood their usage intuitively. However, we protocolled several change requests and suggestions for additional features to fine-tune the interaction and added them to our list of requirements.

Overall, the tool was perceived as useful as well as innovative, and all participants stated that they see themselves using the tool for work as well as private purposes as soon as further databases related to their interests are added. (Full overview of the feedback given in the interviews and commentaries on social media is listed in the Supplementary Materials: https://gonku.de/sup-mat-phd-gho/Kubun-Social-Media-Feedback.pdf)

4.6.2. Field Study

Qualitative Results

On Reddit, our posts received 40 upvotes in total (88% of votes were positive), as well as 25 comments. We noticed that on discussion boards that were centered around anime, our demonstrator received more critical feedback, while on design- and web-development-focused ones, we received mostly commendation. Users posted feedback regarding interaction and design, as well as bug reports and suggestions for future development. Those bugs that were reported within the comment sections were fixed right away. Some noteworthy comments include:

"*Looking forward to when you'll add parameter-based clustering, all the more if the parameters can be user defined*" (Ot3n 2019)

"*Awesome!!! I love it. We should collaborate.*" (benthebigblackguy 2019)

"*Awesome idea and I appreciate your time in making this thing and I think you are so close to making an intuitive searching experience! I think there is a great need for this.*" (uxgordonlewis 2019)

"*Just finished playing around with it. and man I'm excited about its applications for other datasets, and I'm glad you guys/gals plan to make it open.*" (bitwhys 2019)

Quantitative Results

We evaluated the data aggregated over the first one and a half months of operation. For this, we used two sources of data: Firstly, our backend (labeled as BE) logs all requests it receives and identifies unique users by a tuple consisting of an IP address, as well as a user-agent. Secondly, we employ Google Analytics (labeled as GA) to log interactions with the user interface. As 58% of our users used ad-blocking, thus preventing us from collecting data about their interaction with the front-end, we use additional backend data to increase the accuracy of our analyses. We set a limit of 7 minutes of idle time to determine that a session had ended, as we had several participants that showed usage over several days. If a person resumed their interaction after the set idle time, we regarded it as a new session. The demonstrator was not developed for mobile use, and we turned off functionality if accessed via a mobile browser. Participants who clicked on the link with their mobile devices were excluded from the data analysis (about 50% of users). Additionally, we excluded our own interactions with the application from the dataset.

We analyzed the time users spent with the tool as well as the number of recurring visits. In total, we logged data of 375 (BE) unique users that interacted with the application at least once (users that only requested the website but did not interact by searching, for example, were excluded from our evaluations to avoid counting automated requests like those from bots and crawlers). On average, the application was used 1.2 times, and 51 users returned at least once (BE). The highest number of total sessions of a unique user was 5 (BE). Visitors used the application for an average of 5:41 minutes, new users stayed for 4:31 minutes, and returning visitors for 7:28 minutes (GA). The longest session lasted for 28:04 minutes.

On average, users viewed a set of results for 43 seconds before selecting a new center or changing the weights of the current one. The longest browsing of a result-set took 6:36 minutes (BE). 6.5% of users started their session on an anime directly, presumably because they received a shared link from an acquaintance (GA). Finally, 1% of users revisited a node that they had selected as the center at a later date (BE).

In total, users selected 843 distinct anime as centers. Three hundred sixty-six of those were centered as a result of a search query, while 477 were selected from the previous result-set (meaning 56% of centers were set while browsing). Per session, users set 4.8 different anime as center on average (BE). Regarding the affordance of weighting, on average, users selected 1.2 weights per request. Considering weights are reset once a user selects a new center, the average number of weights is 2.6 (all BE). Figures 39 and 40 show the distribution of the properties users selected most often during the weighting process and the number of weights set per interaction session.

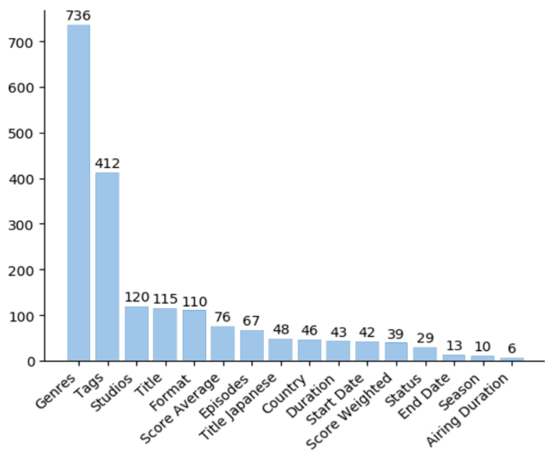


Figure 39 – Distribution of the Properties Selected by the Users

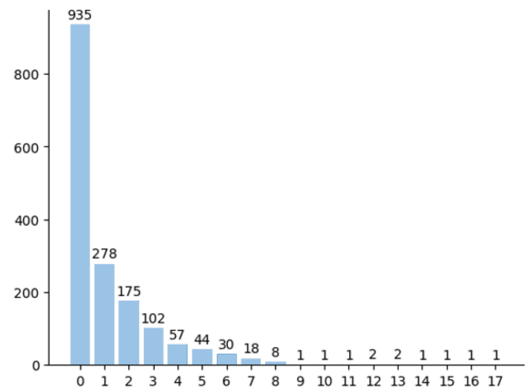


Figure 40 – Distribution of the Number of Weights set by Users per Session

The maximum number of weights that were selected in a single request is 17. Users were most interested in weighting genres and tags, but other properties were selected with non-negligible frequency as well. Furthermore, looking at the distribution of weights within the classes (such as “genres” and “tags”), users selected a wide spectrum of values and did not primarily select the same genres or tags, for example.

From a system perspective, most users used a screen resolution of 1920 x 1080 pixels, but we also observed a large variety of other resolutions. We did not find indications that the average session duration was affected by differences in screen resolution (GA). In the timespan mentioned above, the system was unavailable for 55 minutes in total (<1% of downtime, BE).

4.7. Discussion and Conclusions

4.7.1. User Tests

During the user tests, the overall feedback we received was very positive. Especially the general utility of the tool was praised regarding the new approach the tool offers toward serendipitous, personalized findings. While the mapping was very successful in terms of usability (all participants managed to navigate the visual browser intuitively in terms of moving away from the center and scrolling in and out of the graph), several requests were made in terms of further utilizing the newly afforded space, especially with regards to direction. Thus, now that mapping is afforded, we have to establish further features for using the established space and further develop the navigational features of the tool. The second main design component, the weighting feature, was also received very well and offered good usability, albeit in most cases only after the second usage. Once understood, the users utilized the feature without any problem or hesitation; however, as the weighting feature affords a new interaction design, we have to design better discovery features to support a quicker understanding of its potential. Several of the identified usability issues were fixed between iterations, and the remaining points, as well as the feature requests and visionary ideas offered by the users, were integrated into the to-do list of the project and will be addressed as the project evolves.

4.7.2. Field Study

Qualitative Results

The feedback we received from users of the field study further confirmed the usability as well as the usefulness of the tool. Especially users of discussion boards that focus on design and web development expressed their satisfaction with the general idea as well as the tool, posted suggestions for future functionality, and even asked to collaborate. Anime fans were more critical of the tool, especially regarding the results that were spawned by the similarity algorithm when no weights were set and regarding certain entries of the underlying database. This highlighted two relevant factors – first that while the tool affords users to specify their preferences, each database should be pre-weighted by experts in the domain to increase the usability even further. Secondly, while our tool affords a new way of interfacing and navigating databases, the actual entries will need community management systems to ensure data integrity and timeliness.

Quantitative Results

We were positively surprised with the number of users that we managed to attract by posting about the field test and the many messages with positive feedback about our efforts. Seeing that many users decided to return to the application after their first impression, we feel confident in assuming that this serves as an indication that the tool was of some use to them. This assumption is further confirmed by the duration time of each session, with users even prolonging their sessions in subsequent visits.

When investigating browsing behavior, more than half of the centered items were selected from a previous set of results, implying that the initial results were used to explore other items that users presumably didn't know about before. Since, throughout a typical session, participants set an average of almost five items as centers, we conclude that the initial searching for a known item and following browsing synergized well with each other. We further conclude the tool to be useful in that some users shared links, indicating that the tool was used to recommend items to each other.

When it comes to the weighting feature, on average, users tended to set a low number of weights. However, since weights are reset once a user selects a new center, the data is biased towards zero weights. Thus, from the quantitative analysis, we conclude that setting between two and three weights seems to suffice for steering the results towards results reflecting the users' personal preferences. In this dataset, users predominantly selected two properties for weighting: genre and tags. However, by examining the specific genres and tags that were chosen, we found a broad distribution of preferences that would not be easily generalizable using a non-personalized presetting.

4.8. Limitations and Outlook

One limitation of this study is that we tested the tool with a niche dataset, Anime, targeting a specific online community. While this was done to better assess the underlying similarity algorithm, based on our domain knowledge on the matter, the anime community in the field study reacted almost only in terms of specifics regarding the dataset instead of the tool. On the other hand, the UX-design-related communities reacted very positively. Within the expert interviews as well as the field study, we received suggestions with regard to future features that should be evaluated and tested for inclusion: In the current version of the demonstrator, it is only possible to select existing nodes as a center. Several users requested or implied they might want to model their personal preferences regarding some or all properties of the schema from scratch without having to start their interaction by selecting a starting point for their browsing process – particularly in domains where they have preexisting expertise. Also, it emerged that while weighting offers the affordance of expressing a personal preference, a filter function would have added further usability to the tool. In future studies, combined interactions between these two data arrangement features should be tested. Finally, with regards to known items, a feature request emerged from hiding specific nodes users are not interested in. This approach could be further undergirded by implementing a labeling feature for nodes (e. g. a label like "Movies I have already watched" for a movie-related dataset). In contrast to the significant progress made by search algorithms, the interaction design of database interfaces has not seen large changes despite a growing need and dissatisfaction with existing services. Specifically, the affordance of browsing ("to look over or through an aggregate of things casually, especially in search of something of interest" (Merriam-Webster, 2020)), instead of being improved on, has mostly been substituted by searching affordances, despite constituting of a different process with distinct benefits. With Kubun, we contribute to research and practice by demonstrating that it is possible to create desirable and intuitive interfaces that afford meaningful browsing processes even in large datasets. By affording users input control over the similarity algorithm regarding specific facets of their preferences, our tool offers autonomy and transparency over the sample space. Furthermore, through its map-inspired interface offering visual and spatial meta-information on the result entries, our tool affords serendipitous finds.

The user tests, as well as the field studies, have demonstrated the potential as well as the usefulness of our tool in its current state while equally inspiring several different directions for design and development that the tool can evolve towards in the future.

4.9. Dataset Integration

After designing, implementing, and successfully testing the database navigation tool, we prepared the game design element and playing motivations datasets for integration in Kubun. For this, we first conducted data cleaning on both databases, ensuring uniform notations and correct spellings. Secondly, we extended the database schemas with the metadata we had evaluated during our research (see Chapter 3).

4.9.1. New Schemas

Apart from the fundamental data (name, description, and picture), both datasets feature the "origin" module (metadata on the literature the game design element or playing motivation was extracted from). Building on the algorithm that we developed and tested in the keyword matching studies (see Chapters 3.4.1

& 4.3.2), we extended the schemas for both datasets with a module on linked nodes, thus connecting them with each other and offering the link information as additional metadata. We further added links from the aggregated human needs dataset as well as from a dataset of games that we extracted from IGDB (Twitch, 2006).

We added another module for categories as these were derived via the card sort studies (see Chapters 3.3.2 & 3.3.3) to the game design element schema. We designed an additional "requirements" module to the game design elements schema, as the implementation of certain game design elements can require special skills which might be relevant to know for users that are under restrictions in terms of resources; however, the content for this module will have to be collated in future efforts, potentially through crowd-sourcing based studies. Figures 41, 42, and 43 show screenshots of the datasets integrated into Kubun, and Figures 44 and 45 depict mockup designs for derived full schemas with example data. The digital database can be accessed via this link:

https://next.kubun.io/

4.9.2. Limitations & Outlook

To be successful in its aim to support users in their design process, the ontology should provide the best game design element or playing motivation for a given purpose. For this, the database has to offer ways in which constraints, as well as goals of the final product, can be expressed within the search process. In our research, we looked at several different possibilities for arriving at sensible clusters of meta-data. We added those modules that showed promise as input data for the similarity algorithm as well as search criteria for end-users. While the overall schemas provide a serviceable structure, due to the setup of our experiments and limited resources, only excerpts of the databases are currently labeled in all categories. Thus, while the goal of offering an overview of a larger set of viable game design elements has currently been reached, a dedicated effort will have to be made to reach completion in terms of labels and enriched meta-data. For this, the implementation of community-based editing will have to be the next step in terms of design and implementation. To evaluate our digital ontology from the perspective of a scholar, we conducted a series of in-depth studies on learning-oriented game design elements, which will be presented in part III (Chapters 5 & 6) of this work.

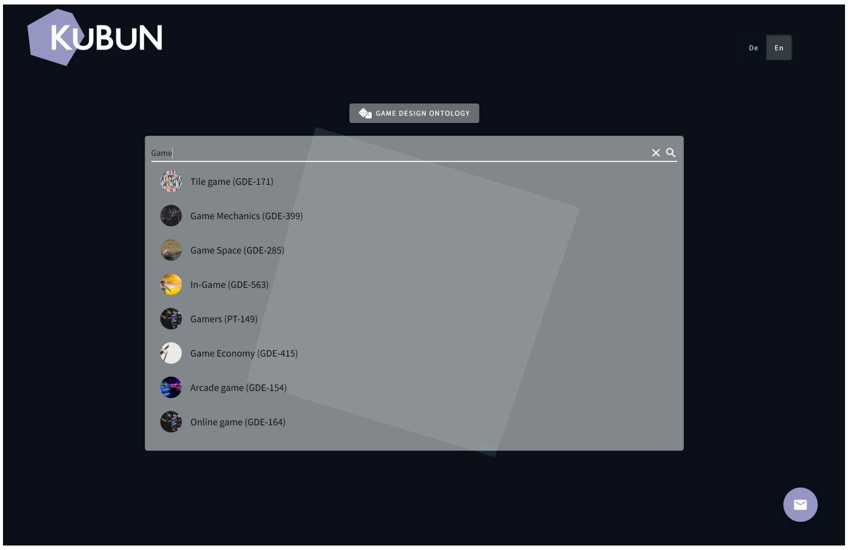


Figure 41 – Screenshot Startpage Game Design Ontology Kubun (*accessed 11.06.2022*)

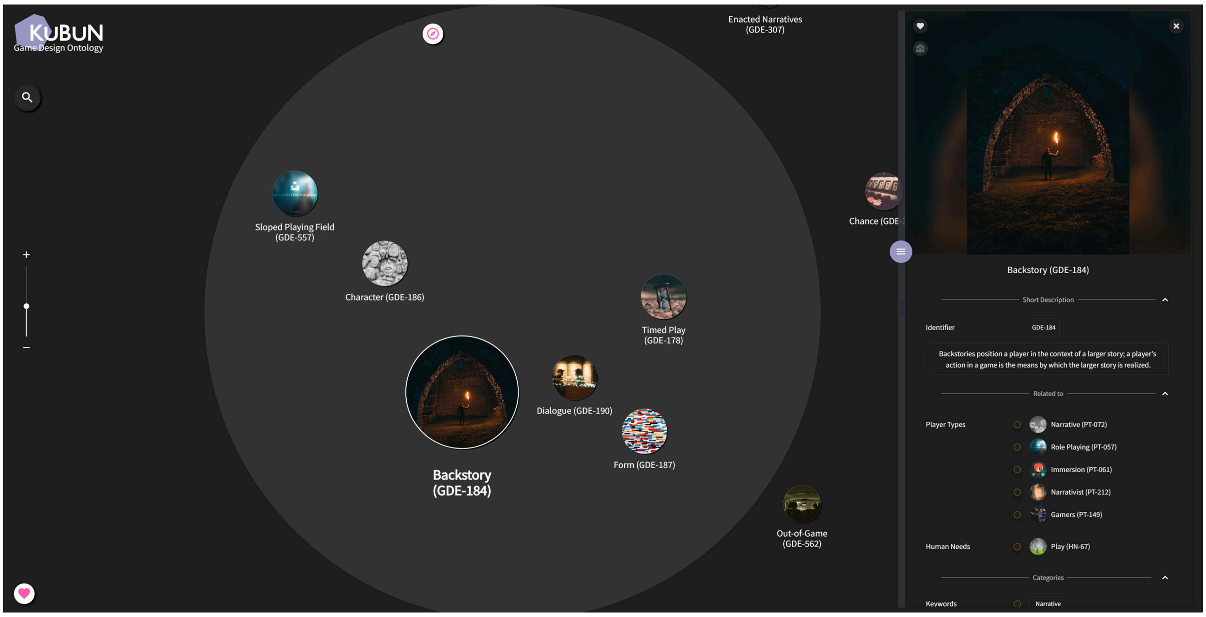


Figure 42 – Screenshot Game Design Element Kubun (*accessed 11.06.2022*)

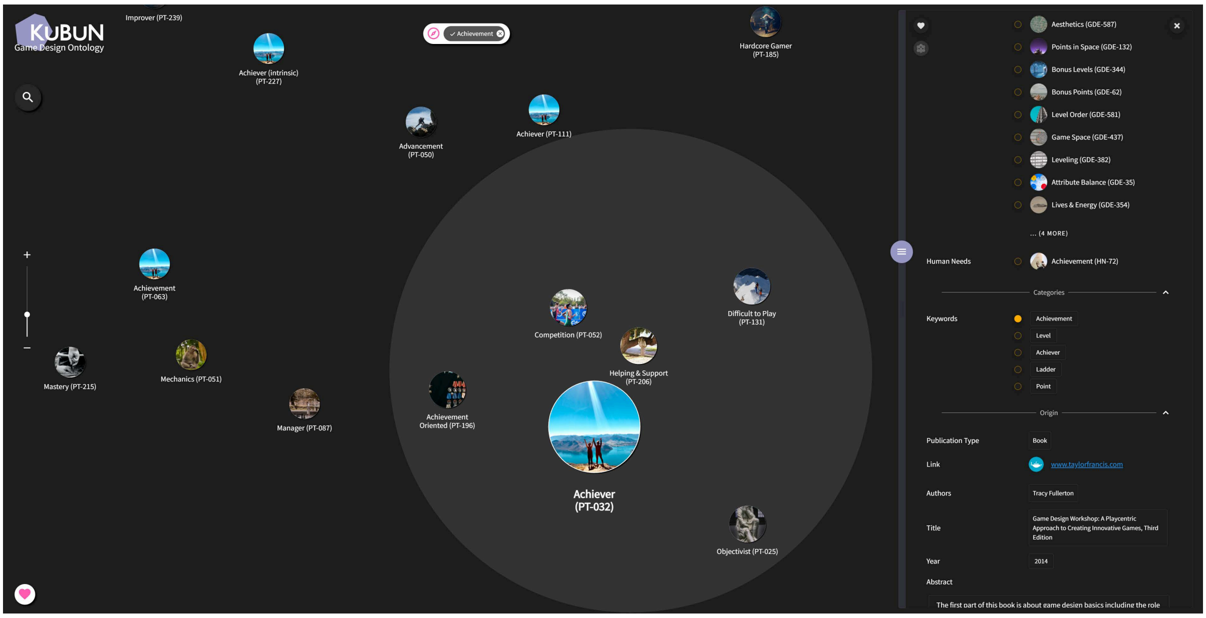


Figure 43 – Screenshot Playing Motivations Kubun (*accessed 11.06.2022*)

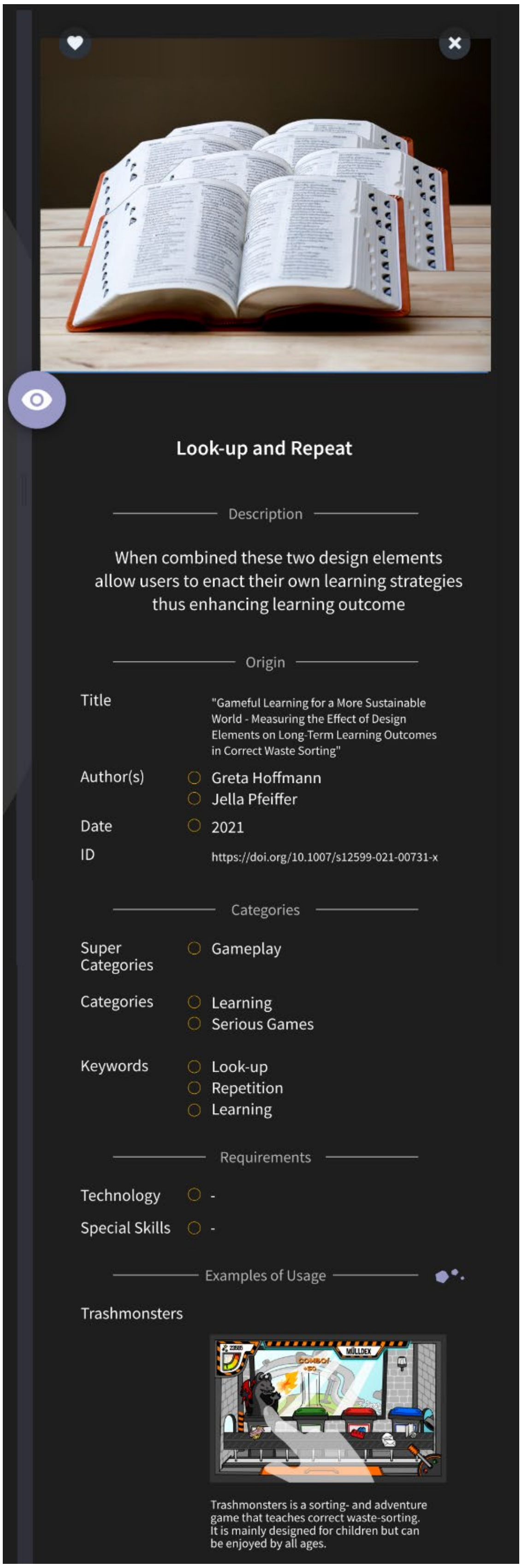


Figure 44 – Mockup: Schema Design for a Game Design Elements Kubun with Exemplary Data

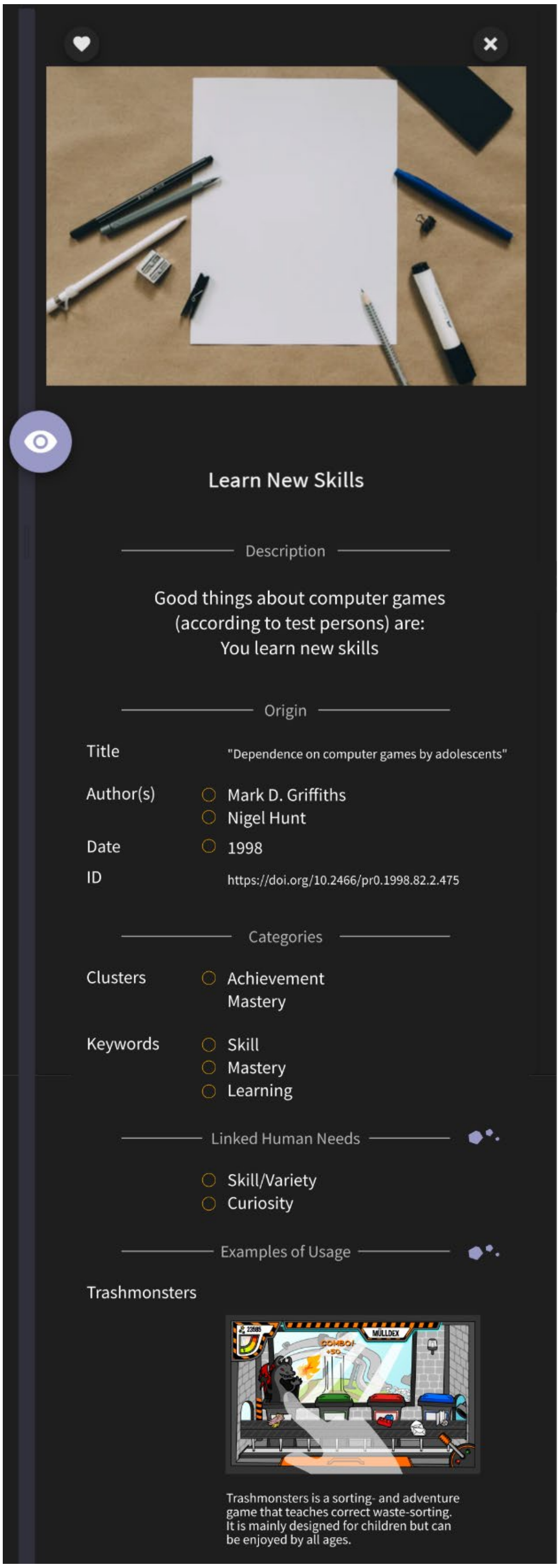


Figure 45 – Mockup: Schema Design for a Playing Motivations Kubun with Exemplary Data

Part III

Populating the Ontology

5. Chapter 5

Application and Testing of Game Design Elements: Perfect Reward

"Have no fear of perfection - you'll never reach it."

Salvador Dali

5.1. Introduction

The goal of this work is to produce a user-friendly ontology of game design elements for practitioners as well as scholars. This is reflected in the third overarching goal of our research: *to add insights from a user-based perspective to the development process*. The first part of our research is focused on game design elements from an outward perspective. To ensure that the needs of scholars are represented in the final ontology as well, in the second part of our research, we took an experimental researcher's perspective on game design element research, looking at the relationships between specific game design elements and learning-oriented outcome factors through in-depth studies.

Using a free, educative mobile game of our making, we gathered and analyzed data from a field study to design and conduct in-depth research experiments. The games' content matter is set in the domain of municipal waste sorting. Players have to correctly sort waste items into their respective bins and are thus trained to know where each item belongs. The game was designed to contribute to educative practice using the intrinsic potential of games to motivate as well as teach; thus, we deemed exploring the actual effects of specific elements in terms of their desired outcomes worthwhile to research as well as practice. For our research, the game was modified and abstracted according to the respective experimental needs (see Chapter 5.3).

We started by analyzing the field data gathered over ten months, where we first focused on emerging playing behaviors – particularly in terms of perfect play. As the game aims to teach correct waste sorting, we were particularly interested in those game design elements linked to incentivizing the best performances (highest sorting correctness) in the game. During the analysis, it emerged that the game provoked significantly different playing behaviors. To further investigate these observations from the field, we set up a laboratory experiment that tested the effect of a game design element rewarding players for perfect play. Therein, we further evaluated players' behavior in the game as well as their personality with three different established personality scales. The results of this experiment were inconclusive, equally with regard to the perfect element as well as in terms of the relationship between the tested personality scales and the emerging in-game behavior. However, we found interesting playing behaviors with regard to two additional game design elements implemented to encourage and enhance learning, namely the option to repeat a wave without penalty if dissatisfied with the result and the option to look up items in an index during play (see Chapter 5.3.4).

Based on our findings, we focused our follow-up research on these two elements (repeat and look-up) and their effect on learning outcome. We compared the findings of the first experiment with the field data and found similar behavior patterns to emerge. To measure the effect of these game design elements on learning outcome, we designed and conducted a full-factorial experiment where we compared their isolated as well as their combined effect in three different media: in-game, as a multiple-choice test, and in real life. We found significant indications that these elements had a positive effect on learning outcome, particularly when combined. However, different results were obtained between learning outcome within the same medium in contrast to learning outcome in real life, as certain configurations did not translate in one medium or the other.

5.2. Contextual Background[16]

In their set of goals for sustainable development, the UN listed targets for different areas of human and environmental wellbeing, one of which concerns waste, sustainable consumption, and production (United Nations, 2020). Acknowledging the insufficiency of the status quo in terms of waste management, the EU created a plan to raise EU-wide recycling to 55% and decrease landfill use to 10% by 2025 (European Parliament 2018). However, recent studies have shown that global progress is slow, partly due to a lack of appropriate legislation, insufficient financial resources, poor infrastructure, poor environmental attitudes and social norms, and a lack of knowledge about what goes into which bin (Filho et al., 2016; Luo et al., 2019; Schultz et al., 1995; Thomas & Sharp, 2013). A contributing factor is that many recycling and waste sorting facilities are as yet unable to reach maximum efficiency without pre-sorting measures (Bucciol et al., 2015; Hawlitschek, 2020). Countries like Germany, Austria, and Switzerland have tackled this issue by making domestic pre-sorting a citizen's responsibility. However, incentivizing citizens to correctly and consistently dispose of their household waste continues to be a challenge for society, as it is a task requiring individuals to perform for the benefit of society, often without rewards for compliance (Abdel-Shafy & Mansour, 2018). Furthermore, successful compliance requires citizens to first gain the fundamental knowledge to fulfill the required task. Yet, municipal waste sorting authorities often fail in their education attempts, partly because of outdated measures of communication and information like analog, paper-based flyers (Luo et al., 2019). Such materials are insufficient for knowledge transmission as they lack incentives to engage mentally, particularly given the amount and depth of information that people need to retain. Of the hundreds of potential waste items, more than 200 are listed on many websites for German waste management organizations as being fundamental to sufficient municipal waste sorting (e.g., Berlin and Hamburg).[17] While citizens do not have to know each item by heart, they need to understand the underlying principles that link

[16] This chapter comprises of excerpts of the authors accepted manuscript of an article published as the version of record in Business & Information Systems Engineering © 2021. https://doi.org/10.1007/s12599-021-00731-x. Full reference: Hoffmann, G., and Pfeiffer, J. "Gameful Learning for a More Sustainable World." Business & Information Systems Engineering (2021): 1-24. https://doi.org/10.1007/s12599-021-00731-x Note: The original manuscript was restructured for the context of this work, placing excerpts from the "Introduction" and "The Design Artifact" in this introductory chapter to establish the context in which all in-depth studies in Part III of this work are conducted. Figures, and were renamed to fit the structure of the thesis. Chapter and section numbering and respective cross-references were modified. Formatting and reference style was adapted and references were updated.

[17] https://www.bsr.de/die-berliner-stadtreinigung-in-leichter-sprache-24048.php
https://www.stadtreinigung.hamburg/privatkunden/abfallabc/

different types of waste to their respective bins. To engrain the knowledge in the long term, such extensive amounts of information require adequate training measures.

5.3. The Artifact – Die Müll AG (Trashmonsters)

We created a waste sorting training game based on best practices of game design as well as learning theories to address the prevalent lack of waste sorting knowledge. The downloadable app is a complete and complex game that was released in 2015 and, as of January 2022, was downloaded over 63,333 times on the Apple, Microsoft, and Windows mobile app stores ("Die Müll AG"/ "Trashmonsters" (Clocher & Hoffmann, 2014)). While playable on a PC, we designed the game with touch interaction in mind, focusing on mobile devices. We embedded the core gameplay into a small and interconnected world that represents the broader cosmos of waste sorting. The full game features an overarching story narrated through a consecutive quest structure. We aimed to motivate prolonged play through an interplay of unlockable minigames, collectible accessories, and an underlying discoverable mystery. We added these elements for players to alternate the core gameplay with additional activities connected to the general theme of waste sorting. We made each design decision with metaphorical mapping in mind.

5.3.1. Design Process

As stated by Bellotti et al. (2013), a serious game's purpose is twofold: to be fun and entertaining as well as educational. In contrast to entertainment-oriented games and applications where all aspects of the artifact can be adjusted solely to the desires and tastes of the target audience, gamified applications and serious games face the challenge of having to maintain their applicability to their original problem. For this, special effort needs to be put into the design process to ensure that the actual function of the application is not impeded but enhanced. When it comes to the goal of teaching and training content matter, games share a lot of common factors with engaged learning: *focused goals* (1), *challenging tasks* (2), *clear and compelling standards* (3), *protection from adverse consequences for initial failures* (4), *affirmation of performance* (5), *affiliation with others* (6), *novelty and variety* (7), *choice* (8), *authenticity* (9) (as aggregated from Jones (1994) and Schlechty (1997) by Dickey (2005) in her book "Engaging By Design"). This makes games inherently fundamental and powerful tools for learning (Koster, 2005). However, when it comes to specific design decisions, factors of distraction and diversion from the actual training matter can be a concern. Thus, the focus needs to be set on the specific implementations of design elements and their potential outcomes.

As the central goal of a serious game is typically related to solving or improving a real-life issue, the design process of a serious game should ensure that the design of the main activity(ies) (core gameplay) of the game ties in with desired real-life activities and effects. Best practices dictate that only after ensuring cohesion between the real-life goal and the in-game pursuit of the abstracted goal, as well as establishing intrinsic fun with this activity should other supplemental design activities follow. Such supplemental design can pertain to secondary goals that support the success of the fundamental objective. For example, when it comes to achieving a learning objective, like in our case, it can be necessary to make additional adjustments to ensure longer engagement with the main task. According to the literature on long-term retainment of knowledge, for a training measure to be effective, the brain must be exposed to the learning content for a certain amount of time before the trained information is pushed from the short-term to long-term memory (Atkinson & Shiffrin, 1968). Reports vary in regards to the specific amount of time that needs to pass as many factors can influence retention positively or negatively depending on the situation and person; however, for most teachable content, a certain amount of exposure must be ensured (Hintzman, 1976). Thus, one of the biggest design challenges in learning-oriented applications is to ensure continuous motivation to interact with the artifact itself for the learning contents to be moved from short-term to long-term memory (Baddeley & Hitch, 1974; Kelley & Whatson, 2013).

Another important factor for training and retention is rehearsals (Bygate, 1996). Depending on the amount of content, several sessions might be necessary to transmit the desired learning content to the user/player. Since the speed of learning and internalizing depends on different user-inherent factors like intrinsic interest in the topic, mental processing speed, focusing capacity, mode of learning (Curry et al. 2005; Lim et al. 2007), as well as extrinsic factors like location and environment (Hanrahan, 1998), it is important to implement a functioning composition of design elements that can succeed in overcoming the user's stagnation in learning motivation.

Based on these factors, the design of the game was divided into three parts: 1) deciding on the *boundary conditions* and *overarching design choices,* 2) establishing solid *core gameplay* that directly works towards the main goal of the game – teaching and training correct waste sorting and 3) designing *complementary supportive components* contributing to the main goal of the game by achieving secondary goals like retaining player interest to ensure *continuous engagement* with the content, encouraging them towards *optimal performances* and building up an *emotional attachment* to the in-game characters and world to support long-term behavior change in real-life through affect and empathy.

5.3.2. Boundary Conditions & General Design Decisions

We started our design process by consulting with the local waste management institution and conducting a set of ten informal interviews with a variety of members of the public (n=10, in 5 Minute Street-Interviews). Resulting, we aggregated the key problems concerning incorrect waste sorting. 1) A lack of participation due to *misinformation*: several interviewees named disbelief in the effectiveness of the domestic waste sorting process, citing reasons that were often founded in false information (most commonly that all waste was being burned afterward anyway) – highlighting a lack of awareness of the post-processing of different waste categories. Similar insights have recently been confirmed by the 'Waste separation works' initiative, listing several common prejudices concerning waste sorting in their online campaign (Gemeinsame Stelle dualer Systeme Deutschlands GmbH, 2021). 2) A lack of participation due to *lack of habituation*: The experts at the waste management institution indicated that those citizens that had moved in from regions with little or no preexisting waste sorting procedures were less likely to participate in the process successfully. 3) *Insufficient knowledge* about specific waste items: even those participants that claimed to be actively participating in domestic waste sorting were often not able to correctly identify the correct bin when prompted with a sample selection of more ambiguous waste items.

Based on these factors, we derived general design principles for the game:

1) Accessibility: the game should aim to reach as many households as possible, unlimited by financial limitations or language barriers; 2) Habituation: the game should be targeted towards young audiences to habituate sorting principles and benefits from early ages; 3) Transparent communication: the game should address known prejudices and misinformation and clear them up; 4) Light-heartedness: the way problems

and teaching content are addressed within the game should be fun and engaging; 5) Empathy: the game should create feelings of sympathy and empathy within the players towards the world of waste sorting and all the participants within it; 6) Delight: the game should evoke feelings of pleasure and anticipation within the players.

Addressing the first two principles, we chose children aged 8-12 as the target audience. As the topic is typically addressed during elementary school in Germany (BMUV 2021), the game can be embedded into the curriculum. In terms of distribution platforms, we chose the three major mobile platforms (iOS, Android, and Windows) for mobile devices as these were at the time - and are still - the most common game platform across households for this age group (GAME, 2020). To ensure accessibility independent of financial factors, the game is downloadable free of charge. And to address potential language barriers, we added translations for English and French and developed a localization tool to facilitate the inclusion of additional languages.

The next two design principles particularly influenced the tone of the in-game texts: e. g. in the in-game waste index ("Trashdex"), players can look up where each waste item goes and why they belong there and not in another bin. Explanations are given for each item, and common misunderstandings are addressed. All in-game dialogues are designed with different communicative approaches working in tandem through the different characters (e. g. residual waste monster: strict and easily offended by wrong sorting vs. recycling monster: forgiving and calmly explaining vs. biowaste monster: puns and dad jokes serving as comic relief vs. paper recycling: outsiders view – serving as "the inside man").

The final two design principles particularly influenced the aesthetical choices: the visual and the auditive design. The visual style is set in a child-oriented cartoon aesthetic. The color scheme is based on the official regional waste disposal colors and is kept in bright and vivid tones. To allow for HD-optics across the different types of mobile devices, we used SVG-based vector graphics. Special focus, particularly in the design of waste items that might induce discomfort in handling (e. g., used female products, diapers, hairs, and nails), was set on finding neutral ways of depiction that don't deter from an enjoyable playing experience. The soundtrack consists of over 30 different tracks for each location within the game, as well as different states of the core gameplay and over 260 sound effects. Also, many small features (like small animations, different weather states, as well as a day and night cycle) were added for the simple purpose of inducing delight in the players. As Adams states in his work on game design: "An ugly or awkward video game is a bad one, no matter how innovative its design or impressive its technology. Part of your job is to give your players aesthetic pleasure." (Adams 2014, p.21).

General Setting

We designed the overall setting as a waste-themed fantasy-world set on a small planet that hosts locations for its different waste-related inhabitants. The overall world represents a metaphorical holistic view of the waste management process. The planet is inhabited by monsters that represent the different waste recycling processes (see Figure 46).

Their character design in terms of looks and personality was linked to the type of waste they represent – Recycling – resourceful, creative, optimistic; Bio – derpy, smelly but likable; Paper – young, clean, fresh, and Residual waste – grumpy, cynic but secretly gold-hearted. They all live and work together on the planet as it is their job to take care of the different city planets' waste. Their homes can be visited by the player where they can talk to the monsters, receive and complete quests and gift them items of clothing. Through these different social interactions with the monsters, we aimed to achieve emotional involvement and commitment to the topic. The planet overview screen connects the games' different locations. These locations are i) the waste sorting facility, where the core gameplay takes place, ii) the monsters' living spaces, where players accept quests and different minigames can be played, and iii) the info center, where players can get information on the current state of the game (pollution, sorting correctness and unlocked quests). Adjacent to the waste planet is another smaller planet that represents the respective waste supplier (in the current version, the waste system of Karlsruhe). As each region in Germany has autonomy in its choice of a waste management system, the game is designed to switch systems according to the city-planet to which it is connected.

Figure 46 – Game Aesthetics of the Unabridged Game (World and Characters)

Player Character

We chose a 2D first-person perspective where the players interact with the game as themselves and are addressed directly. We decided against an intermediate character to keep the attribution of all in-game actions and successes as close to the players as possible to facilitate and suggest a reproduction of their in-game actions in real life. According to Gee (2003), effective learning involves "playing a character." For example, learning in a science class works best if students "think, act and value like scientists." This assumption is supported by the findings of a psychological study where participants who were given a virtual body (avatar) communicated as Einstein (signifying super-intelligence) performed significantly better than participants of the control group, considering prior cognitive ability (Banakou et al., 2018). Studies concluded that players learn best when they are engaged in meaningful, goal-directed activities within the identities of experts (Shaffer et al. 2005). As such, we designed the monsters' dialogues to directly communicate to the player in their role as a new and essential member of the workforce. From the beginning, players are directly involved by the monsters to help with their struggle to deal with the overwhelming amounts of waste they receive by presorting them and being thanked for their hard work.

5.3.3. Core Gameplay

Core gameplay (or core gameplay loop) has been defined as the primary game system or mechanic that defines your title (Guardiola, 2016) or a set of activities that the player will undertake most frequently during the game experience and which are indispensable to win the game (Fabricatore, 2007). Establishing fun core gameplay is of essential importance before proceeding to the design and implementation of any other design elements (Järvinen, 2007; Sicart, 2008). Since the core gameplay is the task players will be engaging with the most in the game, the task itself must be fun-inducing, not -draining. Especially since the motivation to interact with the content is solely intrinsic (Deci et al. 2001), the core gameplay has to be engaging enough to sustain the ongoing motivation to interact with the game system. Thus, before going into full development, we tested an isolated version of the core gameplay with 20 play-testers. The game went through several iterations before the parameters were finally set. The tests were conducted in the manner of the quiet observer, as is common in user experience testing, with a follow-up session to discuss the highlights and flaws and make suggestions for the gameplay mechanism.

Setting

Figure 47 – Core Gameplay

To keep a coherent metaphor with the general setting, we set the core gameplay within a factory. Four waste bins (paper, recycling, bio, and residual waste—reflecting the system in Karlsruhe) are placed next to each other behind a conveyor belt. Each bin is inhabited by a monster representing the respective post-process of the waste. They react to each sort depending on the correctness of the players' decisions (see Figure 47). Serving as a more global indicator of the players' current sorting trajectory, a window in the back shows the state of the outside planet. If the players keep the ratio of correctly sorted waste high, the world outside looks pristine, but if they sort too many waste items incorrectly or put too many items into the residual waste, it deteriorates into declining degrees of being trashed and polluted.

Core Mechanic

The core interaction consists of picking up waste items that come down the conveyor belt and dragging and dropping them onto the correct bin. By using a touch-and-swipe-input as the main game interaction, we aimed to build a closer connection between the in-game action and the real-life action of moving the waste to the right place. The items spawn on the right side of the screen from where they are moved by the belt to the left side of the screen and need to be sorted during this time. An item that drops off the conveyer belt on the left counts as unsorted and raises the counter of the waste pollution bar, leading to a littering-based Game Over. If, on the other hand, it is sorted incorrectly, it is counted towards an air-pollution-based Game Over. The over 200 waste items are split into waves of 15 to 35 waste items to pace the game and achieve more winning states. At the end of each wave, the number of correctly sorted waste items is displayed. Depending on the percentage, the monsters react differently to that number.

Feedback System

Feedback should be immediate and comprehensible in terms of the failure or success of the given task (Sicart, 2008), with rewards and advancement in the game carefully bound to it (Bellotti et al., 2013), which is an established rule in games. Thus, we implemented a positive/negative reinforcement system: points (+10/-3) for right vs. wrong sorting of an item, visual/audio feedback of the monsters (joy/anger), combos (+50 points for a correct three-item streak), and combo-breakers (disruptions of the combo counter upon missorting within a streak). A numeric score and a pollution counter (top left-hand corner in Figure 48) provide feedback on the overall performance, warning players of an impending Game Over. This counter fills up each time an item is placed in an incorrect bin or drops off the lane and is reduced when an item is placed in the correct waste bin. An appropriate chunking of tasks helps provide a flow experience (Csikszentmihalyi et al., 2005). Inspired by the successful two-minute format of game applications like Angry Birds (Rovio Entertainment, 2009), we chunked the learning content into waves that do not exceed playtimes of two minutes to encourage shorter but more frequent playtimes. Following advice by Wolfe et al. (1998), we implemented a structure blending the previously learned items with newly introduced ones.

Tutorial

As is common practice within games, the first three waves serve as tutorials and differ from regular gameplay (Gee 2003). In the first wave, we present the main types of waste (recyclable, bio-degradable, paper, and refuse) with an explanation of the underlying attributes with which players can infer the correct bin for each waste item (e.g., inextricably compounded materials go to residual waste). In the second wave, players are supposed to familiarize themselves with the core gameplay through representative waste items for each type. In the third wave, we introduced additional design elements that accompany the core gameplay: the look-up element and the pollution counter, which indicates how close players are to a potential Game Over. In the experiment, we introduced only the pollution counter but not the look-up element to the groups without the look-up element.

Depiction of Knowledge Items

"Advertisers have learned through trial and error, focus groups, and intuition that people 's behavior and attitudes are governed by a cognitive system that is more responsive to pictures than to words." (Epstein 2014, p.31). For experimental evidence, see Clark and Paivio (1987). When deciding on the presentation of knowledge items for the game, we consulted the literature on the mental representation of knowledge. During the learning process, different types of memory connections are formed (e.g., typical connections in mathematical didactics are numeric, graphic, situational and algebraic (Nitsch et al., 2016)). Two of the most common items are words (designated representation) and pictures (iconic representation) (Kolers & Brison, 1984). According to Mayer's theory of multimedia learning (2002), active learning entails the coordinated stimulation of both channels of the human information processing system (visual/pictorial and auditory/verbal processing). For our game, we chose to depict our knowledge items (waste items) with a combination of iconic and designated memory connection items through sticker-like pictures and by displaying the name of the waste item when picked up (see Figure 48). We selected the 223 waste items used for the experiment from a list provided by the Amt für Abfallwirtschaft (AfA) Karlsruhe based on the following criteria: 1) relevance (loss of precious resources if sorted incorrectly), 2) frequency of appearance in common households and 3) difficulty (frequency of missorting in real life).

Figure 48 – Metaphorical Representation (Mapping) of the Waste Sorting Process in the Core Gameplay of the Artefact

5.3.4. Complementary Supportive Components

Learning Enhancing Design Elements

During the initial testing sessions of the core gameplay, we found that to successfully increase knowledge on correct waste sorting, we needed to improve the game design concerning player motivation. The first goal was to enhance players' motivation to increase their sorting performances, achieving *depth of knowledge* (the likelihood for an item to be allocated correctly). We added the following game design elements:

Perfect Reward

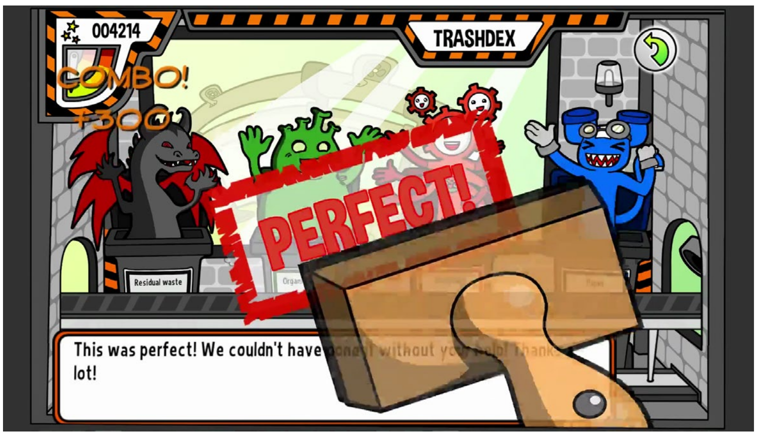


Figure 49 – Perfect Reward Element

To reward players for finishing a wave perfectly, we added a design element consisting of a "perfect!"-stamp combined with a gratifying sound effect and specific victory music (see Figure 49). If awarded, players move straight on to the next level without being given the option to repeat.

Repetition-based Design Element

If a level is not completed perfectly, the game shows players how many items they sorted incorrectly and offers them the chance to repeat the level without penalty (see Figure 50). We strategically placed and colored the "yes" and "no" buttons to favor repetition. If players choose to repeat, their level of pollution is reset to the level when that wave was played for the first time. We were inspired by the quick trial, immediate performance feedback, and low inhibition retrial-loop pattern of games like Cut the Rope (ZeptoLab, 2010) and Angry Birds (Rovio Entertainment, 2009).

Figure 50 – Repeat Element

Look-Up-based Design Element

In his article, Gee (2003) elaborated on the placement of information: that it should be given "on-demand" and applied soon after having read it. He based this on people's poor understanding, and retention of information received out of context (Barsalou, 1999; Brown et al., 1989; Glenberg & Robertson, 1999). The look-up element (see Figure 51) is an index (called "Trashdex" in the game) that can be used to find all previously encountered waste items. For each item, it shows the correct target bin, as well as additional information on why the item belongs there and not in another bin. It can be accessed at any point throughout the game by simply opening it or by pulling an item on top of it (it then scrolls directly to that item). It is introduced in the tutorial, and its usage is penalty-free. For the mechanics of this look-up design element, two game design elements that serve to offer additional information to the players inspired us. First, we drew insights from the interactive "hint" functions found in puzzle games and point-and-click adventures like Machinarium (Amanita Design, 2009). These hints are designed to reduce frustration by guiding the players with incremental tips. They are optional, so players decide for themselves when and if they want to use them. The second inspirational game design element is the pokédex used in the Pokémon (Game Freak, 1996) game series: a lexicon-based design element that gradually lists all monsters and their related meta-data that players encounter during the game.

Figure 51 – Look-up Element

Content for Prolonged Engagement

The second goal was to have players engage with the content as long as possible, increasing the *breadth of knowledge* (number of trained items). For this, we focused on design elements tying the core gameplay to the outside world of the game.

Story, Quests, and Mystery

Building on the incentive of exploration, we created a whole planet to explore and interact with in between sorting sessions. Players can visit each monster in their home and explore their personalities through conversation. A questline is connected to each monster, resulting in the unlocking of quests, minigames, and additional areas within the game. The questline and story progress are regulated through the game waves, and new content is unlocked after each wave. Underlying the more mundane story points of the quest structure that relate to everyday occurrences like repair and match-making, a global narrative is introduced in the middle of the game: the planet's volcano starts to reactivate, threatening to obliterate the planet and its inhabitants. The monsters and the player work together to prevent this; however, towards the end of the game, it still happens: The volcano erupts into a fountain of waste that had been accumulating within the core of the planet over many years. This initiates the final waves of the game, where players have to sort double and triple amounts of waste at maximum speed to get on top of the emergency. Related to this main story, there is an underlying mystery surrounding the planet and its history that only curious and meticulous players will unveil (see Figure 52).

Figure 52 – Mystery Surrounding the Volcano

Minigames, Accessories, and Upgrades

As a means to embed other waste management processes as well as offering another incentive to come back to the core gameplay, we implemented several minigames that are unlocked during the progression of the story. Each minigame relates to other waste categories that are not represented within the waste-bin system (like battery- or glass recycling). They require a certain number of their respective special waste items (glass bottles, twigs & leaves, batteries) that are collected during the core gameplay. Thus, players are incentivized to come back and play another wave until they have enough items to play the respective minigame. The first minigame represents the inner workings of a composting plant and is inspired by the mobile game Fruit Ninja (Halfbrick Studios pty. ltd., 2010), where players have to cut down the garden waste until it is compostable (see Figure 53 right). The second minigame represents the process of glass separation at the glass container with a quick-sort mechanic (see Figure 53 center), and the third represents operating principles of a battery recycling process where players have to manipulate heat and cold to break them down into their chemical components. Every time players successfully complete a minigame, they are awarded one of nine accessories that they can present to any monster on the planet as wearables. By collecting a set of valuable items during the core gameplay, players can also unlock three upgrades to the core gameplay: a lever that is unlocked in two parts and allows players to either slow down or speed up the conveyor belt and a second conveyor belt that transports waste that would otherwise have fallen off the first one back across the screen. These items give players more control over the main game, allow them to explore the outer edges of their competency (Gee, 2003), or take away some of the pressure (see Figure 53 left). Other extra interactions include a guide for paper recycling at home or the training of a dog to pick up bulky waste that occasionally blocks the main game.

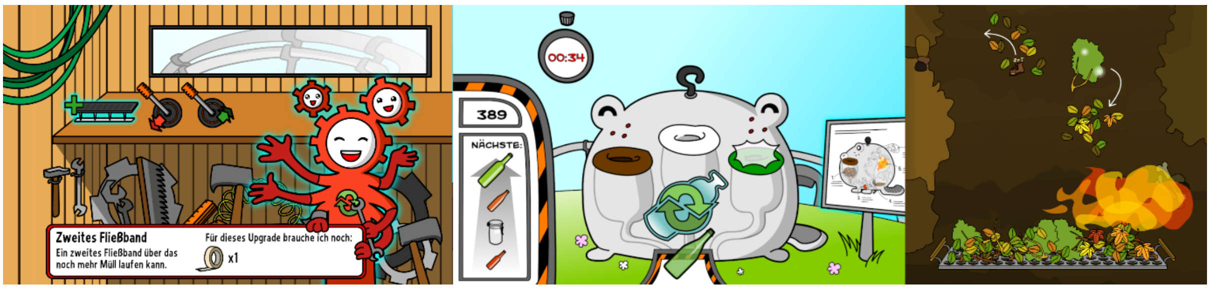


Figure 53 – Upgrades, Bottle-Sort and Garden-Waste Minigames

5.4. Pre-study – Evaluation of Field Data Concerning Motivations for Perfect Play[18]

The overall goal of the game is to teach its users correct waste sorting. In preparation for the development of the game, we evaluated lists of the most incorrectly sorted waste items provided by the department for waste management in Karlsruhe (Amt für Abfallwirtschaft (AfA) Karlsruhe, n.d.). Of the over 220 waste items we included, 150 were items often sorted incorrectly. In our design, we were concerned about motivating players that encountered an item they did not know where to sort to either directly look up its correct bin via the look-up element or repeat and try again (with or without looking it up). To encourage this behavior, we implemented a suggestive final screen at the end of a sorting (wave), showing players the number of items they missed for a perfect run (see Chapter 5.3.4). To further reward and reinforce this behavior, we added the perfect reward game element. To assess the effectiveness of our design choices in terms of our goals – achieving *depth of knowledge* (quality-related: certainty on single items regarding their correct bin) and *breadth of knowledge* (quantity-related: having encountered many waste items) in our players, we set up a series of studies where we evaluate the effects of single design elements with regards to their intended outcomes.

Our first focus was to look deeper into the depth of knowledge we can achieve in our player base. We started by assessing the success of our design in terms of the degree we managed to incentivize players to aim for perfect sorting results. For this, we wanted to understand to what degree opting for perfection is an intrinsic quality relating to personality and interest and what type of incentive would be able to alter non-perfectionistic behavior towards a more desirable outcome. To gain first insights into an expectable baseline of perfectionist behavior in our game, we conducted an exploratory data analysis of data collected during a field study of the game. We also wanted to assess the effectiveness of changes we had made after a pretest (slight visual workovers, adjusting the number of items per wave, and reworking the tutorial). For this, we differentiate between data collected during the pretest and data retrieved during the following field study.

5.4.1. Method

We started the field study with a preliminary study (pretest) in November 2015 that ran for three months. During this time, we fixed smaller design issues (visual workovers, adjusting the number of items per wave, and reworking the tutorial) and then launched the actual field study. The data we used for the following analysis consists of data aggregated during the pretest that we compare with data aggregated throughout the first six months of the field study. For the analysis, we first used Unity Analytics (Unity Technologies, 2014) to access player data and identify unique players by their IDs and then conducted our analyses with Python. We filtered out corrupt datasets where data was transmitted incompletely. The tutorial waves (the first three waves of the game) were discounted. Players who quit during this time were not included in the analysis.

[18] This chapter comprises excerpts of an article that was published in the following outlet under the following title: Hoffmann, G., Martin, R., Weinhardt, C. "Perfectionism in Games-Analyzing Playing Behaviors in an Educational Game." 2019 11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games). IEEE, 2019. 10.1109/VS-Games.2019.8864542. Note: Tables and figures were renamed, reformatted, and newly referenced to fit the structure of the thesis. Chapter and section numbering and respective cross-references were modified. Formatting and reference style was adapted and references were updated.

In this pre-study, we were particularly interested in two parameters: *performance* (percentage of sorting correctness) and *commitment* (number of waves played). As players could repeat waves to achieve higher final scores, we included the number of repetitions in our analysis and used the *score of the last played wave* as the *performance metric*. Using these parameters, we created playing behavior plots for each participant (see e. g. Figures 54 and 55). Based on the generated plots, we isolated players with a perfect track record: ending each unique wave with a perfect score before proceeding to the next. During this process, we stumbled upon another distinct but opposite behavior pattern with regards to the repetition function. While the players we isolated as a group that we refer to as "Perfectionists" often repeated waves several times to achieve their perfect performance, this other group stood out by never repeating a single wave. We accordingly refer to them as "Rushers." After assessing the data in terms of repeated emergence of these behaviors, we defined that players fall under the "Perfectionist" group if their performance was 100% in more than 80% of the waves they played before quitting or finishing the game (see Figure 55). We allowed for a 20% latitude due to deviations stemming from i) special waves that introduce special mechanics as well as ii) an observed drop in behavior shortly before the end of a playing session. A similar occurrence of loss of contribution to an activity can be found in game theory and are there referred to as an end-game effect[19] (Selten & Stoecker, 1986). We further defined the "Rushers" as players that have not used the repeat option in 80% or more of the waves they played (example plot: see Figure 55). Here the 20% latitude was included affording for cases of unintentionally hitting the repeat button due to its suggestive design (see Chapter 5.3.4). We chose the 20% threshold based on observations in the plots, given the number of players that deviated from the behavior patterns by not more than one- or two outliers.

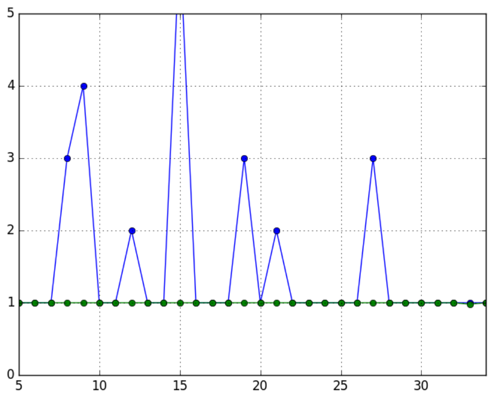


Figure 54 – Playing Behavior Perfectionist
(green: Score, blue: Number of Repetitions)

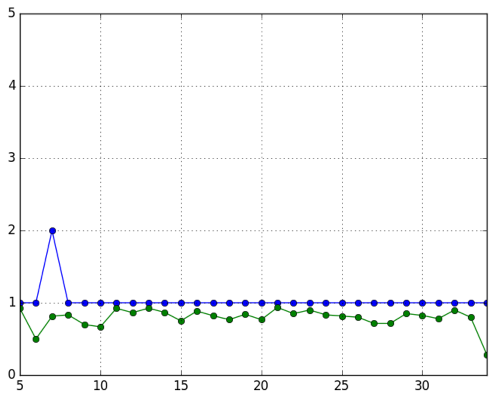


Figure 55 – Playing Behavior Rusher
(green: Score, blue: Number of Repetitions)

We compared both groups regarding their commitment by dividing the players into groups according to exit peaks (waves after which a big group of players left); see Figure 56. We chose this metric as the exit spikes seemed to indicate collective losses of interest and thus a potential underlying factor contributing to this behavior.

[19] While the endgame effect refers to a loss of cooperation at the end of an interaction, we argue that a transfer can be made given the contribution of players' time to the waste sorting effort within the game.

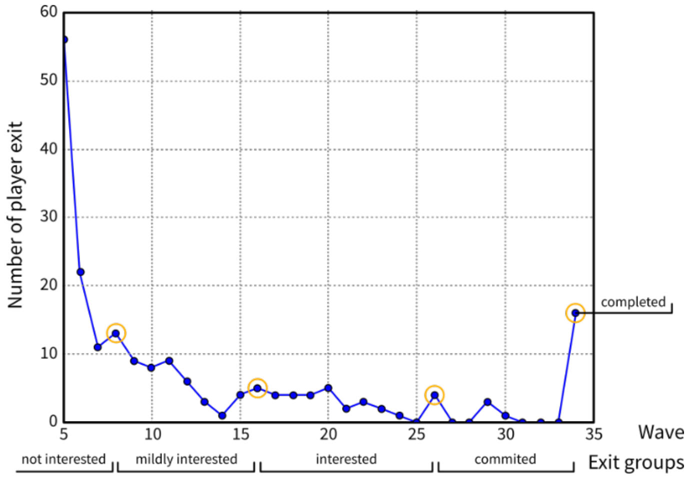


Figure 56 – Plot Player Exits per Wave during Field Study

We labeled the groups "not interested" (left between waves 1-8), "mildly interested" (left between waves 9-16), "interested" (left between waves 17-26), "committed" (left between wave 27-34) and, as a *subgroup of the "committed" group*: "completed" (finished the game: wave 34). A complete playthrough of the game takes an inexperienced player between 4-6 hours. To better compare the differences between these two groups, we further introduced the parameter of *tenacity* (repetitions per wave) as a measure of comparison.

5.4.2. Results

The findings of this pre-study are based on the analysis of the aggregated, anonymized user data from a period of nine months: one batch of data from the pretest (three months) and one from the field study (six months). Of the total of 1660 downloads the game could accumulate to the end of the study, data of 1045 players was successfully collected via Unity Analytics, of which 184 had to be discarded because of incomplete datasets. The pretest accumulated 139 identifiable players and therein a total of 118 uncorrupted datasets, and the field v 906 players with a total of 743 uncorrupted datasets. The distribution of the behaviors in terms of commitment is depicted in Table 13:

Table 13 – Distribution of Players that fall into Playing Behavior Categories by Commitment

	Pretest (n=118)				**Field Study (n=743)**			
	Rushers		Perfectionists		**Rushers**		**Perfectionists**	
Exit-Groups	n (% of Rushers)	n (% of total)	n (% of Perf.)	n (% of total)	n (% of Rushers)	n (% of total)	n (% of Perf.)	n (% of total)
Not interested	8 (44%)	6.78%	2 (33%)	1.69%	99 (70%)	13.32%	31 (56%)	4.17%
Mildly interested	6 (33%)	5.08%	2 (33%)	1.69%	18 (13%)	2.42%	11 (20%)	1.48%
Interested	3 (16%)	2.54%	1 (17%)	0.85%	14 (10%)	1.88%	4 (7%)	0.53%
Committed	1 (6%)	0.85%	1 (17%)	0.85%	11 (8%)	1.48%	9 (16%)	1.21%
(Completed)	0 (0%)	0%	0 (0%)	0%	11 (8%)	1.48%	9 (16%)	1.21%
Total	**18**	**15%**	**6**	**5%**	**142**	**19%**	**55**	**7%**

A paired t-test of the average *last performance* (performance of each last played wave if it was repeated) shows that the Perfectionists' performance was significantly higher than that of the Rushers in both the pretest (T = 3.83; $p < .001$) and the field study (T = -10.61; $p < .001$). The measured *tenacity* also showed a significant difference between the groups in the field study (T= -7.19; $p < .001$) but not in the pretest (T = 1.80; p = 0.085).

Table 14 – Comparison of Playing Performance & Effort

	Perfectionists		*Rushers*	
	Average last performance	**Repetitions per wave**	**Average last performance**	**Repetitions per wave**
Pretest	99.9%	1.75	76.9%	1.43
Field Study	99.0%	2.14	81.3%	1.74

5.4.3. Discussion & Conclusion

In this study, we gained first insights into the effectiveness of our game design decisions in terms of increasing the breadth and depth of waste-sorting knowledge. Consistent with other studies that evaluated perfect playing behavior, the game in its tested configuration managed to incentivize a certain number of players to optimize performance during their playthroughs (see Table 14 – Perfectionists). During the selection process, we identified a secondary group (Rushers) three times the size of the Perfectionist group. These players seemed to be mainly incentivized to interact with the secondary content of the game (the explorative design elements like minigames and side-quests all relating to the topic of waste management). This group showed the opposite of the intended behavior, not responding to the performance-enhancing design elements at all, but proceeding further in terms of the overall game before quitting.

We compared the pre-test to the field data, as we had made adjustments to the game and additionally wanted to assess their effects with regards to emerging differences in playing behavior. Analyzing the datasets of the pre-test, we found that a total of 5% of the players behaved according to the *Perfectionist* pattern, while 15% of the players could be distinguished as *Rushers*. In the actual field study, both numbers increased. Now 7% of the players were identified as *Perfectionists* and 19% as *Rushers* indicating a positive tendency with regard to our design decisions. Even though the pretest delivered a smaller set of data compared to the field study, the same types of player behavior could be identified in both versions. The percentages indicating the prevalence of both the *Perfectionists* as well as the *Rushers* in the player pool were slightly higher in the field test compared to the pretest. In terms of *commitment* (how long the players continued to interact with the game), more players in the Rusher group played the game from beginning to end. Overall, more Perfectionists completed the game (see Table 13). To summarize, while the Perfectionists dominated in terms of performance and tenacity, Rushers tended to see more of the game as evidenced by their higher commitment (in this study, commitment refers to seeing all of the game, not necessarily to the underlying learning content itself). It is possible that the Perfectionists' compulsion to repeat a wave until it is perfected might have disrupted the game fun and thus influenced their motivation to progress in the game as the originally designed flow of playing the core gameplay and then taking players' minds off of the core gameplay on the main planet was skewed towards the core gameplay. It is possible that their strive for perfection quickly depleted the general motivation to follow through with the whole game.

The findings of this first analysis confirm the emergence of Perfectionist behavior in a small but consistent percentage of the playing popularity of our game. We further found indications for another complementary playing behavior (Rusher) that, despite not aiming for high performance, gained approximately equal amounts of overall playtime through their continued interaction with the other parts of the game, thus potentially having learned about waste sorting in equal amount if not depth.

This field study served as a first step toward testing the effects and outcomes of our design decisions. In the following study, a laboratory experiment designed to test the internal validity of our findings, we aim to quantify the role of personality in playing behavior and the effect of our perfect reward.

5.5. Experimental Study 1 – Influence of Player Personality on Perfect Play

After conducting the field study as described in the previous chapter, we were intrigued by the results in terms of the two antithetical playing behaviors we identified from our data analysis. As those two behaviors seemed to occur mutually exclusive from each other, we aimed to gain further insights into potential connections of different personality types related to these behaviors.

The theory of intrinsic and extrinsic motivation (Ryan & Deci, 2000) depicts a motivational scale that progressively moves from amotivation over different stages of extrinsic motivation toward intrinsic motivation. According to the theory, extrinsic motivation can negatively affect existing intrinsic motivation, as proven through different psychological studies (Loveland & Olley, 1979). In the design of our game, we intend to nudge players towards optimal performances while trying to prevent them from losing interest and quitting altogether. Thus, the question arises whether rewarding in-game efforts, particularly for perfect play, is beneficial or detrimental to motivation.

In the game, if players do not sort perfectly, we incentivize perfect play through a feedback element at the end of each wave, where we highlight the distance between a perfect and the actual result and ask players if they would like to repeat (see Chapter 5.3.4). Perfect play is rewarded with positive visual and auditive feedback through a perfect stamp and cheers from the monsters. Seeing that a comparatively small amount of the player base was motivated to aim for perfect play (7% in the field study – see Chapter 5.4.2), we wanted to build on this finding through an experimental study to gain more in-depth insights into the effectivity of our current set of design elements.

The overall research interests of this study are twofold. First, we want to gain insights into the effectiveness of the reward element in incentivizing perfect play. For this, we conducted a lab experiment to measure players' motivation for perfect play under the following research question: *Does rewarding perfect play result in continued willingness to achieve perfect play?*

Secondly, we want to understand the potential effects of underlying personality traits on the playing behavior patterns we found in the field study (see Chapter 5.4) under the following research question: *Can factors of personality be identified that influence the willingness to repeat a task to achieve a perfect run?*

To gain qualitative insights into the effectivity of our game design, we designed and conducted a laboratory experiment with a control and treatment group where we analyzed playing behavior with and without the reward element. To build on the findings of the field test and test their internal validity, we added a segment with three psychological studies (PID (Betsch, 2004), DMT (Misuraca et al., 2015), and BFI-10 (Rammstedt & John, 2007) to identify potential connections between these tests' items and the playing behaviors we had identified in the field test. Finally, we added a qualitative segment at the end to gather rationalizations behind players' in-game choices.

Our results showed that the perfect reward element did not affect the willingness of players to opt for perfect scores. However, in the experiment, the playing behaviors found in the field study (Perfectionists and Rushers) could be reproduced in significantly higher numbers. In terms of personality, we found a significant connection between Perfectionists and preference for intuition, which in addition to our insignificant results in terms of the perfect reward, implies that Perfectionist behavior is intrinsically driven and does not benefit from external rewards.

We contribute to the literature by adding a study on a specific game design element analyzed under controlled conditions, which, particularly on the topic of perfectionism, is currently still very rare. By linking in-game playing behaviors to psychological tests, we further add to player-type research through two consolidated playing behaviors, one of which can be linked to a psychometric construct.

5.5.1. Related Literature

We started our research with a literature review of studies that analyzed perfect play within gameful setups.[20] We narrowed down the search results to the 48 papers most suitable to our research context (using experimental setups to evaluate specific game design elements in terms of performance-related parameters) that we then cross-examined for relevance to our research. Of these, only three studies emerged that suited our research context of evaluating gameful design on behavioral outcomes, specifically perfectionism. First, in a study conducted by Lisitsyna et al. (2015), they summarize the results of three experiments analyzing participants' performances in an online course (n=11.319, with n=417 finishing all exercises), of which two investigate the effects of perfectionism (Lisitsyna et al., 2015). The course was designed to allow for penalty-free repetition to reach a high or best score. At the end of the study, participants were clustered according to their performance, where 13% fell into a group referred to as "idealists" (participants that finished an exercise with a perfect score). These participants spent more time on the exercise, exerted more effort in looking for additional materials, and showed more repeated attempts to achieve a better score. It is worth noting that no special affordance was given to incentivize perfect play. Second, Rose et al. (2016), in their study, measured the effects of gamification via an a/b test in an online quiz. Both treatment groups were afforded to repeat the task penalty-free, with one group receiving a gamified version. Here, the results showed a significantly higher number of participants in the gamified treatment group to repeat until reaching a perfect score, demonstrating the success of the gamified version in incentivizing perfect play. In the third manuscript, we found that perfect play was successfully incentivized through a three-star reward system implemented into the serious game Foldit (Gaston & Cooper, 2017). In the experiment, player behavior is compared regarding three different implementations of a perfectionism incentivizing game design element: NO-STAR (basic game without a three-star system), 3-STAR (star rewards according to their number of moves), and 3-STAR-R (star rewards according to their number of moves including a forced reset after a specific number of moves). Significant differences were observed in the amounts of extra moves (number of moves above the perfect

[20] Search terms: (perfectionis* OR idealism OR compul* OR obsess* OR "perfect score" OR personality OR "player type") AND ("design element" OR "game element" OR "design mechanic" OR "game mechanic" OR "design pattern" OR "game pattern" OR affordance OR "grading system" OR "academic grading"). Search platforms: ACM, IEEE Xplore, JSTOR, ScienceDirect, Web of Science and Google Scholar.

value) and time per move between the NO-STAR and the 3-STAR version. In total, the three-star mechanism doubled the number of re-completed levels (Gaston & Cooper, 2017).

In summary, while not directly comparable to our setup, particularly the study by Gaston and Cooper (2017) shows that it is possible to nudge player behavior towards more perfectionistic outcomes. We find that while there is profound literature on perfectionism in psychology and sociology, only very few studies look at this phenomenon through the lens of game design, warranting further research as provided in the following study.

5.5.2. Theory and Hypothesis Development

To gain insights into expectable in-game behavior regarding perfect play, we base our theoretical foundations on the motivational theory by Ryan and Deci (2000). Motivation is the foundation of action: "To be motivated means to be moved to do something." (Ryan and Deci 2000, p.1). But not all motivation arises from a person's own intrinsic willingness to act on something. Oftentimes, particularly in education, students need to perform actions that are not inherently interesting or enjoyable; thus, it becomes essential for educators to know how to promote more active and volitional forms of extrinsic motivation to achieve successful teaching. In their research, Deci and Ryan differentiate between the level of motivation (in terms of quantity) and orientation of motivation (qualitative differentiation of types). In their taxonomy of human motivation Deci and Ryan map these types on a continuous scale from motivational incentives located solely outside of a person's motivational locus up towards a purely internal locus born from interest/enjoyment and inherent satisfaction. The different types of motivation are influenced by the underlying attitudes and goals that drive the individual as well as external factors.

During the early development of their self-determination theory, Deci conducted studies in which they found that monetary payments induced a change in the perceived locus of causality from internal to external, resulting in decreased intrinsic motivation for the activity (Edward L Deci, 1971, 1972). Such detrimental factors typically come from an outside source – either through the offer of an alternative, extrinsic reward which overwrites the former intrinsic impulse (if existent) or through direct negative input (Reiss & Sushinsky, 1975; Ross, 1975). In later studies, however, Deci et al. find that over time and through the application of self-determination (by strengthening autonomy, control, and relatedness), motivation can move from extrinsic towards intrinsic motivations, highlighting the potential for positive as well as detrimental affection of motivation (Deci et al., 1999).

We designed the perfect reward element in a way where it acknowledges perfection (through a pleasurable visual and auditive animation) but does not offer additional value towards other facets of the game (no in-game money or assets). As such, it could be argued that the element is less of a reward and more like a stimulating acknowledgment of the game that the state of perfection has been reached. When looking at different definitions of reward, we find that both arguments are valid. While the Cambridge Dictionary definition seems to revolve more strongly around the concept of a thing given in exchange ("something given in exchange for good behavior or good work, etc.", "to give someone a reward," "an advantage, for example, more money or a better job, that someone receives if they are successful, work hard, etc.") (Cambridge University Press, 2022), the definition by Merriam-Webster includes the concept of recognition ("something that is given in return for good or evil done or received or that is offered or given for some service or attainment") (Merriam-Webster Incorporated, 2022a). Given the softer nature in terms of extrinsic compensation, the perfect element has the potential to nudge players towards stronger intrinsic motivation but not lessen any exciting motivation as it does not replace intrinsic satisfaction through extrinsic rewards.

In the game, players are afforded full control and autonomy on their in-game behavior choices, which, according to Deci and Ryan's findings that self-determination serves as a beneficial driver towards intrinsic motivation, should compensate for potentially detrimental effects of the external locus of the reward. Further speaking for the game as an environment that nurtures intrinsic motivation is the fact the overall game has been perceived as enjoyable (as according to the very positive reviews on the stores, it is uploaded on: 4.4/5.0 on Android (n=68), 5.0/5.0 (n=4), as of January 2022)[21]. As enjoyment is another essential factor of intrinsic motivation, we are building in-game behavior on a baseline of intrinsic motivation to try and reach high scores. Finally, the studies we found on the effect of gameful design on perfectionism-oriented behavior showed significant effects in their design of nudging players to higher performances (Gaston & Cooper, 2017; Lisitsyna et al., 2015; Rose et al., 2016). Thus, we hypothesize that

H1: *Players that are rewarded for reaching a perfect score through acknowledgment achieve a higher performance overall compared to players that are not rewarded for reaching a perfect score.*

With regards to our second research question on the potential psychological foundations of the in-game behaviors we discovered in the field study, we first directed our interests toward models that could be related to perfectionism. Given the contrasting nature of the behaviors, we found we also looked into studies and theories based on dual models. Finally, we considered theories of decision-making strategies to find indications of relatability to the in-game choices players make. For this, we conducted a second literature study on suitable theories and related questionnaires for psychological measurement.

Need for Closure, Need for Structure, and Consistency Orientation

We started our research by looking for constructs that might fit this compulsion to complete a level or wave with a perfect score. The first need theory we took into consideration for this is the Need for Closure (NFC) by Kruglanski (1990), where they describe the item as a desire for a definite response on some topic in contrast to confusion and ambiguity. Building on this, Webster and Kruglanski (1994) developed a scale that measures the degree to which a person has a desire for certainty. High scores on this scale correlated with a preference for order, a dislike for ambiguity, making decisions and forming impressions quickly, and having strong opinions. While the construct shows an interesting premise, it has been criticized by Neuberg et al. (1997) for lack of construct validity in that it is treated as a one-dimensional construct while being de facto multidimensional. They further highlighted a redundancy with the already established Personal Need for Structure (PNFS) by Thompson et al. (1989, 1994), a scale that assesses preferences for structure and clarity with ambiguity and grey areas being perceived as troublesome and annoying. While both scales seemed to look at possibly related constructs, they did not exactly seem to describe the need for perfect completion that we were looking for, as the scores within the game are always unambiguous – independent of full, high, or low scores. The third need we evaluated for inclusion was the preference for consistency by Cialdini et al. (1995). It is a threefold construct as it consists of i) the need to align personal attitudes with one's behavior

[21] These are comparatively high values, given an average of 4.53 star ratings for an average iOS app and an average 4.05 star ratings for an average android app according to a benchmark study conducted in 2021 (Finixio Ltd, 2021)

(internal consistency), ii) consistent appearance towards others (public consistency), and iii) the desire for consistency of important related people (other's consistency). Like the other two, this score seemed to be related to the construct we were looking for but did not fit our case exactly as it dominantly related to social facets of consistent behavior more than the internal need to solve a problem as perfectly as possible.

Unfortunately, we did not find a construct for measuring a need for perfection. After considerations with regard to experimental duration, we decided to discard these singular constructs for our experimental design and instead focus on broader theories that could explain the dichotomous behavior structure we found in our player base. For this, we looked at different dual-processing theories.

Cognitive-Experiential Self-Theory

The Cognitive-Experiential Self-Theory (CEST) by Epstein (1994) is a dual-process model of perception and information processing based on three assumptions. The first assumption (based on the general dual process theory) is that social judgment and behavior result from an interplay of two interacting but independent systems of which one operates automatically (experiential) and the other willingly (rational) (Gawronski & Creighton, 2013). The second assumption is that the experiential system is driven by emotions. The third assumption is that (in contrast to other dual-processing theories) the following four basic needs are equally important in the interplay of the system: i) the pleasure principle (maximize pleasure, minimize pain), ii) the need for relatedness, iii) the need for stability and coherence of one's conceptual system and iv) the need for self-esteem (Epstein, 2003). The primary mode of processing happens through an adaptive cognitive system that is characterized by subconscious, rapid and holistic processing. Emotions and subconscious feelings/tendencies are connected to associationistic thinking, broad generalization, and categorization. The experiential system is slower to change as it does not relate to logic but to the outcome of repetitive and intense experiences (Epstein, 2003). The secondary mode of processing is an inferential system that is experienced actively and relates to logic rather than emotions as it operates on cause-and-effect connections. It processes thoughts slower but more extensively but is quicker to change as its focus is on the process rather than the outcome (Epstein, 2003). The two systems interact with each other and either can influence the other positively or negatively. A 40-item Rational-Experiential Inventory (REI) was developed by Epstein et al. (1996) to test the dominance of one of the two systems in a person. Therein, the rational scale is positively related to intellectual performance, self-esteem, and conscientiousness and negatively related to anxiety and naïve optimism (Epstein et al., 1996), while the experiential system is related stronger positively to extroversion, emotionality, and creativity and negatively to distrust and intolerance (Epstein, 2003).

Preference for Intuition and Deliberation

Another dual processing theory we evaluated for inclusion relates to individual strategy preferences and decisional fit by Betsch (2004). This theory suggests that there are two independent preferences in human decision-making that are stable over time: intuition and deliberation. Intuition describes the immediate feeling to make a certain decision and is, therefore, an affective mode. Deliberation is a decision mode that requires cognitive efforts of analysis and evaluation. It is measured by a self-report 16-item inventory (Preference for Intuition and Deliberation (PID)). Preference for intuition and deliberation is measured with the PID inventory by Betsch (2004). It has two subscales of intuition (PID-I) and deliberation (PID-D) with eight items each and only one recoded item in total. Its validity has been established with studies of more than 2500 people. All items were measured on a scale from 1: "very much disagree" to 5: "very much agree."

While building upon the same principles, this theory adjusts and extends certain aspects of CEST, specifically in terms of intention. Betsch (2004) states that their theory PID-inventory is based on motivation and preference, whereas CEST focuses on facets of ability and enjoyment (Betsch & Kunz, 2008). A further criticism of the REI given by Betsch is that the experiential scale of the REI confounds the two concepts of heuristic processing and intuition (Epstein, 2003; Keller et al., 2000). They argue that PID is an advancement to CEST in that it includes cognitive and behavioral dimensions and specifically isolates intuitive decision-making from heuristic processing. Despite the conceptual differences, faith in intuition (FI: experiential subscale in REI) correlates strongly with intuitive decision making (.67, $p<.001$), and the need for cognition (NFC: rational subscale in REI) correlates moderately (.20, $p<.001$) with the deliberation subscale of PID (Betsch, 2004). While REI correlates to logical thinking ability, the PID scale shows no relation to it on either subscale. Furthermore, the rational system of REI does not correlate with perfectionism, while deliberation of PID correlates to perfectionism and conscientiousness. Both scales, REI and PID, are widely used for measuring intuition and deliberation (Mikuskova et al., 2015), but given the higher focus on intention and the inclusion of perfectionism, we chose to include the PID and not the REI scale in our experimental design.

Taking into consideration that Perfectionists seem to be willing to spend extra time to achieve a perfect score given only a minor reward (seeing a “PERFECT” stamp and short celebratory sequence without further influence on the rest of the game), we associate the Perfectionist playing behavior with the construct of intuition. The willingness of Perfectionists to compulsively repeat for their well-being indicates an affective reaction to the game design element. Subsequently, the Rusher playing behavior is associated with the construct of deliberation due to its underlying utilitarian motive and rational efficiency. We thus hypothesize that:

H2: *Players with high intuition score high on perfectionism in the game.*

H3: *Players with high rationality score high on rushing in the game.*

Decision Strategies (Maximising, Satisficing, and Minimising)

As the core difference between the two player groups we found relates to the choice of repeating or not repeating for a higher overall score, we approached the domain of behavioral economics for concepts of outcome-oriented decision strategies. We found an interesting lead through Simon’s (1955) theory of bounded rationality that introduced the concept of maximization as a more realistic approach to searching and decision making than suggested by rational choice theory. Its core assumption is that individuals differ in terms of their personal decision goals. The theory discriminates between two overall behavioral profiles: i) *maximizers* who are eager to find the optimal decision by evaluating as many alternatives as possible and ii) *satisficers* that have a personal threshold above which they are content with their decision. While different scales have been developed to measure the tendency for maximization and satisficing, the Maximisation Scale (MS) by Schwartz et al. (2002) emerged as the most prominent (Schwartz, 2016; Schwartz et al., 2002). The different scales vary in terms of definitions and conceptualizations for maximizing. Some studies show maximizers to be more present-focused, neurotic, more likely to divagate from factual thinking (Besharat et al. 2014), feel more distressed about their decisions (Dahling and Thompson 2013), and at last, be more perfectionistic than satisficers (Schwartz et al. 2002), while others have found maximizers not be dissatisfied with their lives (Highhouse et al., 2008; Purvis et al., 2011).

In their study and related inventory (Decision Making Tendency Inventory (DMTI)), Misuraca et al. (2015) expand on the concept of maximizing and satisficing. They add to the theory by introducing two subcategories of maximizing as well as satisficing. They further include the concept of minimizing, which is also divided into two subcategories adding up to a total of six sub-categories. The following personality types are discriminated: *Resolute maximizers* have high standards and seek alternatives. Their behavior is defined by conscientiousness, perseverance, and scrupulousness. Misuraca et al. (2015) find that resolute maximizers are more goal-oriented and more tenacious. *Fearful maximizers* engage deeper in their search for alternatives and experience greater decision difficulties than resolute maximizers, although their behavior correlates less strongly to high standards. In contrast to resolute maximizers, fearful maximizers do not follow clear goals, and their fear of failing can further weaken their efforts. *Less ambitious satisficers* do not follow a clear plan and have lower perseverance, while more *ambitious satisficers* have higher standards. Minimizing is referred to as using the minimum amount of resources to gain the minimally acceptable outcome (Schwartz, 2016). *Indolent minimizers* make fast decisions and act according to this principle. *Parsimonious minimizers,* on the other hand, aim at spending the least money and evaluate options by that single criterion.

These constructs are tested through the Decision Making Tendency Inventory (DMTI) which consists of 29 items, of which seven of the eleven items measuring maximization are taken from Schwartz et al. (2002) and one from Highhouse et al. (2008). Misuraca et al. (2015) developed a new satisficing scale to increase construct validity, while the category of less ambitious satisficing has been shown to mostly match the previous studies. The inventory is not domain-specific, showing good psychometric properties with items that range from consumer-over professional- to academic backgrounds. The alpha-values of the six subcategories range from .6 to .81. Many different search and decision strategies challenge the idealistic rational choice theory (satisficing, maximizing, and minimizing being one subset of such concepts), describing different algorithms of how decisions are made under time and knowledge constraints (for an overview, see Gigerenzer and Selten, 2002, and an in-depth study see Pfeiffer et al. (2014)). As the repeat and reward-setting is a binary choice situation with (yet) unknown attributes, the satisficing, maximizing, and minimizing heuristic appeared to be the most applicable based on the behavioral findings of the pre-study, which is why we included the DMTI in the experimental design.

Due to the subcategories introduced by the theory of Misuraca et al. (2015), we can deduct two motifs for perfectionistic playing behavior: performance-orientation (neurotic maximizer) and mastery-orientation (goal-oriented maximizer). While the first construct is closely linked to neuroticism, the second construct is related to personality traits of resolution and conscientiousness. Underlying both motifs is the overall pattern of a high willingness to invest extra resources for the best outcome (Hewitt & Flett, 1991; Misuraca et al., 2015; Purvis et al., 2011). We thus see a strong connection between the measure of maximization and perfectionism. On the other hand, the measure of minimization is defined by the participants' unwillingness to invest more than an absolute minimum of resources to reach their goals which fits the behavior pattern shown by the Rushers. We thus hypothesize that:

H4: *Players that have a tendency for maximizing score high on perfectionism in the game.*

H5: *Players that have a tendency for minimizing score high on rushing in the game.*

Big 5 – OCEAN model

As we aim to understand the underlying personality-related motifs for the two playing behaviors, we wanted to include a standard personality model to test a general correlation between personality and playing behavior. We chose the Big 5 model or OCEAN-Model as it is one of the most widely used personality models (Zillig et al., 2002). The model measures five basic personality dimensions that were originally obtained by factor analysis. Its dimensions are *conscientiousness* (e. g. reliability, discipline), *agreeableness* (e.g., altruism, empathy, trust), *neuroticism* (e.g., nervousness, fear, irrational behavior), *openness to experience* (e.g., curiosity, creativity, independent judgment), and *extraversion* (e.g., sociality, activity, optimism) (Goldberg, 1990). The model is typically tested through the BFI-44 scale (John et al., 1991). Due to the time constraints of the experiment, we looked at shortened versions, settling on the self-evaluation BFI-10 by Rammstedt and John (2007), as according to the authors, the scale covers 70% of the variance of the full version. The scale introduces two representative questions for each dimension, with one always being reverse coded. Each sentence starts with "I see myself as someone who..." and is followed by two items describing typical characteristics of each of the five dimensions (e. g. "...tends to be lazy"). The items are measured with five-point scales ranging from 1: "disagree strongly" to 5: "agree strongly".

Psychological studies have linked perfectionistic behavior to conscientiousness and neuroticism, with neuroticism as the dominant trait (Betsch, 2004; Frost et al., 1990; Hewitt & Flett, 1991). We thus expect players that fall into the Perfectionist behavior to score high in these dimensions. Respectively, we expect Rushers to score negatively on conscientiousness. This would match the results of Misuraca et al. (2015), who found their indolent minimizer to correlate strongly negatively to conscientiousness. We hypothesize that:

H6: *Players that are highly conscientious and neurotic score high on perfectionism in the game.*

H7: *Players that are low in conscientiousness score high on rushing in the game.*

Including all personality-related hypotheses, the overall framework consists of seven hypotheses relating to the relationship between game design elements and performance mediated by matters of personality and decision strategies (for an overview of our research model, see Figure 57).

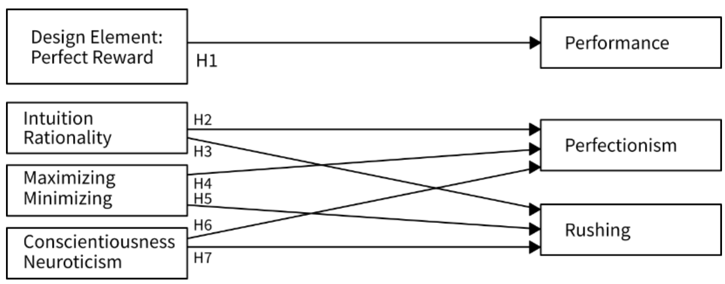


Figure 57 – Research Model Experimental Study 1

5.5.3. Experiment

Experimental Design

We designed the laboratory experiment to test the effect of the perfect reward element on performance as well as the effect of three scales for measuring personality-related constructs on emerging in-game playing behaviors. We designed a between-subject experiment in two stages, where the perfect reward element was manipulated. Depending on their group, participants received a link with a different version of the game. While the treatment group was shown the perfect reward for perfect wave completion, the control group received a neutral text to inform them they had passed the wave. The app was locked by a server gate until the experiment started. To be able to match the experimenters' in-game data with the respective surveys in the lab, a unique, individual code was displayed by the app, which participants had to enter into the survey at the beginning of the experiment. The name of the app and all potential identifiers were anonymized. On location, we provided backup mobile phones for participants that failed to follow the instruction for downloading and installing the app beforehand (this happened with one person). The study was conducted with LimeSurvey (LimeSurvey GmbH, 2003), and the complete study was held in English. As it was held in a German university (Karlsruhe Institute of Technology, 2007), we set fluent knowledge of the English language as an exclusory requirement for participation. An overview of the Operationalization of Control and Additional Variables can be found in Table 51 in Appendix C.1.1.

For the experiment, we used an abbreviated version of the game. We excluded all elements outside the core gameplay and only kept the sorting game, the repetition element, the look-up element, and the perfect reward. The additional game design elements of the core gameplay (repetition and look-up) were kept to ensure comparability to the findings of our field test. The original level system of the core gameplay as it is used in the field study was cut to 4 tutorial waves and ten mandatory waves consisting of 15 incoming waste items each. These values were chosen due to the time constraints of the experiment and for consistent measurement. We included neutral, informative screens at the beginning and end of the game as additional instructions for the experiment. We also added an option for players to continue playing without disclosing any information on how long that would be possible after the mandatory ten waves participants were told that the official part of the experiment was done but that they could continue playing if they wanted to. After seven more waves, the app would tell players that the practical part of the experiment was now fully finished and to close the app. We included this optional part to explore how strongly participants would want to continue outside of the mandatory setting. The user behavior data was gathered using a centralized logging server provided by Unity Analytics (Unity Technologies, 2014). We controlled for the demographic factors of age and gender at the beginning of the experiment and the construct of enjoyment of the game based on the model by Koufaris (2002).

Experimental Procedure – Pretests

We conducted two pretests to evaluate the chosen inventories and the overall experimental design, improve the planned experiment process and gain feedback on the tested gameplay elements. Both pre-tests were conducted with five student participants at the KD^2 Lab (Karlsruhe Institute of Technology, 2007). The pretests were conducted in three parts and lasted 45 minutes in total. Participants were incentivized with a flat fee of 10€. Each participant was placed in an isolated cabin with a computer and their own tablet or cell phone for the experimental task. After the instructions and confidentiality information was read to the participants, they were to start with the survey. In the first part, participants were asked to state their age and gender at the beginning of the survey. They then were asked to work through the prepared inventories (REI, PID, DMTI, and BFI-10). Next, participants were asked to play a minimum of eleven waves. Afterward, participants were notified that they could play until they felt like quitting. Once all participants indicated that they were finished with the experiment, we had a protocolled discussion with all participants in an open session. There we asked the participants about their perception of the implemented design decisions, what they liked and disliked about the game, if and why they tried to sort perfectly, how they perceived the reward at the end of each round and how they could be incentivized to opt for higher scores. After evaluating the participants' feedback on the first pretest, we adjusted matters of wording and user experience within the survey. As four of the five participants had criticized the overall number of questions, especially in regard to certain redundancy between the dual processing inventories (REI and PID), we chose to curtail the experiment. Contrary to the REI, the deliberation scale of PID has been shown to correlate to a higher need for structure and a higher tendency for decision outcome maximization and perfectionism (Betsch & Kunz, 2008). We thus decided to omit the REI from the experimental setup and focus on the PID inventory in terms of the inclusion of the dual processing theory. The number of inventory items used in the second pretest was reported as adequate. Further, while there were no criticisms of the game design decisions, some more wording adjustments were made to the final questionnaire. With those adjustments, the overall test design showed a good fit for actual testing and was reported as adequate in length by the test participants.

Experimental Procedure

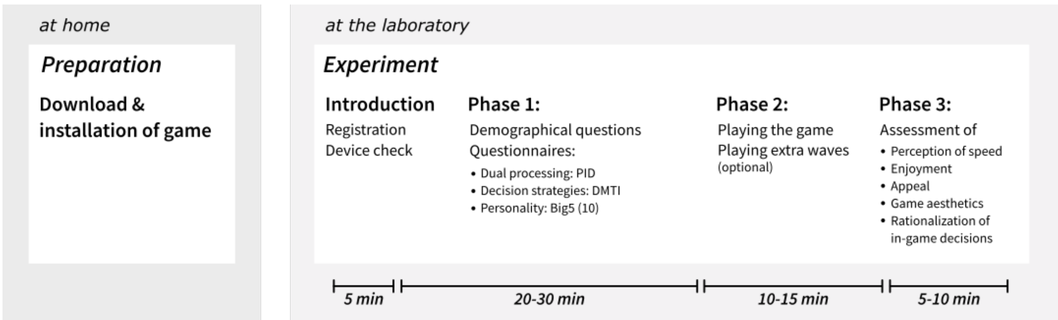


Figure 58 – Experimental Procedure Experimental Study 1

We recruited participants through the organizing and recruiting software hroot (Bock et al., 2014) provided by the c in which the experiment was conducted. Participants were offered a flat fee of 10€ to prevent any changes in behavior through extrinsic incentives. After randomly selecting applicants, we used a script to randomly assign them to the treatment- and control groups. One week before the start of the experiment, we sent an email to both groups, asking them to download and install the experimental artifact on their private mobile phones, linking to different versions of the game, respectively. They were assigned to one of two sessions, the first of which hosted ten and the second twenty participants.

On location, we started the experiment with a short introductory phase, where we registered the participants and ensured that the app was running correctly and that notifications were turned off to prevent interruptions during the practical part of the experiment. Next, participants were led to their single cabins, where they sat in front of a computer screen that displayed the introductory screen of the survey. Once everyone was seated, and all cabins were closed, participants were asked to start the experiment by clicking on the start button. In the first phase, participants had to answer demographical questions on their age and gender, followed by the three questionnaires on personality and decision strategies (PID, BFI-10, and DMT) on the screen in front of them. For the second phase, they were instructed to take up their phones, open the app and play the game until they received further instructions. The game started with a four-wave in-game tutorial, followed by eleven mandatory waves. Once finished, an informational text in the game informed them that the obligatory part was over and that they could either continue with the survey right away or continue playing a few more waves before finishing the experiment. For those who continued to play to the end, a final screen instructed them to now proceed with the experiment by finishing the survey. The final phase happened back in front of the screen, where participants were asked to answer some questions on their perception of speed, enjoyment, the general appeal of the task, and the game aesthetics, as well as elaborate on certain in-game decisions: ("if you chose to repeat, please elaborate, why"; "If you played extra-waves, please elaborate, why"). Lastly, they were given the option to give general feedback through a free text. For an overview, see Figure 58. The full Operationalization of Control and Additional Variables can be found in Appendix C.1.1.

Operationalization of the Dependent Variables

As we are interested in players' motivation to aim for perfection, we measured performance by looking at the final performance players reached before choosing to move to the next wave. As the final seven waves were optional, performance is only measured for the ten mandatory waves. Furthermore, based on the metric we used to identify the Perfectionist and the Rusher behavior, we measured perfectionism by the number of perfect waves players reached during a playthrough and rushing by the number of waves players chose not to repeat. An overview of the treatments' structure for this experiment is provided in Table 15.

Table 15 – Dependent Variables for Measuring Performance and Playing Behaviors

Dependent Variable	**Range and Meaning**
Performance	Continuous value between 0 and 1: The percentage of correctly sorted waste items of the last repetition per wave over the ten mandatory waves.
Perfectionism	Continuous value between 0 and 1: The number of perfect waves divided by 10.
Rushing	Continuous value between 0 and 1: The number of non-repeated waves divided by 10

5.5.4. Results

The experiment was completed with thirty participants. However, one data set had to be excluded because of technical problems, leaving 29 datasets for the final evaluation. The mean age was 22.61 (min:18, max:28[22]). In terms of gender distribution, 86,2% of participants identified as male, and 13,8% of participants identified as female, no person identified as "other."

Table 16 – Descriptive Statistics of Dependent Measures Experimental Study 1

		In-Game Performance	**Perfectionism**	**Rushing**	**Perfectionists**	**Rushers**
	n	mean (std. dev.)	mean (std. dev.)	mean (std. dev.)	n (% of all players)	n (% of all players)
Treatment (perfect reward)	14	13.99 (.898)	.571 (.292)	.750 (.277)	3 (21,4%)	8 (57.1%)
Control (neutral text)	15	14.24 (.571)	.620 (.286)	.729 (.204)	5 (33.3%)	5 (33.3%)
Total	29	14.12 (.744)	.597 (.285)	.735 (.238)	8 (27.6%)	13 (44.8%)

[22] One person reported their code instead of their age. As this was the only missing data-point in the otherwise complete and coherent dataset we decided to include this person's data, given that the age-range was not expected to influence our primary research interest.

As can be seen in Table 16, both behaviors we identified in the field study emerged in the experiment as well. The overall performance was very high, with participants missorting on average 1 item per wave. No participant ever encountered a game-over-state (50% or more items missorted). With regards to H1, an independent t-test comparing the last performance of each wave showed no influence of the reward design element; thus, it can't be supported (t (27) =.453, p=.666) (see Figure 59).

A Kolmogorov-Smirnov test showed, however, that the data was normally distributed. When looking at the aggregated last performance per wave, we see that, while not significant, the treatment group (A) performs consistently worse than the control group (B) (see Figure 60).

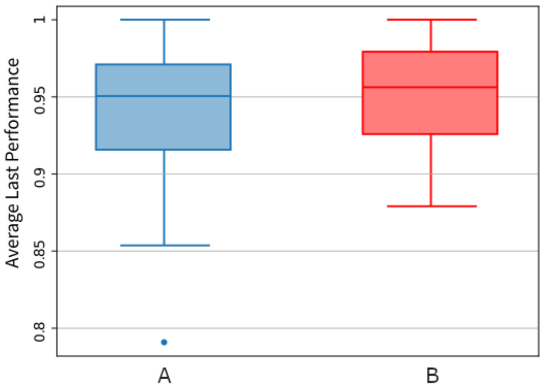


Figure 59 – Main Effect for the Treatment Group (A) and the Control Group (B)

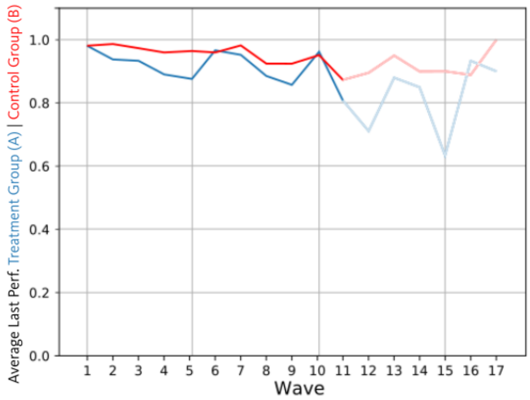


Figure 60 – Comparison of Last Performance per Wave for the Treatment Group (A) and the Control Group (B)

In terms of playing behaviors, we found eight Perfectionists (players that repeated every single wave until they reached perfect scores) and thirteen Rushers.

Table 17 - Regression Analysis Summary for Personality Traits Predicting Perfectionist Behavior

Variable	B	Std. Error	t	p
PID-D	-0.01	0.182	-0.056	0.956
PID-I	0.255	0.13	1.952	0.067
BiG1-E	-0.054	0.067	-0.797	0.436
BiG5-C	-0.024	0.072	-0.325	0.749
BiG5-N	-0.06	0.083	-0.72	0.481
BiG5-O	0.001	0.058	0.024	0.981
BiG5-A	-0.02	0.083	-0.243	0.811
DMT-Max	-0.108	0.115	-0.94	0.36
DMT-Sat	-0.06	0.188	-0.318	0.754
DMT-Min	-0.058	0.113	-0.512	0.615
(Constant)	1.03	0.921	1.118	0.278

Adjusted R^2=-.027

Table 18 - Regression Analysis Summary for Personality Traits Predicting Rushing Behavior

Variable	B	Std. Error	t	p
PID-D	0.178	0.172	1.035	0.314
PID-I	0.08	0.123	0.648	0.525
BiG1-E	0.008	0.063	0.124	0.902
BiG5-C	0.035	0.068	0.508	0.617
BiG5-N	-0.057	0.078	-0.734	0.472
BiG5-O	-0.059	0.055	-1.079	0.295
BiG5-A	-0.017	0.079	-0.217	0.83
DMT-Max	-0.072	0.109	-0.659	0.518
DMT-Sat	-0.162	0.178	-0.91	0.375
DMT-Min	0.104	0.107	0.977	0.341
(Constant)	0.586	0.87	0.673	0.509

Adjusted R^2=-.071

With regards to the personality-related hypotheses (H2-H6), we looked at the personality tests as single groups without including all the controls, as given the small sample size, the data collected in this experiment is underspecified. We conducted OLS Regressions for each personality test group with each behavior group (Perfectionist value and Rusher value); however, we did not find significant correlations (see Tables 17 and 18).

Looking at the usage of the design elements that incentivize performance enhancement (repeat option and look-up), the look-up element was used by almost every user at least once (n=26) with, on average, 22.19 lookups throughout the main ten waves of the experiment (mean: 22.19, min: 0, max: 65, std.dev.: 16.49). We found a significant positive correlation between the number of look-ups and last performance with a correlation coefficient of 0.56. The spearmen correlation was used for the regression analysis (see Figure 61). Each dot in the graph represents one participant.

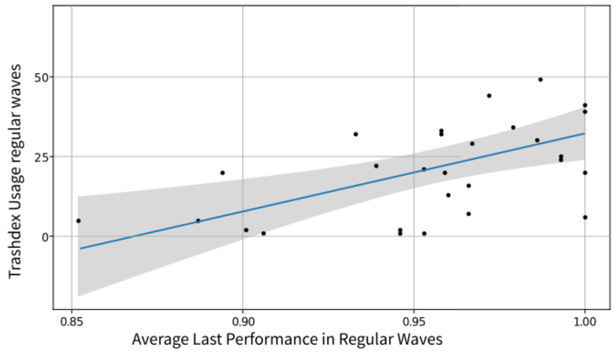


Figure 61 – Correlation: Look-Up Use and Performance

The repeat element was used by fewer participants (n=23). Participants repeated, on average, 3.62 waves over the main ten waves of the experiment (mean: 3.62, min=0, max=15, std.dev.: 3.86). No significant correlation was found between the use of the repetition element and the last performance.

Overall, 18 participants continued to play at least one extra wave after the formal experiment, with four participants continuing to play all seven extra waves. On average, the number of played extra waves was 2.24 (min=0, max=7, std.dev.: 6.50).

On average, it took participants 11.8 minutes (min: 8.1 and max: 17.5) to complete the mandatory playing task. In total, it took participants, on average, 43 min to complete the whole study. On average, it took players 52.76 seconds to play a wave (min: 33.00 s, max:125.00 s, std.dev.:6.54). During the optional waves, it took the remaining players on average 56.42 seconds (min: 33.00 s, max: 272.00 s (4,53 min), std.dev.: 13.14).

Explorative Research

As we did not find an effect in our main hypothesis (H1) and few correlations between the psychological tests we conducted and the in-game behaviors we had identified, we decided to conduct additional, explorative research on the data to understand our players' motifs on a qualitative level. While the main behaviors we were looking for emerged from an interplay of the core gameplay and the option to repeat, the game features an additional element that affords players to look up items they don't know where to sort. We were interested in how players used the look-up element and how it influenced their outcome. In this explorative analysis, we found an interesting pattern with regards to the combined usage of the repetition and the look-up element. In three players' plots, we identified a behavior where the participant would first play a wave without using the look-up element, then repeat and, in the following wave, look up every item they had sorted incorrectly beforehand (see Figure 62, one bar means no repetition, a second bar means that the wave was repeated, blue indicates that the look-up element was used). The outcome of this behavior always resulted in an increase in performance compared to the first wave.

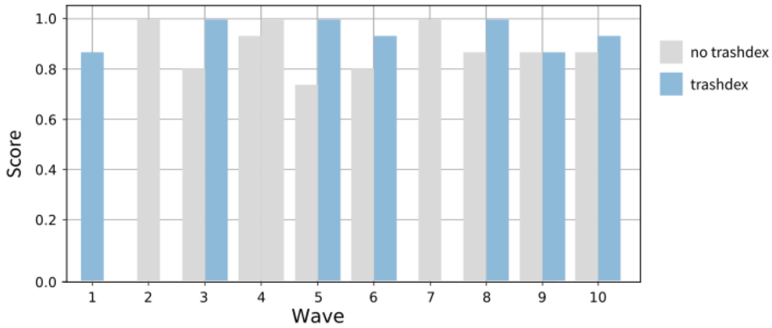


Figure 62 – Example Plot for an 'I'll Try It by Myself First' - Behavior

Analyzing the rationalizations participants stated for repeating or not repeating a wave, three (10,35%) participants stated that they always repeated until they had a perfect score, and seven (24,14%) stated that they repeated if they had sorted more than 1 item incorrectly, four (13,79%) stated that they repeated if they had sorted more than two items incorrectly, and seven (24%) never repeated their wave irrespective of their score. The other statements were less clear in terms of the algorithm of choice (examples include: "if I made a stupid mistake," "if I felt cheated by the game because it treacherously depicts things," and "If I made too many mistakes, I repeated"). These statements were in accordance with their in-game behavior.

Analyzing a potential relationship between our playing behaviors (Perfectionists and Rushers), we did not find a linear relationship between the behaviors and the number of additional waves played nor the other two learning-enhancing design elements (repeat and look-up). When looking at the performance distribution in terms of perfectly completed waves, we found a non-linear performance pattern with significant drops in performance at two points during the task (see Figure 63).

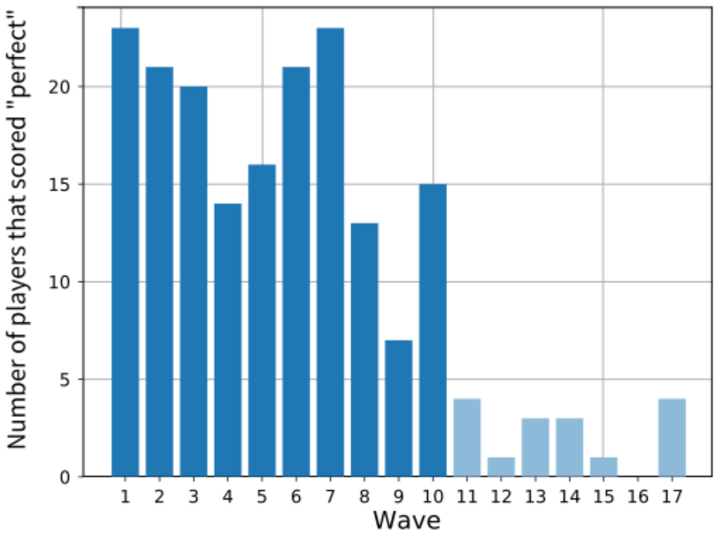


Figure 63 – Number of Players that Scored Perfectly per Wave

We also asked for rationalizations on why they continued to play if they did so. Of the given reasons, "curiosity" was the most stated or implied reason for continuation with six participants, and "ambition" and "to reach a score of over 10.000" were the two second-most stated reasons with two participants each. Other reasons included "for experimentation," "fun," "accident," and "hope that it might raise the experimental fee." Of the reasons not to continue, "no reason," "not enough fun," and "accidentally quit" were the three most stated reasons by three participants each; other reasons included "not enough fun," "too repetitive," and "all energy was used up for repeating waves." In terms of general qualitative evaluation of the design, 19 of the 29 users were either satisfied or strongly satisfied with the visual design of the game (see Figure 64). The feedback on the monsters was perceived as important or very important by 15 of the users (see Figure 65).

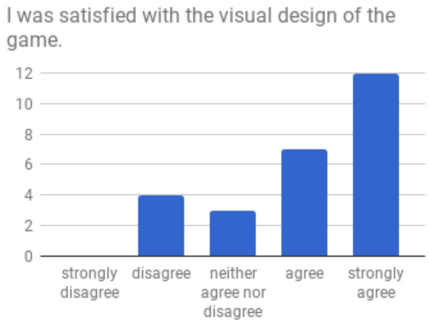


Figure 64 – Satisfaction with Visual Design

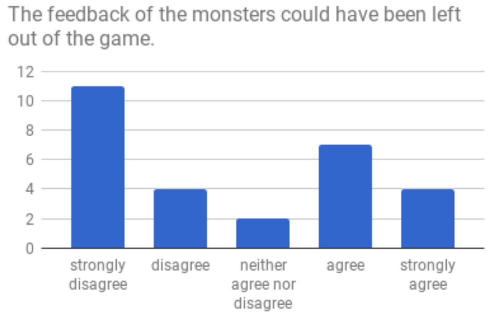


Figure 65– Importance of Monster Feedback

5.5.5. Discussion & Conclusion

The main goal of this experiment was to gain insights into the effectiveness of our game design decisions in incentivizing perfect performance and completion, particularly regarding the reward for perfect play. During a second analysis of the field study,[23] we observed an effect where once players managed a perfect wave for the first time, their performance in the next waves improved significantly afterward (a two-tailed t-test of the aggregated waves previous to the first perfect wave and the waves afterward showed a p-Value of 1.573e-10 (<0.05)). However, when we tested this in terms of the perfect reward design element, the effect could not be reproduced. Our analysis found that the perfect reward element as tested in our experimental setup produced no significant effect compared to the control group.

Generally, the overall low number of participants contributes to the weak outcome values the experiment produced. However, as the two groups were so homogenous in their average last performance, it is likely that a higher number of participants would have produced a similar outcome. Furthermore, while the experimental outcome seems to suggest that the perfect reward element was too weak as a stimulus to positively affect behavioral change (as would a successful nudge), it also did not have a significantly detrimental effect: we found the same ratio of players showing perfection-oriented playing behavior within the experiment as in the field study. Given the volatile nature of motivation, while not a success, it is also not a failure in terms of the design of the game element.

In terms of in-game behavior, both formerly isolated behaviors reemerged in the experiment, with an almost quadrupled amount of Perfectionist behavior (Perfectionists in the experiment: 27.6%, Perfectionists

[23] At the point of this analysis, the number of downloads had accumulated to 7029 of which we could retrieve 4529 unique and uncorrupted datasets (through Unity Analytics)), 1176 of which proceeded past the tutorial waves.

in the field study: 7%) and more than double the amount of Rushing behavior compared to the field study (Rushers in the experiment: 44.8%, Rushers in the field study: 19%). It seems that the strong increase in these numbers is related to the nature of the experiment and its participants. First, an experimental setting can influence behavior towards socially desirable actions, thus incentivizing more conscientious play. On the other hand, as the experiment was rewarded with a flat fee, participants could be incentivized to maximize their outcome by rushing the process. Regarding the participant selection (students recruited via the university lab list), their behavior will likely generally differ from the field study, given that the games' target audience is children. As such, the field study is likely to be interspersed with many irrational and spontaneous playing decisions. In the personality test on deliberation and intuition, the distribution for the deliberation construct is skewed to the high end of the scale, indicating that according to their evaluation, most participants see themselves as rational individuals (see Figures 94 and 95 in Appendix C.1.2). We found further evidence of the rationality of participants' in-game playing choices in the evaluation of the given reasons for playing behaviors where most participants stated a clear cut-off point for satisfaction in terms of performance (missorting not more than zero, one, or two items per wave). Several participants explicitly mentioned extrinsic or reward-based factors as components that drove them to aim for higher scores or play longer (reach a score over 10.000, hoping for more pay at the end of the experiment). This indicates that a rational player base could be incentivized by such extrinsic rewards to play longer and achieve higher performances.

On the other hand, looking at the significant relationship between Perfectionist behavior and the preference for intuition, it seems that aiming for perfect scores is less based on rational choice and more on the intrinsic compulsion to do so. This is a finding corroborated by literature, where perfectionism is often linked to neurotic and compulsive tendencies (Hamachek, 1978). Another facet of perfection was further brought to our attention through the free-text commentary of the experiment. Several participants commented that they would have liked to see a completion bar or any indication of overall progress. This sets some light on the general role of completionism, which was also touched upon by Gaston and Cooper (2017) in their 3 Star progress reward. Our results do not indicate that the two constructs, completionism, and perfectionism, are identical (Perfectionists were not necessarily inclined to complete all waves in contrast to Rushers, that explicitly forfeited any option to raising their score but finished the game, respectively, the experimental task up to the last wave). However, testing for the need for completion would be an important measure for a future study. Given the consistent emergence of this playing group while not being linked to its respective design element, we suggest that the other components that make up the overall design pattern (letting players know that they have not reached 100% and allowing them to further pursue that state without penalty) should be looked at for further understanding of how Perfectionist behavior is incentivized.

In terms of the rushing playing behavior, we did not gain new insights from the results of our experiment. Approximately half of the players (independent of treatment group or control) fell into that behavior pattern. As the experiment was incentivized with a flat fee, it would be natural for participants to optimize their return by finishing the task as quickly as possible. It is further likely that those who proceeded with the additional waves did so mostly due to the circumstances of the experiment and factors of social desirability (see Podsakoff et al., 2003).

By using the player plots to conduct further explorative analyses with regards to the optional learning-enhancing design elements, we stumbled upon some additional intriguing behaviors regarding players' use of the game. Several players showed behavior patterns that indicated an additional, self-adhered set of rules that, while being afforded through the design elements of the game, were not explicitly incentivized through any experimental instruction. One such emerging pattern related to the usage of the look-up and repeat element, where the rule seemed to be that the participants first tried to sort everything without help and then, after gaining the unbiased feedback on their performance, looked up the items they didn't seem to know, achieving a perfect wave in the second run. For one participant of such behavior, we found an explicit explanation about their perception of the look-up element (Trashdex) as an "inofficial cheating device only to be accessed in need."

When looking at the performance distribution in terms of perfectly completed waves, we found a non-linear performance pattern with significant drops in performance at two points during the task (see Figure 63). The observed inconsistency could be an indicator of a motivational pattern (motivated at the beginning with a decline of motivation from there, followed by another motivational surge followed by an overall drop in motivation for perfect play). However, any reliable interpretation would warrant further data.

This study contributes to research by being one of a small sample of studies that analyze perfectionism in relation to specific design configurations – in our case, an acknowledgment reward. We further contribute by corroborating the findings of the field study, where we identify perfectionism and rushing as two consistently emerging playing behaviors. Particularly perfectionism distinguishes itself as a behavior pattern related to intuitive decision-making. This implies that designers of gameful artifacts should expect a certain amount of their player base to behave perfectionistic – independent of set incentives. If a perfect state can be reached, these players will try to reach it, independent of a final reward. Also, our study shows how, given the freedom to interact or not interact with certain design elements, players will come up with their complex, intrinsic rule-sets depending on their beliefs and values – as evidenced by the "try-it-by-myself-first" behavior pattern we identified in this experiment. Seeing the overall high performance outcomes (on average 14.2/15 correctly sorted items in the last wave per round) and the high thresholds players set for themselves (not more than zero, one, or two mistakes), we are satisfied with the game's ability to incentivize high-performance play.

5.6. Post-study – Evaluation of Learning Enhancing Game Design Elements in the Field Study

Building on the findings of the experiment – particularly regarding the additional behavior patterns we found in relationship to the optional, learning-enhancing design elements, we wanted to gain further insights regarding the primary goal of the game: its effect on learning outcome. For this, we conducted another analysis of data collected through the field study. We were particularly interested in the following parameters: ongoing engagement motivation (how far into the game did players manage to progress (counted by waves)), repetition (how much and to what effect was the repetition element used), look-up (how much and to what effect was the look-up element used) and learning effect (how often did players sort an item correctly after having sorted it incorrectly the first time).

5.6.1. Descriptives

As of April 2018, the field study managed to gather 8041 Downloads over three years (the study was launched in May 2015). Of the 4529 data points we managed to retrieve from unity analytics, we were able to use 1176 unique IDs for our additional analyses. In terms of invested playing time, players of the field study interacted with the game on average for 5.48 different days, the minimum was one day, and the maximum of individual days of interaction was 79 days.

In contrast to the high-performance output of the experiment where no game-over-state was reached, we found 229 players in the data of the field study that suffered one or more game-overs in their playing session in contrast to the experiment. While we do not have any insights into the personal data of the users of our field study, the game is explicitly tailored to be children-friendly. We thus assume that the high number of game overs, as well as some of the extreme outliers and more erratic behavior patterns of the field study, can be attributed to the younger part of our audience.

5.6.2. Learning effect

To gain first insights into the efficacy of our design elements in terms of increasing the retained knowledge (*learning effect 1*), we devised the following definition: the average difference between the performance value after finishing a wave the first time (first performance) and after the last repetition before moving on or quitting (last performance).

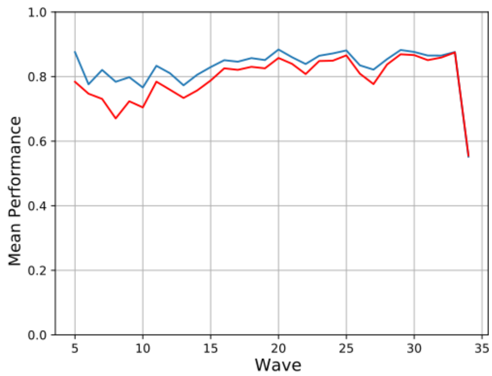


Figure 66 – Comparison Between First and Last Performance (Field Study)

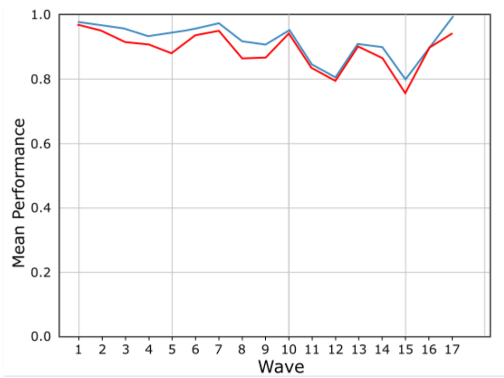


Figure 67 – Comparison Between First and Last Performance (Experiment)

Figures 66 and 67 show the comparative averages of the players' first (red) and last performances (blue) in the field study and the experiment. In the experiment, the average learning effect amounts to 1.3%, in contrast to the field study, where it was 11%. This indicates a higher learning effect in the field study in comparison to the experiment. However, as stated in Chapter 5.5.4 and as can be seen in Figures 66 and 67, the overall performance in the experiment is already very high in contrast to the field study. Further complicating the comparison is the fact that, in contrast to the experiment, the number of players in the field study is not constant but, in fact, decreases per game wave.

To gain a more objective value of learning, we devised another metric for measuring the learning effect (*learning effect 2*): the average sorting correctness of a waste item after an initially wrong assignment. To gain a consistent player base, we chose to analyze a player group that saw the first quarter of the game (n=780). At that point, this group of players had already seen and sorted from 25 up to 68 different waste items. The average performance of this group was measured at 73% (correctly sorted items/total items per wave), meaning that, on average, players in this group should be able to sort 50 different waste items correctly outside of the game. This group's in-game learning outcome was measured at 51% (average sorting correctness after an initially wrong assignment).

On the other hand, players that played to the end of the game (n=234) were exposed to the full number of 285 waste items. The average performance of this group was measured at 87% (correctly sorted items/total items per wave), indicating a learning benefit through the longer content exposure (14% difference to players that quit during the first quarter of the game).

Interestingly, a general tendency toward the overall increase in performance and an overall decrease in learning effect can be observed in the field study (the extreme drop at the end of the game is accounted for by waves 31-34 being unrepeatable final-boss waves with large amounts of items). In contrast, we observe the opposite tendency in the experiment (note: waves 11 to 17 were optional waves). This difference could be explained by the non-intrinsic nature of the experimental setup itself and the resulting drop in interest after the mandatory waves.

Concluding, we see indications that even in an erratic player-base with an inconsistent setting if played for an approximate duration of 30 minutes (~amounting to a quarter of the overall playing time of the game), the game manages to increase correct waste sorting knowledge by approximately 50% and increased values with increased playing time. This construct of learning-effect is, of course, an in-game learning measure and does not reflect on improvements in waste-related behavior outside of the game.

5.6.3. Ongoing Engagement Motivation

Despite a decline in player base per progressive wave, a total of 234 players reached the end of the game (5,1% of all accounted data sets). Given that the field study offered several game design elements that were implemented to encourage prolonged play (narrative elements) that were not prevalent in the experimental study, we were interested in the relationship of these elements to the willingness to prolonged play. When correlating the number of screen loads per player of all narrative-related screens with the number of waves played, we found a positive Spearman correlation of 0.52 (see Figure 68).

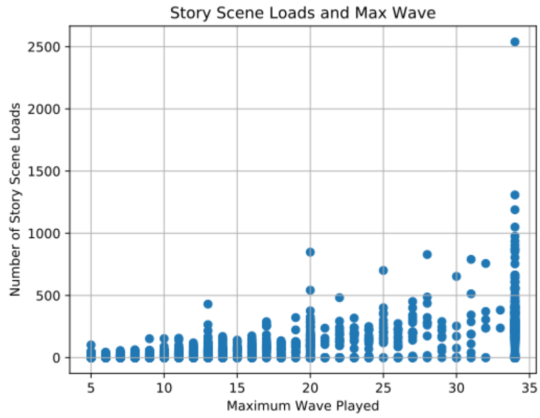


Figure 68 – Correlation: Story Scene Loads

This finding highlights the importance of these additional elements for the achievement of the *breadth of knowledge* as opposed to the elements implemented in the core gameplay that aim for *depth of knowledge* in terms of correct waste sorting.

5.6.4. Repetition and Look-Up

In the design of our core gameplay, we wanted to afford players to be able to learn as much as possible according to their playing styles. For this, we had implemented two additional, optional design elements, repeat, and look-up (see Chapter 5.3.4). No changes were made in terms of the design of these two items in the implementation of the experimental version. While both design elements afford and incentivize playing behavior contributing to the depth of learning, their usage might produce different learning outcomes, as one requires more effort to reach perfect play while the other reduces this effort. While they were not manipulated in the experiment (and thus, no causal relationships can be inferred), we gained some insights from the comparison of the descriptive results.

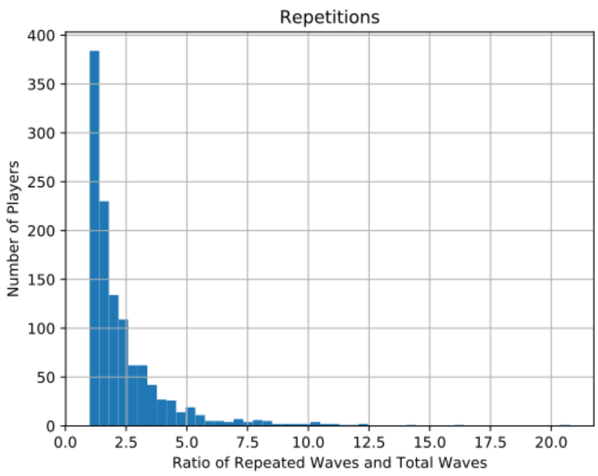


Figure 69 – Number of Players Related to Average Number of Repetitions per Total Waves Played

When comparing the experiment and field study in terms of overall repetition, the repeat ratio of the field study was 2.39 repetitions per wave (variance=1.90, min=0, max= 90), while the experiment produced an average repeat ratio of 3.62 repetitions per wave. As reported in Chapter 5.5.4, most players had a maximum threshold of 2 missorted items before they chose to repeat. For comparison, we measured the average repetition performance threshold in the field study by averaging the performances of the next to last and last repetition of each wave. Here we found an average performance threshold of 78%. However, as stated earlier, the data of the field study shows a larger ratio of potential outliers, as can be seen in Figure 69. While most players show an average repetition rate of between 0 and 2.5 repetitions, a small number of players seemed to be willing to repeat a wave between 5 to 20 times.

Overall, while the repetition values were higher in the experiment, given the lack of social desirability and accountability that can sometimes occur in an experimental setting, the acceptance of the repeat element in the field study was relatively high, serving as another indicator for an intrinsic willingness of participants to achieve a certain outcome at the end of a wave.

In terms of the look-up element, the lookup ratio of the field study was 1.41 look-ups per wave (variance=2.64). In contrast, the experiment produced an average look-up ratio of 22.19 per wave. This difference in usage is the strongest indicator of the influences of the experimental setting and the higher ambitions of the experimentees to achieve high outcomes.

Summarizing these additional, comparative evaluations, we are satisfied with the design as well as the consistency of the overall effects and usages of the game design elements in terms of performance and ongoing engagement motivation. However, the actual learning outcome of the game should be assessed with another experimental study set in a real-life scenario.

6. Chapter 6

Application and Testing of Game Design Elements: Repeat & Look-Up

"Repetitio est mater studiorum"

Horace

6.1. Experimental Study 2 – Effect of Repetition and Look-Up on Long-Term Learning Outcomes in Correct Waste Sorting[24]

6.1.1. Introduction

As stated in an interview on "The Future of Waste Management," Information Systems (IS) can teach citizens where exactly to dispose of different types of waste (Hawlitschek, 2020). Multi-disciplinary research has shown that games, in particular, are successful educational tools and supplements (Fileni, 1988; Van Eck et al., 2017). By applying gameful design to a real-life context, education can be effectively manipulated, whether as fully conceptualized games or as strategically implemented gamification affordances (Barata et al., 2013; Landers & Landers, 2014). In their meta-analysis on digital games and learning, Clark et al. (2016) found significant correlations between quality of design and learning outcomes, highlighting the value of deliberation on specific design decisions.

We created a waste sorting training game based on best practices of game design as well as learning theories to address the prevalent lack of waste sorting knowledge. The game's release was in 2015, and, as of April 2021, it was downloaded over 31,684 times on the Apple, Microsoft, and Windows mobile app stores. As stated by Bellotti et al. (2013), a serious game's purpose is twofold: to be fun and entertaining as well as educational. Thus, we must assess both aspects. While the field data allowed us to ascertain the game's success in matters of game fun and engagement with a certain degree of external validity, we could not reliably infer the game's efficacy in terms of the intended learning outcome. As this is the game's primary aim, we prepared a lab experiment to measure the game's learning outcome under the following research question:

Does gameful design afford learning about the correct sorting of waste items into their target bin?

[24] This chapter comprises the authors accepted manuscript of an article published as the version of record in Business & Information Systems Engineering © 2021. Reprinted with permission. https://doi.org/10.1007/s12599-021-00731-x. Full reference: Hoffmann, G., Pfeiffer, J. "Gameful Learning for a More Sustainable World." Business & Information Systems Engineering (2021): 1-24. https://doi.org/10.1007/s12599-021-00731-x Note: The supplemental material of the authors accepted manuscript in Appendix C.2. The appendix is also based on joint work by the authors. Tables, figures, and appendices were renamed, reformatted, and newly referenced to fit the structure of the thesis. Chapter and section numbering and respective cross-references were modified. Formatting and reference style was adapted and references were updated. Opening quotation was not part of the article.

Gamification/gameful design[25] is a praxis that consists of designing suitable "service bundles" (Blohm & Leimeister, 2013) by adding game design elements to the respective core offer—or core gameplay when the gamified product is a game in itself. The core interaction—core gameplay—of our game is based on a combination of sorting and feedback, the latter being particularly beneficial for knowledge transfer and player engagement (Bellotti et al., 2013; Sicart, 2008). However, during the development of the first prototype, our user tests found that the core gameplay by itself did not engage players long enough to benefit from a long-term learning effect. We decided to add optional design elements that would offer players more choices on how to engage with the learning content. We based this decision on motivational theories that highlight autonomy as a fundamental factor of intrinsic motivation (Ryan & Deci, 2000). We first chose a repetition element that would allow players to repeat a level—or wave—without penalty. The overall learning benefits of repetition are well-documented across different learning domains (Ahmadian, 2012; Bygate, 1996). However, as its inherent repetitiveness could interfere with the game fun, we wanted to gain insights into the potential detriments and benefits of including such a design element. We then added an index element, where waste items can be looked up penalty-free during the core gameplay (look-up element). This was inspired by testers frequently asking why certain items were assigned to bins other than expected. We conducted a literature review to find theoretical leads on the expected learning outcome of such an index-based design element. Lacking a related foundational theory, we analysed research on related contexts: instructional explanations, dictionaries and help tools (Miller & Gildea, 1987; Ryan & Shin, 2011) finding mixed expectable outcomes. Thus, we designed our experiment in a way to further answer the second research question:

What effect does a repetition-based and a look-up-based game design element have on the learning outcomes of correct sorting of waste items into their target bin?

The experiment consisted of five treatment groups to reflect both research questions. The first four learned to correctly sort waste items by playing the game. They differed with respect to whether participants played only the core gameplay, one or both design elements (repetition, look-up, or combined). The fifth group completed the training with common, paper-based teaching materials on waste sorting. As we wanted to ensure long-term retention—long-term memory storage of the learned content—we measured learning outcomes 10 to 12 days after the participants had been trained. Also, while the training was conducted with a game, the learning outcome was supposed to be translated into real life. To test if participants successfully managed this knowledge transfer, we measured the learning outcome in three different ways: first, by testing knowledge retention within the training medium itself in a slightly altered version of the core gameplay designed to test each training item exactly once. Second, we measured knowledge transfer in an abstracted setting via a multiple-choice test featuring only the names (written words) of the trained items. Third, we measured knowledge transfer to real-life through a sorting task with real-world waste items.

Our results showed that the treatment trained with the game significantly benefited with regard to learning outcomes of waste sorting knowledge compared to the treatment group given the non-game materials. This is especially remarkable as, in contrast with other studies (Größler et al., 2000; Luo et al., 2019),

[25] In this manuscript, we refer to both practices under the umbrella term gameful design: "affording ludic qualities or gamefulness (the experiential qualities characteristic for gameplay) in nongame contexts" (Deterding, 2016). The term encompasses the practice of creating as well as research into the effects of serious games and gamification implementation.

we demonstrated that knowledge transfer to real-life can be successfully achieved with a gameful application. We further found that the combination of the repetition and look-up game design elements showed significantly higher learning outcomes within the original content domain as well as the reduced setting. Interestingly, this combinatory effect was lost in the transfer to real life.

We contribute to the literature and practice in several ways. While growing in numbers, assessments of the effects of specific game design elements are still rare (Bellotti et al., 2013; Kim & Shute, 2015). This makes it difficult for both researchers and practitioners to derive informed design decisions from research. We identified a research gap with regard to expectable effects and outcomes of an optional look-up element and found that its implementation contributed to the learning outcome, especially when combined with a repetition-based design element. We also tested learning outcome in terms of long-term retention as well as knowledge transfer to ensure that an actual learning outcome was achieved. While our game successfully achieved the transfer of content to long-term memory, the differences in outcomes between the three different measures highlighted the importance of testing and ensuring successful knowledge transfer in multimedia contexts such as ours.

By constructing and testing a serious game on teaching correct waste sorting and detailing design rationales for future reproduction, we contributed to the ongoing effort of enhancing sustainable IS (see, e.g., Elliot and Webster (2017) and Stanitsas et al. (2019)). Our results showed that successful learning outcomes could be achieved through meticulous gameful design even in less intrinsically motivated and attractive domains such as waste management and even outside a socially mediated learning context.

6.1.2. Related Work

Empirical Findings on Gameful Design in IS Research

As early as 1991, Duffield showed that computer games provide great learning opportunities for students, as games motivate them to learn, provided that they are adequately designed and that the content of the software, problems presented, and instructional methods are carefully aligned (Clark et al., 2016; Duffield, 1991). Later, Rieber (2005) and Gee (2003) recommended games as potential learning tools, reasoning that gaming is a complex social practice in which players engage in high-order thinking and where they need to make a complex cognitive effort. Studies have shown that games entertain, instruct, change attitudes and enable skills development. Studies successfully correlating participants' previous gameplay experiences to related real-life skills, e.g., reaction games and driving skills (Vichitvanichphong et al., 2016) and strategy games with management skills (Simons et al., 2020), supported this finding.

In terms of application domains, we found certain topics especially prevalent, such as education (Fotaris et al., 2016; Sanmugam et al., 2016), fitness (Jang et al., 2018; Kappen et al., 2018), health (Allam et al., 2015; El-Hilly et al., 2016; Kurtzman et al., 2018) and the economy (Hamari, 2017; Luís Filipe Rodrigues et al., 2016). Most sustainability-related studies are connected to broader economic domains like sustainable transport education (Putz et al., 2018) or domestic energy engagement (Gustafsson et al., 2009). Studies solely focused on topics of sustainability are scarce, particularly with regard to sustainable waste management (Elliot and Webster 2017).

Gameful Design and Waste Sorting

In a literature review of serious games in the general domain of sustainability, Stanitsas et al. (2019) found that there has been a radical increase in the development of sustainability-related games since 2010. However, of the 77 listed games (starting in 1990), only two are related to waste management: a board game that teaches about industrial waste management (Jürgen Strohm, 2001) and a role-playing game that educates on irrigation management (Burton, 1994). Both games provide a broad perspective on the topic but do not specifically teach about municipal waste sorting. As an extension to Stanitsas et al. (2019), we conducted an additional literature review looking for research studies with a focus on gameful design and waste sorting. In total, we found nine more research studies somewhat related to our topic (see Table 19).

Table 19 – Overview of Other Gameful Design-based Studies on Waste and Recycling

Authors	Sub-Domain	Digital/ Analogue	Publishing Domain	Country	ql/qt*	Study Design	Study Goal
Berengueres et al. (2013)	Gamification	Analogue (digitally enhanced)	Human-Robot Interaction	United Arab Emirates	ql	n not reported (two waste bins installed)	Evaluate effective system to increase usage of recycling bins
Sreelakshmi et al. (2015)	Gamification/ Game-based Learning	Digital (unity 2D)	Computing Communication and Networking Technologies	India	ql	n=20 participants	Highlight the success of game-based learning through a waste sorting game
González-Briones et al. (2018)	Gamification	Analogue (digitally enhanced)	Distributed Computing and Artificial Intelligence	Spain	ql	n not reported (30 waste bins installed)	Generate motivation for citizen participation in the recycling chain
Bifulco et al. (2011)	Serious Game	Digital (3D virtual environment	Distributed Multimedia Systems	Italy	ql	No report of scientific testing	Present the main concepts of waste collection and garbage recycling to primary school students
Lotfi and Mohammed (2014)	Serious Game	Digital (browser-based)	Computer Applications	Morocco	ql	n=20 participants	Help instructors and experts improve teaching strategies
Menon et al. (2017)	Serious Game	Digital (Microsoft Kinekt)	Serious Games and Applications for Health	India	ql	n=9 participants	Raise awareness of the importance of trash removal, initiate recycling programs, and teach basic hygiene practices
Whalen et al. (2018)	Serious Game	Analogue (board game)	Resources, Conservation and Recycling	Sweden	ql	17, 18, 36 participants (total: n=71)	Teach the benefits and complexity of the circular economy
Idrobo et al. (2018)	Serious Game	Digital (3D)	Telematics and Computing	Colombia	ql	n=5 participants	Teach correct waste sorting
Luo et al. (2018)	Digital Sorting Game	Digital (2D)	Environmental Management	Canada	qt	n=50, n=100, n=308	Evaluate the effect of immediate feedback on recycling and composting accuracy
Hoffmann 2021	Serious Game / Gameful Design	Digital (2D / mobile)	Information Systems	Germany	qt	n=215	Evaluate game / game design elements affording learning of waste sorting knowledge

*ql= qualitative, qt= quantitative

Eight of the nine studies did not prove entirely useful to our research efforts, as they either presented their gameful approach without actually evaluating the effectiveness of their design (Berengueres et al., 2013; Bifulco et al., 2011; González-Briones et al., 2018), evaluated their design qualitatively as a whole with a small number of users (Idrobo et al., 2018; Lotfi & Mohammed, 2014; Menon et al., 2017; Sreelakshmi et al., 2015) or touched on a relevant but adjacent topic (Whalen et al., 2018). Except for one (Luo et al., 2019), these studies did not provide insights into the effect of single design choices on learning outcomes but instead looked at their gameful implementations as a whole.

The study closest to our setup was Luo et al. (2018). In their series of experiments, the authors assessed the effect of immediate feedback as a game design element on the learning of accurate recycling and composting. In their lab experiment, they asked 100 students to sort 80 pictures of different waste items into one of the four bins shown on screen. The learning condition received feedback on the correctness of their sorting, while the control condition did not. The learning outcome was tested after one week in the same game to assess long-term retention. Results showed a positive influence of immediate feedback on the learning outcome. Yet, they could not replicate this result in a real-life follow-up experiment. The authors hypothesized that the reason for this null effect could be related to the logistics of accurately measuring changes in real-life waste containers. Evaluating their design artifact—as presented in the manuscript—from a game design perspective, we think that the lack of game design elements like world-building or colorful aesthetics could also have contributed to this outcome. We came to this conclusion because user tests of the early iterations of our game indicated that feedback alone lacked incentives to continually engage with the content.

We conclude our literature review with the insight that research on gameful design—particularly with regards to the analysis of game-specific design elements—in the domain of waste management and sustainability can benefit from further research in terms of expectable outcomes on learning. We will discuss this in the following section.

6.1.3. Hypothesis Development

We based the theoretical foundations of this work on learning theories, particularly on models introduced by instructional design and the didactic method (Nitsch et al., 2016; Wittwer & Renkl, 2008). Most learning theories have been developed in and for contexts where social interaction is interwoven into the learning process (e.g., Chi et al. 2001). However, multimedia learning is often designed to work outside of social interaction contexts. We thus built on learning theories and strategies that have proven effective outside of a socially embedded learning context, namely elaborative encoding (for Hypothesis 1) and repetition (for Hypothesis 2). While these strategies afford an empirical learning process, rationalization (the process of sense-making or understanding "why") is equally essential for providing meaning and context to learning matter (Wittwer & Renkl, 2008). To offer explanations without overwhelming players, we implemented an optional, dictionary-inspired look-up element as a complementary game design element and elaborated on its supposed learning outcome in the development of Hypothesis 3.

Depending on the context, the learning outcome can be measured with regard to different facets. In terms of memorization, the learning outcome often differs when tested immediately after the training phase rather than during a post-training transfer phase (Keith and Frese 2008). By measuring the learning outcome in terms of *long-term retention*, we wanted to ensure that the content was memorized in the long term to achieve successful change in real-life waste sorting behavior.

Furthermore, in contexts like ours, where the training medium differed from the application context, it was particularly important to assess the learning outcome with regard to *knowledge transfer* (Barnett &

Ceci, 2002). In language transfer theory, knowledge transfer happens within the context of translation, where words and meanings are connected between two languages, typically native and second (Mahmood & Murad, 2018). In the didactics of mathematics, this transfer is referred to as "conversion," or the mental merging of different representations—such as graphs and functions—of the same mathematical concept (Dreher et al., 2014). Expanding on the concept of conversion, Nitsch et al. (2016) differentiate two phases of the transfer process: identification (comparison of the incorporated information and identification of similarity with existing schema) and construction (transferability of the incorporated information into new situations). Differentiating the two is important because construction measures whether the content has been understood deeply enough for reapplication in new contexts. Therefore, as later explained in the section on the operationalization of our outcome variable, we measured knowledge transfer with different measurements that capture both identification and construction.

Elaborative Encoding

The overall story and game design rationale of the game are based on elaboration strategies. Elaborative encoding belongs to the category of learning techniques known as mnemonics. In this learning strategy, loosely adjacent content items are added to the learning matter. Offering more associations that may connect to the learners' existing knowledge facilitates embedding new information into prevalent mental structures (Bradshaw & Anderson, 1982). Examples of mnemonics involve meaning-enhancing additions: constructions or creations that improve one's memory of what is learned (Levin, 1988). Mnemonics can range from acronyms and rhymes to complex strategies for remembering numbers (Putnam, 2015) and character designs (such as the mascot designs commonly found in Japan).[26] Elaborative encoding encompasses the purposeful addition of information, whether visual, semantic, spatial, or acoustic, to create more retrieval paths in the mind of the learner from existing knowledge structures to the learning matter (Bradshaw & Anderson, 1982). In his meta-analysis of elaboration studies, Mayer (1980) concluded that associative elaboration "increases retention performance as compared with control or simple repetition procedures" (p. 771). This technique is particularly valuable in the context of gameful design, as elaboration can occur on several layers at once: the game's mechanics (rules and systems), aesthetics (visual/auditive representation), and narrative form a multi-sensory context for knowledge transmission (Hunicke et al., 2004; Westera et al., 2008). This is particularly important in serious games like Re-Mission (Hopelab, 2006) that translate a specific problem context into gameplay: adolescents are incentivized to take their cancer medication on a regular basis by transferring the game setting to their own bodies and providing them the medication as ammunition against destructive cancer cells. Re-Mission has proven highly successful in improving the health outcomes of its players (Kato et al., 2008). In sum, due to their multimodal elaborative encoding of real-life activities, principles, and systems, games have been found to be an effective medium for teaching. Thus, we hypothesize that:

H1: *Learning waste sorting through a game rather than with state-of-the-art paper-based information on the correct sorting of items increases learning outcome.*

[26] For example: "[Morio-kun is] a vampire bat who promotes paying taxes by direct debit in Chiba, Japan. He uses direct debit because he's nocturnal and can't get to the bank in the daytime." (https://twitter.com/mondomascots/status/1020495338644230144).

Repetition Strategies

However, in terms of the core teaching effort—correctly sorting waste items—studies have shown that a single exposure to new content is not enough for learners to effectively encode that content into memory (Bygate, 1996). Repetition is a learning activity in which students repeat individual facts to create firmly anchored connections in their long-term memory. Repetition has long been acknowledged as a powerful learning mechanism. As a universal principle, it is part of all prevalent learning theories: behaviorism (e.g., in Pavlov [Dunsmoor et al. 2007] and Skinner (1936), cognitivism (e.g., in schema theory), constructivism (e.g., in Piaget [Greenfield and Savage-Rumbaugh 1993]) and social learning (e.g., Vygotsky (1967) on child development). The underlying theme relates to the formation of memory in the brain. In their research on working memory, Baddeley and Hitch (1974) stated that by repeatedly forming mental connections through reflection and deliberate recall, the stored information gets retrieved more easily and quickly. However, different studies have shown that repeated exposure to the same content does not necessarily lead to improved learning (Crowder & Melton, 1965; R. S. Nickerson & Adams, 1979). Memories are formed more precisely and hold for longer in long-term memory if learners are interested in the content and pay attention (R. S. Nickerson & Adams, 1979). This is where games might have an additional advantage compared to learning content presented in a classroom setting.

In their study, Bygate (1996) found that repeating the trained content three days after the initial task led to improvement in fluency and accuracy as well as a marked improvement in repertoire due to growing familiarity with the content. The given reason is that on first contact with the material, learners are primarily concerned with the heuristic planning and understanding of the content matter (Bygate 1996). Ahmadian (2012) corroborated these findings, arguing that it is difficult for learners to focus on form and meaning at the same time. Thus, repetition allows them to gain understanding in both facets. Overall, studies on task-based language learning have reported repetition-based improvements for output factors of accuracy, complexity, repertoire, and task success (Lynch and Maclean 2000; Pinter 2005). According to Driskell et al. (1992), there are even benefits to repeating the content beyond perfect retention. In their study on overlearning, they found a significant overall effect: the greater the degree of overlearning, the greater the resulting long-term retention. By raising the number of occurrences in the brain, the significance of the information is enforced, and so the content is retained longer. Repetition has been proven to be an effective learning strategy in learning tasks across domains (e.g., education (Johnson, 2004), civic knowledge (Ivancic and Hesketh, 2000), and games (Clark & Sefton, 2001)). Building on the theoretical foundations of repetition, we hypothesize that:

H2: *Repetition as a game design element increases the learning outcome.*

Instructional Design, On-Demand Help, and Look-up Strategies

During the teaching process, one of the central functions of the teacher or tutor is to provide context and explanations to learners (Leinhardt & Steele, 2005). In non-social contexts, this function must be substituted within the training medium. While there is no dedicated educational or psychological theory for this construct (providing relevant information/answering the "why" question), research on instructional explanations provides foundational insights. Empirical studies show that instructional explanations have often not been successful in terms of raising the learning outcome (Jonassen & Rohrer-Murphy, 1999;

Leinhardt & Steele, 2005). One explanation was that learners merely engage in superficial processing of instructional explanations (Berthold & Renkl, 2010) and do not attend to the content of the explanations in a meaningful manner (Roelle et al., 2013). However, the learning outcome was positively influenced by instructional explanations if learners rationally engaged with the content of the explanations (Wittwer & Renkl, 2008) and if there was a meaningful follow-up activity after receiving the explanations (Webb & Farivar, 1999). The way the game is designed, each interaction with the look-up element is followed by the sorting of a waste item. Thus, in our case, the instructions can be processed meaningfully.

In self-regulated learning contexts, studies have found help-seeking to be a successful strategy for learning (Ryan & Shin, 2011; Webb et al., 2013) if help-seekers were oriented on independent problem-solving (Gall, 1981) and if the process included asking for explanations and hints (Mäkitalo-Siegl et al. 2011). In summary, if learners are invested in the learning process, giving explanations when needed raises the learning outcome. This connection produces positive indications for the success of our look-up element. However, most studies on help-seeking are embedded in a social context: the help is provided by another person. Thus, the expected effects might be weaker outside of a social context. On the other hand, the same studies found that the social context of help-seeking produced a different problem: those learners needing help the most (students with low self-efficacy) were less likely to seek it out, as they feared being perceived as lacking in ability and thus lose social standing (Ryan & Shin, 2011). This negative effect could be neutralised in our case, as the game provides social anonymity within the look-up process, potentially resulting in lower inhibitions to use the look-up element and benefit from its content.

Interestingly, the IS literature on help tools (Clarebout and Elen (2009); Größler et al. (2000); Mäkitalo-Siegl et al. (2011)) did not confirm these positive expectations of optional help-seeking tools on learning outcomes. The most common reason provided was that tools were barely used (Aleven, Stahl, Schworm, Fischer and Wallace 2003; Größler et al. 2000; Liu and Reed 1994). The general unwillingness of the participants to accept help partly explained these findings, as has been found in various educational settings (Aleven et al., 2003; Newman, 2000; Ryan et al., 2001). One explanation for such usage inhibitions was that the help function was sometimes perceived as cheating (Clarebout & Elen, 2008). The factors found to influence how students behaved in open learning environments were the students' self-efficacy, motivation, and perception of the task. If they felt the task was performance-oriented, they were less likely to use the help tools than when they perceived it as learning-oriented (Clarebout & Elen, 2008). As our game is not only learning-oriented but related to a serious and meaningful task, we believe that such inhibitions regarding the look-up element might be alleviated.

Finally, while looking at the literature on cognitive psychology, we found a dichotomy of two error-related learning strategies: errorful and errorless learning. The former—also referred to as trial-and-error learning—is "the process of making repeated trials or tests, improving the methods used in the light of errors made, until the right result is found" (Webster's New World College Dictionary 2005). It builds on the repetition-based learning strategy that the repeat element of our artifact is based on. Interestingly, we found that this strategy was juxtaposed with an opposite strategy—errorless learning—which is defined as "an approach whereby the task is manipulated to eliminate/reduce errors. Tasks are executed in such a way that the subject is unlikely to make errors" (Fillingham et al. 2003 p. 339). This was partially fitting for us, as the look-up element would allow players to play the game without error if they chose to use it before every decision. However, when comparing studies that used errorful vs. errorless teaching strategies, neither one was found to be more effectual (e.g., Clare et al. 1999) or the results were inconclusive (e.g., Johnson 2004) (see Table 52 in Appendix C.2.3).

Looking at the volatile nature of instructional explanations and help/look-up tools, we believe that, in particular, the optional and anonymous nature of the look-up element as well as the fact that the game affords a meaningful follow-up activity (sorting the waste after the explanation has been provided) can alleviate some of the negative effects listed in the mentioned theories and studies. Furthermore, given citizens' almost daily interactions with waste, the look-up design element can add meaningful context to already existing knowledge structures. Finally, as our learning element only offers information when the learners reach an impasse and are actively inquiring for context and a solution (as recommended by instructional design theory; Wittwer and Renkl 2008), we hypothesize that:

H3: A look-up game design element increases the learning outcome.

6.1.4. Experiment

Exclusion of Game Design Elements for the Experiment Version

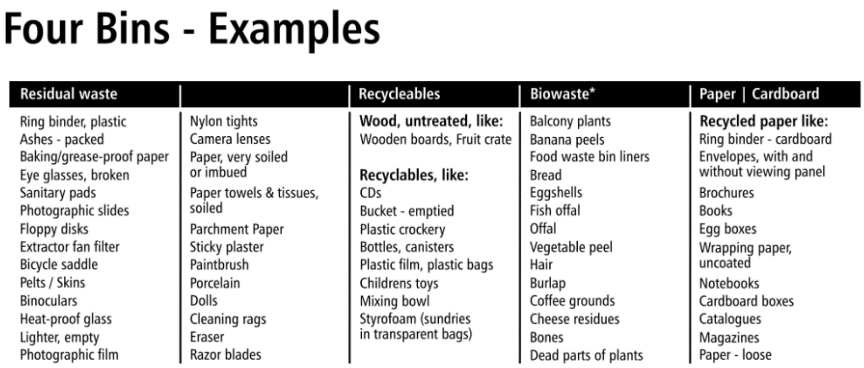
For this experiment, we compiled another abridged version of the game that only included the design elements we were testing for: the core gameplay—including the tutorial—as well as the perfect stamp and the two additional learning enhancing design elements—the repeat option and the look-up feature. We shortened the core gameplay from 34 levels to 10 and from 201 waste items to 108 (eight were used as exemplary items in the tutorial, and the remaining 100 were distributed over the 10 waves, introducing 10 new items and reusing five previously seen ones per wave). To avoid confounding influences, we stripped the experimental version of all design elements that related to motivation enhancement (narrative elements and unlockable features). We wanted to ensure an isolated observation of the effectiveness of the core gameplay in producing a learning outcome. We kept the underlying worldbuilding and setting (monster design and waste sorting plant) as they are integral to the game feel.

Experimental Design and Independent Variables

We designed the laboratory experiment to test the effect of the game in general as well as two independent variables (look-up and repeat) on the learning outcome. We designed a between-subject experiment in three stages where the 10-12-day duration between Phases 2 and 3 served as the retention period. We designed the experiment with four treatments in a full-factorial design with an additional fifth control group (from now on referred to as non-game material) that was given exemplary teaching material as used by waste management institutions. The used non-game teaching material consisted of the three informative flyers conventionally provided by the city of Karlsruhe to teach citizens correct waste sorting. The first flyer listed exemplary waste items for each bin (see Figure 70 (excerpt) and Figure 96 in Appendix C.2.2, the second informed on the general categories of waste that go into each of the four bins, and the third served to differentiate the general waste categories in combination with the underlying rules of what waste belongs where (see Figures 97 and 98 in Appendix C.2.2)). An overview of the treatments' structure is provided in Table 20.

Table 20 – Treatment Overview

Treatment group	Implementation
Control group: Non-game materials	This group received non-interactive learning materials as currently provided by the municipal waste department of Karlsruhe, which consisted of two flyers introducing the general waste assignments and an exemplary list of the items with their correct bins (see Figure 54).
Game group: Core gameplay	This group was given an instantiation where only the core gameplay was implemented (see Figure 34).
Game group: Repeat element	On top of the core gameplay, at the end of each wave, the players of this group were given the option to repeat the wave without penalty (see Figure 36).
Game group: Look-up element	On top of the core gameplay, the players of this group were introduced to and had permanent access to the look-up element, giving them the option to look up the correct bin for any waste item they encountered (see Figure 37).
Game group: Combined repeat and look-up element	On top of the core gameplay, the players of this group could access the look-up element at any time, and after each wave, they were given the option to repeat without penalty.



Four Bins - Examples

Residual waste		Recycleables	Biowaste*	Paper | Cardboard
Ring binder, plastic	Nylon tights	**Wood, untreated, like:**	Balcony plants	**Recycled paper like:**
Ashes - packed	Camera lenses	Wooden boards, Fruit crate	Banana peels	Ring binder - cardboard
Baking/grease-proof paper	Paper, very soiled or imbued		Food waste bin liners	Envelopes, with and without viewing panel
Eye glasses, broken		**Recyclables, like:**	Bread	
Sanitary pads	Paper towels & tissues, soiled	CDs	Eggshells	Brochures
Photographic slides		Bucket - emptied	Fish offal	Books
Floppy disks	Parchment Paper	Plastic crockery	Offal	Egg boxes
Extractor fan filter	Sticky plaster	Bottles, canisters	Vegetable peel	Wrapping paper, uncoated
Bicycle saddle	Paintbrush	Plastic film, plastic bags	Hair	
Pelts / Skins	Porcelain	Childrens toys	Burlap	Notebooks
Binoculars	Dolls	Mixing bowl	Coffee grounds	Cardboard boxes
Heat-proof glass	Cleaning rags	Styrofoam (sundries in transparent bags)	Cheese residues	Catalogues
Lighter, empty	Eraser		Bones	Magazines
Photographic film	Razor blades		Dead parts of plants	Paper - loose

Figure 70 – Flyer on General Waste Sorting in Karlsruhe, Excerpt, Translated to English

Experimental Procedure

We recruited participants from a large German university using the organizing and recruiting software hroot (Bock, Nicklisch, Baetge 2012). Potential participants in the experiment had to meet three requirements to participate: they needed to own a smartphone with an Android-based operating system running on a version higher than 2.3.1 (Gingerbread), be willing to download and install the application on their phone and be fluent in German. We conducted the experiment in three stages (see Figure 71).

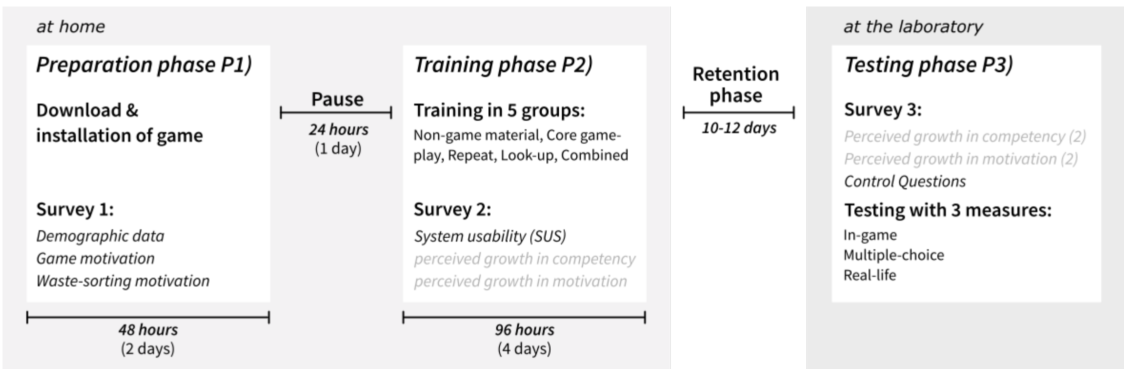


Figure 71 – Experimental Procedure Experimental Study 2

P1: After closing the registration for participation in the experiment, we randomly assigned participants to one of the five experimental groups in a between-subject design. We sent instructions via email for participants to fill out the first survey and download the app (as an apk-file). The game within was locked by a server to prevent premature play and could only be accessed once the first phase officially started. In the survey, we assessed demographic information (age, gender, how long they had been living in Germany, how long they had been living in the city in which the experiment was conducted), participants' game motivation (how much they were involved in and how they felt about these games in general) and their general waste sorting motivation (how they felt about municipal waste sorting). We also included several controls checking language proficiency and conscientiousness in answering the questions. To ensure absolute anonymity in the datasets when linking the game data to the survey entries, each app showed a unique code that participants had to report in each respective survey. For this phase, we set a 48-hour timeframe followed by a pause of 24 hours that allowed for troubleshooting.

P2: In the second phase, we sent the next set of instructions as well as another survey link via email. We instructed the participants on the four game-based treatments to open the application and play it through to the end and then complete the survey. In contrast, we told the control group with the non-interactive materials to attentively read through the teaching materials provided through the link for 25 minutes (this time was derived from the average playtime of the experimental version of the game during the pre-tests) and then complete the survey. The last part of the survey was the same for all treatments: we measured the perceived usability of the application—or the materials in the case of the non-game material treatment—with the system usability scale (Brooke, 1996) as well as self-stated perceived growth in competency and growth in motivation. To adapt the 30 minutes of focused attention to the survey and training, we gave participants a four-day timeframe—including a weekend—to finish the task. We scheduled the final sessions 10-12 days after the deadline for the second phase, depending on the day of the assigned session.

P3: The experiment took place in a laboratory in 19 experimental sessions. Each participant was seated in a cabin where they were guided through the first part of the experiment with the final survey. We first asked participants about their perceived growth in competency and growth in motivation, and there was a final control question on any prior knowledge about the project. Next, we tested the learning outcome in three different performance measures. First, the participants completed a multiple-choice test in which they had to match all 108 trained waste items. Second, we asked all participants to take their phones and start the game application, where they had to sort all 108 items in a special version of the game. Here, each item appeared only once in one big game wave without the two additional design elements. Third, we called the participants into a separate room, where we asked them to sort a selection of real-life waste items.

After the experiment, participants received a flat payment of €15 for their time. This experimental design was pre-tested with seven participants.

Operationalization of the Dependent Variable: The Learning Outcome

We measured the learning outcome with special regard to two factors: long-term retention and knowledge transfer. According to cognitive theory, long-term retention can be tested as soon as two or three days past the training period (Schmidt & Bjork, 1992). For our study, we chose an extended retention phase of 10–12 days to ensure the success of the transfer to long-term memory (see also Luo et al. (2018); Parkin and Streete (1988)). In their work on training evaluation, Kraiger et al. (1993) highlighted the importance of conceptually sound measures of learning that ensure training effectiveness with regard to knowledge transfer. We tested knowledge transfer in three ways: first, by testing identification (Nitsch et al., 2016) of knowledge by evaluating if players can reproduce the learned content within the training medium. For this, we used a special version of the game (game measure) featuring one wave where all 108 trained items appear one by one from the right side of the screen and then have to be sorted into the correct bin before they drop off on the left side (see Table 21). We then tested knowledge transfer via a multiple-choice-based test measure as a power test (number of items answered correctly in an unlimited amount of time (Kraiger et al. 1993). We chose this testing measure because multiple-choice tests are considered best suited for measuring the retention of declarative knowledge (Bellotti et al., 2013; Gagné, 1984). In this measure, participants were given the names of the waste items but not images like in the game measure. By offering only one of the two memory connection items, we could differentiate the effectiveness of the representational elements (pictures vs. text) (Mayer, 2002). Participants were asked to assign the right bin for each of the 108 trained items (the options were residual, recycle, biodegradable and paper waste, and separate recycling) (see Table 21).

Finally, we measured knowledge transfer to the final application domain: real-life waste sorting. This measure relates to the construction item introduced by Nitsch et al. (2006), where knowledge is retained and understood in a way so that it can be reapplied to a different context. In this third measure, participants had to sort a selection of real-life waste items into the correct bin (see Figures 72 and 73). Seven representative waste items were chosen for the real-life sorting according to the participants' performances measured in Phase 2 of the experiment: one from the top five items of best average sorting performance (aluminum), two from the average of their sorting performance (adhesive tape and milk cartons), and four that belonged to the group of the 20 worst-performing items (CDs, thermal paper, empty ring binder, and wood shavings).

Figure 72 – Representative Waste Bins: Bio, Paper, Recycle, Residual

Figure 73 – Real-Life Waste Items

To increase the comparability of the three measures in consideration of the different number of items, we decided to use percentages of correctly sorted items. Thus, for each person and measure, we divided by the number of items sorted. For example, a measure of 85.71% for the real-life sorting performance meant that the participant sorted five out of the seven items correctly (see Table 21).

Table 21 – Dependent Variables for Measuring the Learning Outcome

Dependent Variable	Range and Meaning	Theoretical Construct
In-Game Performance	Continuous value between 0 and 1: the percentage of correctly sorted waste items out of 108	Identification
Multiple-Choice Test	Continuous value between 0 and 1: the percentage of correctly sorted waste items out of 108	Knowledge transfer with reduced stimuli
Real-Life Sorting	Continuous value between 0 and 1: the percentage of correctly sorted waste items out of 7	Knowledge transfer to real-life / construction

Control Variables

Apart from controlling for demographic factors (age, gender, how long the participants had been living in Germany, and the city in which the experiment was conducted), we controlled for the following: **Gaming motivation.** Since gamified systems were previously perceived as less serious than traditional teaching content (Brigham, 2015; Hanus & Fox, 2015), the acceptance of the medium might influence the willingness to learn. We thus measured user attitude towards the medium in general through self-reporting (the full implementations can be found in Table 53 in Appendix C.2.4). **General waste sorting motivation.** Since the personal attitude to the topic plays a role in the learning outcome (Garris et al., 2002), we also measured the general attitude towards waste sorting at home through two questions. **System usability.** The usability of the respective information system plays an equally important role as poor user experience can lead to frustration and thus have a negative impact on user interaction (Bangor et al., 2008). We decided to assess user satisfaction with Brooke's (1996) system usability scale (SUS). This decision was based on its widespread usage in IS for such purposes and to allow for comparability between our artifact and similar studies (Bangor et al., 2008).

6.1.5. Results

The first stage of the experiment was completed by 266 participants. Thirty-one participants did not complete all three stages of the experiment (17 participants did not start or finish Phase 2, and 14 more did not show up to Phase 3 in the lab). Of the 235 remaining participants, we had to exclude 14 further datasets because of transmission errors (e.g., the game data of the second or third stage of the experiment was missing) and one for failing a crucial control question. Finally, of the remaining 220 data sets, there was a minor data transmission error for 23 participants: not all single item sorts for the in-game performance had been transmitted completely. We decided to exclude the datasets where more than 30% of the item sorts were missing (five out of these 23). This decision was backed by a Kruskal-Wallis test that indicated that the performance of the 18 participants with more than 70% but less than 100% correctly transferred item sorts did not differ significantly from the participants with complete sets of item sorts. We thus decided to include them, leaving us with a total of 215 complete datasets. The average age of the participants was 22.72 years old (one person reported the age of 3, which we set as a missing value because this was either a typo or intentionally misreported), and the gender distribution was 66.05% of participants identifying as male vs. 33.49% as female vs. one person (0.47%) indicating "other." Table 22 shows the descriptive statistics of the dependent measures for all treatments. For example, in the treatment with non-game materials, participants correctly sorted on average 70.8% of the items in the in-game performance measure, 59% in the multiple-choice test, and 70.3% in the real-life sorting task. The pattern of having the lowest performance when measuring with the multiple-choice test compared to the other two learning outcome measures is stable over all treatments. The largest value of 78.8% was reached in the combined group for in-game performance. For more details on the descriptive statistics for both the dependent measures as well as the control variables, please see Tables 54 and 55 in Appendix C.2.4.

Table 22 – Descriptive Statistics of Dependent Measures

		In-Game Performance		**Multiple-Choice Test**		**Real-Life Sorting**	
	n	mean (min/max)	std. dev.	mean (min/max)	std. dev.	mean (min/max)	std. dev.
Non-game material	39	.708 (.463/.889)	.100	.590 (.259/.815)	.114	.703 (.286/1)	.181
Repeat element	46	.736 (.509/.926)	.097	.670 (.444/.861)	.108	.776 (.286/1)	.177
Look-up element	45	.748 (.491/.898)	.107	.676 (.380/.870)	.103	.775 (.429/1)	.171
Combined	41	.788 (.574/.917)	.079	.709 (.528/.870)	.092	.767 (.286/1)	.177
Core gameplay	44	.727 (.544/.870)	.093	.645 (.333/.852)	.107	.773 (.286/1)	.142
Overall	215	.741 (.463/.926	.098	.659 (.259/.870)	.870	.760 (.286/1)	.171

For all statistical tests, we computed ordinary least square (OLS) regressions with the three continuous performance measures ranging between 0 (0% correctly sorted) and 1 (100% sorted correctly) as dependent variables. All our hypotheses were directed, and therefore a test was significant if p of the two-tailed tests in the presented tables of the statistical tests was below 10%. Robust standard errors were used in all regressions to account for heteroscedasticity, based on the Breusch–Pagan test (Cohen et al., 2014). Furthermore, we bootstrapped the results with a sample size of 5,000 to account for non-normality of residuals (Pek et al., 2018; Tibshirani & Efron, 1993).

To compute Hypothesis 1, we had to pool the treatments core gameplay, repeat element, look-up element, and combined group into one group because Hypothesis 1 compared the games' performance with the non-game materials. In contrast to the other two hypotheses, it did not focus on the effect of specific game design elements and their related individual treatments. We thus computed a binary variable "Game" that took the value 1 for all observations trained through the game (the pooled group) and the value 0 for the observations in the non-game material treatment. This binary variable was our only independent variable in this main analysis for Hypothesis 1. Table 23 shows the results of the three regressions of this binary variable on each of the three learning outcome measures. We found significant effects on all measures, which supported Hypothesis 1. When tested with the in-game performance measure, the game treatments were estimated to correctly sort 4.1% more items than non-game treatments. For the multiple-choice test, the effects were even larger: the game treatments were estimated to correctly sort 8.4% more items than non-game treatments. Finally, for the real-life sorting measure, the estimate was 6.9%. To sum up, we could fully support Hypothesis 1 for all three performance measures. The effect for in-game performance was surprisingly the weakest, although this was the measurement for which the medium (the digital game) of training and testing was the same.

Table 23 – Effect of the Game in Comparison with the Non-Game Material (OLS regression) (Hypothesis 1)

	In-Game Performance		Multiple-Choice Test		Real-Life Sorting	
Reference category: Non-game material	coef. (bootstr. std. error) [conf. interval]	p (two-tailed)	coef. (bootstr. std. error) [conf. interval]	p (two-tailed)	coef. (bootstr. std. error) [conf. interval]	p (two-tailed)
Game	.041 (.018) [.012, .070]	.020*	.084 (.020) [.052, .117]	.000**	.069 (.031) [.017, .121]	.027*
Constant	.708 (.016) [.682, .734]	.000**	.590 (.018) [.560, .619]	.000**	.703 (.029) [.656, .751]	.000**
N	215		215		215	
R^2	.026		.087		.025	
Adj. R^2	.021		.083		.020	

* p<0.1, ** p<0.01

In contrast to the analysis of Hypothesis 1, Hypotheses 2 and 3 focused on the effect of the examined design elements. Thus, we did not pool all game treatments but rather compared all five treatments with each other. We coded each treatment with a binary variable that took the value 1 if the observation belonged to the respective treatment. The reference category was the non-game material treatment which meant that all coefficients must be compared to the performance in the non-game material treatment.

Table 24 – Effect of the Design Elements in Comparison to the Non-Game Material with OLS (Hypotheses 2 and 3)

	In-Game Performance		Multiple-Choice Test		Real-Life Sorting	
Reference category: Non-game material	coef. (bootstr. std. error) [conf. interval]	p (two-tailed)	coef. (bootstr. std. error) [conf. interval]	p (two-tailed)	coef. (bootstr. std. error) [conf. interval]	p (two-tailed)
Repeat element	.028 (.021) [-.008, .063]	.197	.080 (.024) [.040, .120]	.001**	.073 (.039) [.010, .137]	.058*
Look-up element	.040 (.022) [.003, .076]	.073*	.086 (.024) [.047, .125]	.000**	.071 (.038) [.008, .135]	.066*
Combined	.080 (.021) [.046, .114]	.000**	.119 (.023) [.081, .157]	.000**	.063 (.040) [-.002, .129]	.112
Core gameplay	.019 (.021) [-.015, .055]	.366	.055 (.024) [.015, .095]	.024*	.069 (.036) [.010, .129]	.054*
Constant	.708 (.016) [.682, .055]	.000**	.590 (.018) [.560, .619]	.000**	.703 (.029) [.656, .751]	.000**
N	215		215		215	
R^2	.070		.121		.025	
Adj. R^2	.052		.104		.007	

* p<0.1, ** p<0.01

Table 24 illustrates the results for Hypotheses 2 and 3. When comparing the in-game performance of the treatments with the non-game material treatment, we found a significantly increased learning outcome for the look-up element treatment (estimated increase of 4% of correct item sorts) and the combined one (8%). An additional Wald-test showed that the effect of the combined treatment was larger than that of the look-up treatment (p=0.04). However, the effect for the look-up element was not significantly larger than for the repeat element treatment (again tested with Wald test, p=0.56). Thus, the look-up treatment performed better than the non-game material treatment but not better than the treatment with only repetition. In sum, for the in-game performance measure, Hypothesis 3 was fully supported: we found better performance for both the groups that only had the look-up element by itself or the look-up element combined with the repetition element. Hypothesis 2 was only partially supported: we did not find a stronger performance when only playing with the repetition element. Hypothesis 2 was only supported if repetition was combined with the look-up element. For the multiple-choice test, we found even stronger results and could fully support Hypotheses 2 and 3: all four treatments trained through the game performed better in the multiple-choice test than the treatment trained with the non-game materials. The largest effect was measured for the combined treatment, where on average, participants sorted 11.9% more items correctly than the participants in the control treatment without game materials. For the real-life sorting task, we interestingly found weaker effects for the combined treatment. In detail, we found that only those treatments that had either one design element or neither of those two elements (the core game) performed significantly better than the treatment that did not play the game. Yet, the coefficients also showed that the effects for all four game treatments were rather similar, ranging between 6.3% for the combined treatment to 7.3% for the repeat element treatment. Thus, when conducting further Wald-tests comparing the coefficients of the game treatments with one another, one cannot claim that one game group performed better than another (all $p>0.8$). Thus, all in all, we could support Hypotheses 2 and 3 and found that all game treatments did comparably well.

For game or gamification designers, it is interesting to compare the effects of game design elements not only to the non-game material group but also to the core gameplay group to gain a better understanding of which design elements to include in their gameful applications. Therefore, we want to further focus in detail on the comparison of the different game treatments to the core gameplay group in Table 25.

Table 25 – Effect of the Design Elements in Comparison to the Core Gameplay with OLS (Hypotheses 2 and 3)

	In-Game Performance		**Multiple-Choice Test**		**Real-Life Sorting**	
Reference category: Core gameplay	coef. (bootstr. std. error) [conf. interval]	p (two-tailed)	coef. (bootstr. std. error) [conf. interval]	p (two-tailed)	coef. (bootstr. std. error) [conf. interval]	p (two-tailed)
Repeat element	.008 (.020) [-.024, .041]	.681	.025 (.023) [-.012, .062]	.270	.004 (.034) [-.052, .059]	.914
Look-up element	.021 (.021) [-.014, .055]	.327	.031 (.023) [-.007, .068]	.175	.002 (.033) [-.052, .056]	.954
Combined	.061 (.019) [.030, .091]	.001**	.064 (.021) [.028, .099]	.003**	-.006 (.035) [-.064, .052]	.860
Non-game material	-.019 (.021) [-.055, .016]	.366	-.055 (.024) [-.095, -.015]	.024*	-.069 (.036) [-.129, -.010]	.054
Constant	.727 (.014) [.704, .751]	.000**	.645 (.016) [.619, .671]	.000**	.773 (.021) [.738, .808]	.000**
N	215		215		215	
R^2	.070		.121		.025	
Adj. R^2	.052		.104		.007	

* $p<0.1$, ** $p<0.01$

We found that with the in-game performance measure and the multiple-choice test, the combined treatment achieved a significantly higher learning outcome than the treatment with only the core gameplay available during training, with an increase of 6.1% correctly sorted items with the in-game performance measure and 6.4% with the multiple-choice test. When comparing the groups within the real-life sorting measure, a significantly different learning outcome cannot be discerned. This is a result already highlighted in the analysis above: the game treatments performed comparably well in the real-life sorting task. Thus, for real-life performance, the overall effect of the game itself was much stronger than that of adding the single design elements to the core gameplay.

To further assess the robustness of our results, we also computed robust OLS regressions with these control variables: age, gender, how long they lived in Germany ("Living in Germany"), how long they lived in the city the experiment was conducted in ("Living in Karlsruhe"), their gaming motivation, their general waste sorting motivation and the SUS (for details, see Tables 56 to 58 in Appendix C.2.5). The results were robust regarding the inclusion of these control variables. However, there was one slight change: for Hypothesis 3, the effect on the repeat treatment became significant. Thus, for the statistics with control variables, we could now fully support Hypothesis 3. Regarding the significance of the control variables, we found that the longer participants lived in Germany, the better they performed in-game and in the multiple-choice test. This control variable can be seen as a proxy for prior knowledge about the participants' waste sorting. Furthermore, the general waste sorting motivation value showed a tendency to positively affect the performance measures for all three measures (p ranges between 0.01 and 0.11). The SUS value of the game also had the tendency to positively influence the game performance ($p=0.064$ for all three hypotheses).

6.1.6. Discussion and Conclusions

In terms of our first and overarching research question, we found that the learning outcome for the groups given the game for training was significantly stronger than for the group given state-of-the-art paper-based information during the training phase. This held true across all three measures. Interestingly, the effect was weakest within the in-game performance measure (with 4.1% more items correctly sorted than by the non-game material group) and strongest in the multiple-choice test (with 8.4% more items correctly sorted). This outcome contrasts with the literature on context reinstatement, which suggests that information encoded in one mindset is more successfully retrieved in the same mindset (Fisher and Kraig 1988). This interesting finding was also apparent in the non-game material treatment, where performance in the in-game measure was significantly higher than in the multiple-choice test (Wilcoxon signed-rank test with $z=5.16$; $p<0.01$), although the games' aesthetic and interaction were new to the non-game material treatment.

To gain further insight into this matter, we were interested in whether all participants generally performed better in one measure or the other. We found that performances, when measured in the game (Wilcoxon signed-rank test with $z=12.00$; $p<0.01$) and in real life ($z=7.85$; $p<0.01$), were significantly higher than when measured with the multiple-choice test. We also found that the game-trained group performed comparably well in the game and in real life ($z=1.95$; $p=0.06$) (for the descriptives, see Table 22). A potential explanation for this finding can be linked to the forming of memory connections: the multiple-choice test offered fewer memory connection items (offering only designated connections: words) than the game measure, which presented both iconic and designated connections (words and sticker-like icons) and the real-life measure, which provides real objects. Both the game and real-life objects offered more information items that could connect to existing schemata. This might have helped stimulate memories not activated by the fewer connections offered in the multiple-choice test. This finding is congruent with studies on word and picture learning (Kolers & Brison, 1984; Mayer, 2002) that found that learners performed better through a combination of words and pictures/objects than with words alone. Similarly, in the domain of mathematical didactics, studies have found that using more mathematical representations (like graphs, numbers, and formulas) leads to an increased learning outcome (Ainsworth, 2006). Our results showed that this effect works in both directions: learners retrieved formed memories more successfully if we offered more memory connections with their mental schemata.

In terms of Hypothesis 2—adding repetition as a game design element increases the learning outcome—our results confirmed our conjectures. The group given the additional option to repeat waves showed a significantly higher learning outcome than the non-game material group in two of the three measures (multiple-choice and real-life). This also held true for the in-game measure when inserting control variables. However, when compared with the core gameplay group, the implementation of a repeat option by itself did not increase the learning outcome significantly. The game design elements enhanced learning potential; however, this manifested within the success of the combined design elements. This suggests that the repeat elements inherently lacking in fun can be compensated for better results. This is underpinned by a study by Kim and Shute (2015), who found that changes in just one design element "significantly impacted players' interactions with the game by changing players' mental 'operational rules' during play" (p.351). While the use of the repeat element was optional, it was generally well-received, as 63.95% of players who had the repeat element available used it at least once (mean: 3.88, min: 0, max: 24).

For Hypothesis 3—the increase of the learning outcome through a look-up design element—our results showed that the group given this design element performed significantly better than the non-game material group in all three measures. In terms of usage, the players received it even better than the repeat element, as 67.86% of players who had the look-up element available used it at least once (mean: 14.42, min: 0, max: 85). These are relevant findings given that we found contradictory indicators on the potential outcome in our analysis of related literature (e.g., studies on help tools reported low usage as well as low effects (Aleven et al., 2003; Größler et al., 2000; Liu & Reed, 1994)). When compared to the core gameplay group, the prevalence of the look-up element by itself did not significantly enhance the learning outcome of the game. However, as mentioned above, in combination with the repeat element, this design element created a significantly stronger effect in the in-game and multiple-choice measures. This showed that look-up features should be considered important design elements in learning-oriented gameful applications.

Literature on error management training (Chillarege et al., 2003; Keith & Frese, 2008) provides a potential explanation for the success of the combination of these two design elements. In contrast to errorful and errorless learning, this method (EMT) consists of helping trainees understand why errors occur, indicating how they can be avoided (as afforded by the look-up element), and then applying that knowledge to solve the problem (as afforded by the repeat element). This offers positive indications that if both affordances are implemented at the same time, they could lead to an especially successful learning outcome. This can be further consolidated within the theory of learning styles. In a study conducted by Liu and Reed (1994), which considered affordance combinations in a hypermedia environment, learning was accomplished by offering a diverse set of tools and aides to groups of students with different learning styles. This suggests that offering different optional affordances benefits a diverse group of learners and leads to a stronger overall learning outcome. The combined effect could further assist in preventing the perception of cheating that could come with a help or hint-related design element (Clarebout & Elen, 2008), as it allows players to test their own abilities in the first iteration of a wave before resorting to looking up the correct solution in the repeated wave.

In summary, the results showed that the core gameplay by itself already performed very well in comparison with the non-game materials. However, for the overall game to be more effective, it can be enhanced successfully by the two design elements that we suggested. Particularly, their combination showed their potential as building blocks for successful learning strategies by combining the mnemonic effect of repetition with easily accessible means for understanding.

When analyzing the control variables, we found that the number of years our participants had been living in Germany positively influenced their learning outcome in terms of the in-game and multiple-choice measures. This connection was expected since this particular control variable was implemented to passively enquire about prior waste sorting knowledge (to prevent priming, we decided against a full pre-measure of waste sorting knowledge—see Limitations Section). General waste sorting motivation also proved to have a significant influence over the learning outcome of the in-game measure alone. However, it is difficult to make sense of the fact that this effect was not replicated in the other measures—especially the real-life waste sorting measure. There could be influences in terms of cognitive dissonance of self-belief and self-actualization, but because of the setup of the experiment, we could not derive any personality-based indicators.

6.1.7. Contribution

The central goal of our research is to contribute to the rise of sustainable behavior through gameful design, specifically with regard to waste management. This goal stands in line with point 12.8 of the UN catalog of sustainable goals: "Ensure that people everywhere have the relevant information and awareness for sustainable development and lifestyles in harmony with nature" (United Nations, 2020). Our study showed that gameful design can successfully contribute to better municipal waste sorting, even with regard to a transfer of knowledge to real life. To the best of our knowledge, this is the first study to do so.

We further contributed to the ongoing efforts of investigating the potential of serious games and the implementation of gameful design as powerful teaching devices. In particular, the study showed significant positive learning outcomes within a domain that generally lacks incentives relating to direct personal interest—such as health, or fitness-oriented games would offer—and that is hampered by disinterest or even disgust by their target group regarding the general topic. By successfully translating this into more desirable content matter, our research highlighted the benefits of gameful design for teaching under adverse conditions. In terms of theoretical contribution, by conducting a full assessment of design choices with regard to their different learning outcomes, our research added to the ongoing general efforts of methodically assessing learning through gameplay. In this, our study lined up with a growing amount of research dissipating still-existent doubts about the usefulness of game-based learning (Shute et al., 2009).

A factor that contributes to such doubts is that not all studies in gameful learning test the success of their artifact in connection with its transition to real-life knowledge and applicability (e.g., Kim and Shute, 2015). This measure, however, is very important, as seen in, for instance, Größler et al. (2000), who found in their study on gamification of business simulators that "participants were not capable of accessing the knowledge gained outside the gaming context" (p. 271). Another example is Ball et al. (2002), who concluded that cognitive training may only improve skills that are specific to the trained cognitive domain. Also, Luo et al. (2018) conducted a study with a similar premise and goal to ours and did not manage to reinstate the learning outcome when measured in real life. In contrast, in our study, we found that our game did overall manage to overcome this hurdle. Despite this success, the difficulty of constructing knowledge could be seen in the differences in learning outcomes between the different testing media. Our study highlighted the importance of measuring in the training medium as well as the true context medium (real life) and proving that the transfer is manageable given good design choices (Van Eck et al., 2017).

We also identified a gap in the IS literature on the effectiveness of look-up/help-based design elements and added to the ongoing discussion by conducting an experimental setup that tested this element in an isolated and a combined treatment. Our results showed that affording an optional, learner-moderated look-up element can be a very promising learning-enhancing design element, especially if added to a repetition-based teaching setup. By intricately testing these specific game mechanics, we contributed to understanding how they function to produce meaningful learning experiences, which is a paradigm suggested by the Games, Learning, and Society initiative (Squire, 2007). Regarding the general topic of sustainability in IS, our study was one of few to focus on challenges surrounding the domain of waste management. We hope to inspire further studies in this seminal area of research.

In terms of its practical contribution, we believe that if implemented into the teaching curriculum of sustainability classes, our artifact can have a beneficial impact on the topic of correct waste sorting. Our research aims to support the process of research informing practice and aid designers in optimizing their design decisions, as they have to make efficient decisions under time pressure (Stacey & Nandhakumar, 2009). Furthermore, as stated by Clow (2013), educators need to be given insights about additional tools, as well as their strengths and limitations, which we provide in this manuscript. By affording detailed insights into the rationales behind the design decisions that went into the creation of our game and the design elements used, we facilitated easy means of reproduction for practitioners and researchers. While the mechanisms we looked at are embedded in the framework of a game, any learning or training context can serve as the foundation for the design mechanisms we analyzed in our study (Deterding, 2016). Thus, we argued that in a playful setting that allows a certain degree of make-believe, a wide variety of teaching tasks (e.g., vocabulary, geography training, digital management training, and onboarding) could benefit from applying the findings of our study.

6.1.8. Limitations and Future Work

One potential limitation concerns the fact that we omitted the assessment of prior knowledge on waste sorting. Due to the three-phase setup of the experiment, we consciously decided against this assessment because of concerns about priming the participants and thus skewing the results. While it is common in the assessment of serious games to use pre-and post-testing, "the main problem with the pre-and post-test experimental design is that it is impossible to determine whether the act of pre-testing has influenced any of the results." (Bellotti et al., 2013 p.3). By conducting a prior assessment like completing a survey-based multiple-choice test, we were concerned that participants would influence the actual results by looking up certain items they were unsure of before the first task. Instead, we measured "living in Germany" as a proxy indicator for prior knowledge, which turned out to have a significant influence on the learning outcome.

Because we conducted an experiment by randomly assigning participants to treatments, we trust the internal validity of our results. Thus, the effect should be independent of confounders such as prior knowledge.

We believe the exclusion of prior knowledge as a predictor in our models, as well as the omission of measuring other variables that might influence waste sorting knowledge—e.g., participants' exposure to the topic in school or other contexts or their families' attitudes towards sustainability and eco-friendliness—are the main reasons for the rather low R^2 of our main models that included only the treatment variables. However, a low R^2 is not unusual for experimental research and does not harm the interpretation of the effect of the treatment variables. Our further analyses in Appendix C.2.5 also show that the inclusion of the control variables—e.g., the number of years that the participants have lived in Germany and general waste sorting motivation—helps reduce the unexplained variance substantially, yielding R^2 values around 0.2.

We further see that there could be an underlying cultural bias given the generally high range of results. Frese et al. (1991, p. 90) noted that errors may be perceived as especially stressful in German culture, "where perfectionism is highly valued." For transferability to other cultures with different prior mentalities regarding correct waste sorting, future studies will be necessary to assess mentality as a moderating factor. Another important point is that we assessed the real-life measure with only seven items. While this arguably weakens comparability with the other two measures, practical concerns in terms of implementing a much larger number of items (limited setup and timeframe, participants' resistance to interacting with certain items) limited our options for this measure. It should also be noted that within the non-game materials used during the training phase, one of the three flyers (the one showing examples of waste items for each bin in Figure 96 in Appendix C.2.2) featured more items than were presented to the game groups during the experimental task. When we designed the experiment, we wanted to approximate a real-life scenario and thus chose to use an unabridged set of standard materials provided by the local waste management (see Appendix C.2.2). In hindsight, the overall experimental design would have been cleaner if we had reworked the flyer to feature the exact number of items that were trained in the game. However, it is important to note that the goal of the game was not to teach the specific relationship of each featured waste item to their bin but to help players understand and train them on the rules of the underlying waste systems. As all objects will eventually turn into waste items, citizens need to learn how to correctly sort any object they encounter into the respective system by understanding and internalizing the underlying principles.

We wanted to test the learning outcome in a rigorous and controlled manner to obtain clear and interpretable results, so we decided to conduct a laboratory experiment to provide high internal validity. However, as our findings were based on an experimental setting including mostly students, any gained insights were only applicable to the tested age group (17-41). A future step would be to test whether the effects found are replicable in the field. Another facet of this relates to knowledge transfer. Even though we found that knowledge transfer to real-life (construction) was successfully achieved in the game, we believe that this effect might be enhanced by transferring the game to a virtual/augmented reality environment by bringing the medium of training closer to the actual application context. Finally, while we chose to separate learning from motivation to isolate our findings, this approach might have omitted an important influence on learning. On this basis and because of our overall goal to teach correct waste sorting and to boost the motivation to act upon that knowledge, we want to design and conduct another motivation-focused experiment to build on our findings and enhance gameful design-based learning even further.

6.2. Transfer to Ontological Research & Contributions

In the deep dive studies presented in the previous two chapters, we looked at specific game design elements (particularly a *perfect reward*, *repetition,* and *look-up*) with regards to their success in influencing outcome parameters of interest (like in our case *performance* and *learning effect*). Apart from gaining insights into their effectivity through experimental design and field analysis, we conducted this research to test and substantiate the foundations of our game design element ontology as well as the design of the navigational tool, Kubun. By turning our perspective from the top-down designer view to the hands-on researcher view, we gained first-hand insights into the needs as well as the difficulties of implementing these new game design elements and research outcomes into the existing structure. In the following, we describe the steps we took to implement the findings of our research.

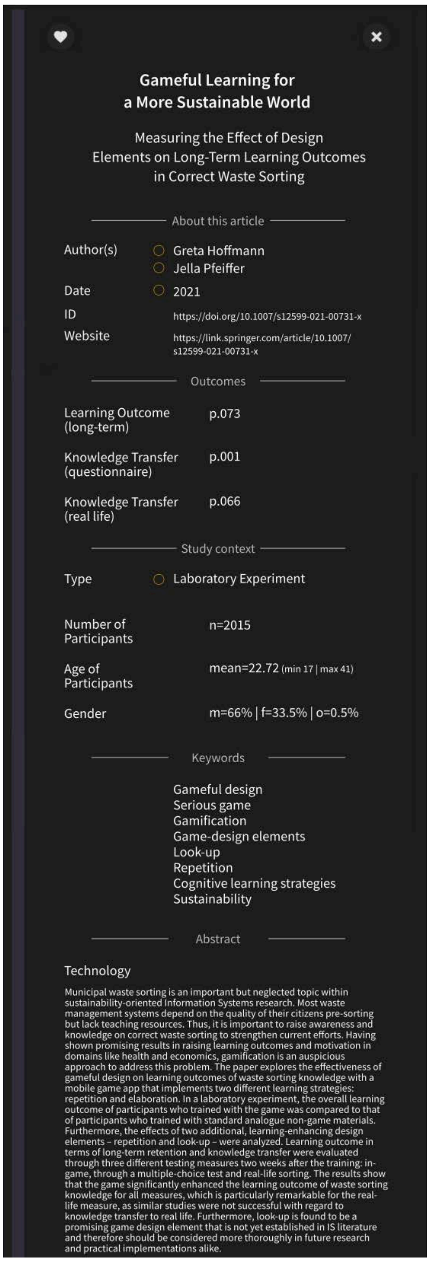


Figure 74 – Design of a Research Report Oriented Schema

As pointed out in the introduction of our research, one major deficit of current collections of game design elements and patterns (like the game design patterns collection by Björk and Holopainen (2005) or the game element classification by Elverdam and Aarseth (2007) concerns the lack of links to current research and conducted studies on respective elements. We concluded that details on scientific findings and their related meta-information on game design elements would help users of the ontology to make more informed decisions. As the ontology presents results to its users through a similarity algorithm, using such information would achieve better search results. In light of these insights, we designed a new schema for research reports, where the specifics of a study can be reported. The choice to design an additional schema was made after we first started to add a complex research module to the original ontology. During the implementation process, we realized that the information we were trying to include was not inherent to the abstract entity of the game design element itself but to the study of the element. Thus, we decided to separate these entities and design a structure that links the relevant data to the abstract entity (Game Design Element Kubun) while keeping the specific data with the specific study (Studies on Game Design Elements Kubun). As such, if several studies report on similar measures for the same game design element, aggregated data could be presented in the parameters module of the game design element Kubun instead of a list of different study outcomes. We designed a new schema for a research-report-oriented Kubun to include the following submodules (see Figure 74):

1) A *study* module ("about this article") presenting core metadata on the study (title, authors, date, DOI/ISSN/ISBN, and publishing website).

2) A *parameters* module that presents the evaluated parameters and the found outcomes (no effect, strong positive effect, strong negative effect). Values can be reported; however, they cannot, as such, be used for the similarity algorithm as such values might not be comparable between studies depending on the experimental setup.
3) A *context* module where the study type is reported as well as the most substantial information on the study's participants (number, age, gender)
4) A *keywords* module that includes all essential keywords that relate to the study's content and topic of interest.
5) A *text* module shows the abstract.

In summary, by conducting in-depth research studies, we gained a better understanding of how the ontology, as well as the navigational tool, should be enhanced and modified to enhance their usability and depth to better serve researchers. The results of the in-depth research highlighted the complexity of reliably attributing expectable outcomes to the implementation of different game design elements as their effect does not only vary between different player types (that may or may not be differentiable by factors of personality) but is also influenced by other game design elements that actively or passively interact with them. Given the complexity of reporting research outcomes and context, we gained an understanding of the limitations as well as the of our tool in how and where information should be presented.

Part IV

Finale

7. Chapter 7
Conclusion, Limitations, and Future Research

"This world is of a single piece; yet, we invent nets to trap it for our inspection.
Then we mistake our nets for the reality of the piece.
In these nets we catch the fishes of the intellect but the
sea of wholeness forever eludes our grasp."

Martha Boles, Universal Patterns

7.1. Summary and Conclusion

In this interdisciplinary dissertation, I set out to design and develop a functional, digital ontology of game design elements, serviceable to scholars as well as practitioners, building on best practices of IS research methodologies and human-centered design. The process was divided into two different overarching sections. In the first section, I use an explorative top-down approach for the aggregation and enrichment of the ontology and the development of an intuitive tool for its navigation. Therein, RQ1 and RQ2 address matters of classification and labeling in terms of the derived ontology, and RQ3 addresses the connection of related datasets. Matters of the design, development, and evaluation of the digital artifact for navigating the ontological data are addressed by RQ4. The second section builds on a quantitative, bottom-up approach for integrating learnings into the ontology from the researcher's perspective when conducting in-depth game design research. For this, I conducted a set of experiments on the outcomes of specific design elements, which were structured through RQ5 and RQ6. These overarching research questions were motivated in Chapter 1 and answered through the studies presented throughout this work. The following paragraph is dedicated to summarizing the main findings and presenting the conclusions I draw from this research.

RQ1. *How compatible are established frameworks from game design theory with a dataset of diversely aggregated game design elements?*

After having collated a dataset of game design elements, I aimed to enrich the ontology through user-relevant metadata. While there are several well-established frameworks in game design literature, they all pertain to different directions in terms of their categorization perspective. With the aim of finding a framework fitting the needs of the game design element dataset, I addressed RQ1 by conducting a closed card sort study based on a process outlined by Rugg and McGeorge (2005) with a set of four representative frameworks from game design theory. In this study, sixteen participants conducted two card sorts, respectively, with two sets of cards. I analyzed the resulting taxonomies to determine the overall suitability of the given frameworks to the dataset in terms of three factors: fitness (how many cards were sorted into the preexisting categories), commonality (how often were the same cards sorted into the same category) and consistency (did the frameworks perform differently between two different presets of cards). Through the analysis, I found two frameworks (MDA by Hunicke et al. (2004) and VAC by Robinson and Bellotti (2013)) that performed better in all of these three factors than the others. However, the overall commonality of sorting items into the same category was relatively low, performing under a threshold of .25 for a two-card commonality for even the best-performing category (VAC). As such, I find that the evaluated frameworks are not very compatible with the game design element dataset. Thus, this study highlighted the complex nature of the game design elements underlying the dataset, as well as the differences in users' understanding of both the dataset entries and the categories.

RQ2. *What categorial viewpoints of game design elements can be identified from an expert perspective?*

Given the volatile results with regard to the preexisting categories, I aimed to inquire further, potentially more suitable categorization viewpoints. To address RQ2, I conducted a second study using an open card sort (see Righi et al. (2013)) to collate and assess viewpoints that users might have outside of established frameworks (see Chapter 3.3). In this study, eight experts sorted cards into a category structure they derived by themselves through the sorting process. Overall, four different types of overarching categorical viewpoints emerged: Bottom-Up, Top-Down, SAM (story, aesthetics, mechanics)/MDA (mechanics, dynamics, aesthetics), and a Design Guide perspective. While oftentimes arriving at similar categories in deeper levels of the hierarchy, the differences in the overarching perspectives the experts chose for establishing a structure for the exemplary datasets of game design elements underlined the complexity of arriving at a singular classification in terms of the dimensions underlying the ontology. In summary, both card sort studies offered valuable insights into the nature of the data and afforded useful labels for the dataset excerpts. However, given the varying labels and perspectives, as well as the fact that the open card sort took most participants more than two hours to complete instead of the one that was scheduled, the outcomes of both studies highlighted the need to explore alternative methods for dataset classification.

RQ3. *How meaningfully can datasets be connected through algorithmic keyword-based matching?*

In Chapter 3.4, I explore an automated process for dataset enrichment through an explorative method of linking datasets via their matching keywords, inspired by a method used in search engine optimizations (Devanur & Hayes, 2009; Uthayan & Anandha Mala, 2015). In a first step, I applied this approach to playing motivations, where I compare items within the set to each other, exploring the method's effectivity in identifying similar or fitting entries. I found eight overarching meaningful clusters of playing motivations: Social Factors (Interacting), Explorative Factors (Exploring), Achievement Factors (Achieving), Escapism Factors (Immersing), Creative Factors (Creating), and Emotional Factors (Feeling). An explorative analysis of the resulting graph showed high potential in terms of similarity, cluster detection, and outlier analysis. Building on the promising explorations of the first matching process, I conducted a second study matching a separate but related dataset of human needs to the playing motivations dataset. The analysis finds an overlap of 89% between the two datasets, highlighting the strong relationship between human needs and motivations to play. Through qualitative evaluation, I further find strong congruence between the linked items, underlining the method's viability in finding meaningful matches. In an explorative clustering built on the combined dataset, I sorted unmatched keywords from the human needs dataset into the formerly established cluster structure. Through this process, I managed to identify additional clusters of potential motivations, for example: Luxury/Possessions, Values, Self-Exploration, Anti, and a special cluster around Fitness – a topic more prevalent in the domain of health games (Adams, 2010). I conclude that the method of keyword matching leads to meaningful connections between thematically fitting datasets and serves as a serviceable foundation for building an interconnected ontology of game design elements enriched with related playing motivations and human needs.

RQ4. *What design factors afford user-friendly browsing processes for database navigation?*

Having built an ontological foundation by labeling and connecting datasets of game design elements and playing motivations, I next focused on addressing the second overarching goal of my research, developing an intuitive and effective tool for navigating ontological data. I address RQ4 by documenting the design process of the digital artifact used for navigating the ontology that I constructed in the studies reported in Chapter 3. Following best practices from human-centered design (Cooley, 2000), I started the process by aggregating foundational requirements and conducting a market analysis of existing tools. Building on these insights, I derived three design components crucial to enhancing existing methods in terms of user-friendliness and efficiency. First, a *map-inspired feature* that builds on eased information processing through visual representation of results and the affordance of meaningful spatial information. Second, a *reference-based search* is inspired by real-life search processes, where an already known item of reference forms the starting point of the browsing, from which similar results are calculated and presented. And third, a *weighting interaction* (in opposite to classic filter-based operations), where user-selected parameters are added to the similarity calculation with a higher weight. The working demonstrator went through an iterative process of design and development, resulting in the final digital artifact. Its usability was ensured and enhanced via a three-stage set of systematically conducted user tests based on semi-structured interviews, further iterating the artifact between each round. Further insights were gained through the evaluation of field study data gained by promoting the online version of the demonstrator in online forums. The results of both evaluations were very positive, especially concerning the usability and usefulness of the tool. The field study resulted in a return rate of 20% and even spawned unprompted offers of collaboration. I conclude that, given the positive evaluation and feedback, the aforementioned design factors were successful in affording user-friendly navigation of ontological data.

RQ5. *What effect does a reward for perfection have on playing performance?*

While Part II of the dissertation focuses on the design and development of a digital game design element ontology from a top-down perspective, part two of the dissertation focuses on a bottom-up approach via in-depth research on specific game design elements. Under an overarching motive of optimizing learning processes through gameful design, in Chapter 5, I evaluated the effect of a reward element for incentivizing perfect play. I start this part of the research through a pre-study where I evaluate field data of a game artifact, thereby identifying two distinct playing behaviors: Perfectionists and Rushers. Building on these insights, I conducted a laboratory experiment to investigate potential coherence between different personality types and the identified behaviors, particularly with a focus on perfect play. The results of the experiment were inconclusive, showing no significant relationships between the tested personality types and in-game behavior. Furthermore, the control group that was not given the reward element did not show lower performance than the treatment group. I found that rewarding perfect play the way I did with the game design element does not result in higher levels of perfect play. Through qualitative analysis of the free-text answers in the experiment, a tendency emerged where users reported a threshold of an 80-90% success rate as satisfactory. I conclude that stronger game design element-based incentives would be necessary to overcome this threshold and incentivize perfect playthroughs. However, given that certain types of perfectionism are classified as compulsive behavior (Hamachek, 1978), they might be inherent to certain personality types while missing in others. As such, it might also not be advisable to encourage such behavior. Given that learning can effectively occur outside of perfect results, I conceded to focus on more effective learning elements in future studies.

RQ6. *What effect does a repetition-based and a look-up-based game design element have on long-term learning of correct sorting of waste items into their target bin?*

Focusing on learning outcome, in Chapter 6, I evaluate two further game design elements (repeat and look-up) designed to support the learning goals of the game artifact they are embedded in. I explore RQ5 through a between-subjects experiment where I evaluate long-term learning outcome through three different testing measures (multiple choice test, in-game, and real-life sorting) to assess potential variations of knowledge transfer. I find that each of the two elements in isolation enhances learning outcome compared to training without a game. Furthermore, a combined implementation of both repeat and look-up surpasses all other game design constellations tested in the experiment in two of the three measures. In accordance with similar findings drawn from literature on error management training (Chillarege et al., 2003; Keith & Frese, 2008) and instructional design (Nitsch et al., 2016; Wittwer & Renkl, 2008), I conclude that allowing for repeating a task and affording a correct answer and explanation affords effective long-term learning outcomes. Overall, the learning outcome for the groups given the game for training was significantly stronger than for the group given state-of-the-art paper-based information during the training phase. This is particularly relevant in terms of the successful transfer of in-game to real-life knowledge. The difficulty of this transfer is highlighted by the fact that all treatment groups performed weakest in the real-life measure. As such, I plan to set a stronger focus on this facet when conducting future research.

I transferred the learnings from part two of the dissertation by integrating the elements researched in Chapters 5 and 6 into the game design element ontology. I further used the insights into the research and documentation process of the in-depth research to design a research outcome-focused module, thus enhancing the schema design of the database navigation tool (see Chapter 6.2).

7.2. Limitations and Future Research

Given the explorative nature of my research, I want to address some of the limitations of this work. First, when looking at the datasets underlying the ontology, I must note that the extracted elements only serve as representative samples, limited by the sources I chose for inclusion. Also, particularly in terms of the game design elements, given the broad definition I chose for including items, the resulting dataset has proven to be difficult to classify. Its elements vary widely in how well they fit the different classification perspectives I evaluated and identified in Chapter 3.3.3. In terms of future development of the game design element dataset, I suggest breaking down the current structure into several datasets, thus building smaller structures on more localized definitions and schemas that can then be connected through an enhanced matching algorithm. Second, this in-depth research highlighted several pending open problems with regard to practical implementations of research findings. Looking at the game design elements through the lens of experimental analysis, I found interactional, combined effects for the repetition and look-up element. This brought forth questions regarding the inclusion of such combined elements, the reporting of interactional effects, and how the platform could be enhanced to best report on both. The current design technically affords different solutions. It is possible to add a new, combined design element and report the effects as if it was a new entity. Alternatively, the combined effect could be reported in each of the singular elements. However, both solutions are unsatisfactory as they fail to highlight the underlying connections intuitively as well as the emergent benefits. This problem is indicative of another, more general limitation towards the undertaking of creating an objective game design element ontology. While game design elements can be listed individually, they never exist without context. Many of the listed elements already consist of smaller elements, as it is difficult to pinpoint when a combination of design decisions emerges as an element and when a collection of such elements arises as a pattern. While I was not able to address these limitations in this work, it is my goal for future research and design to bring both the ontology as well as the navigational tool towards a stage in which the complexities of interconnected design and research can be addressed satisfactorily and intuitively.

In the following, I want to expand on a few features that emerged throughout the overall process that should be addressed in the future of the project.

First, the issue of classification remains yet to be addressed satisfactorily. Given the problems that emerged in terms of the time and resources when conducting the expert-based card sorts, a divide and conquer strategy could be applied to solve this issue. Crowdsourcing (an amalgamation of the terms crowd and outsourcing (Schenk & Guittard, 2011)) is an effective and powerful practice that can be implemented to address any non-trivial problem, from routine to complicated cognitive tasks (Kleemann et al., 2008) through creative tasks or those related to innovation (Reichwald & Piller, 2006) as long as the task is divisible into lower-level tasks that can be accomplished by each individual member of the crowd (Estellés-Arolas & González-Ladrón-de-Guevara, 2012). Building on the existing functionalities of the card sorting tool "Tako" (Hoffmann and Martin 2018, see Appendix A.2.1), a gamified app could be used to break the card sorting task down into daily tasks of singular cards distributed over all users. Furthermore, an "Edit Mode" should be added to the database navigation tool Kubun. Expanding the tool to serve a general purpose, the design artifact should afford users to create their own ontologies by uploading and publicizing their own datasets. In the long run, such additions could build a foundation where users are afforded to improve and enhance existing datasets by adding missing nodes, suggesting new parameters, suggesting alternate information arrangements within the schema and adding or changing characteristics.

Acknowledging growing problems in crowd-sourced information systems with regards to information factuality (Graves & Cherubini, 2016), accompanying the edit mode, the tool should be enhanced to afford transparency by offering a context-relevant, entity-assigned change history per dataset, node, and characteristic. It should further afford single entity authentication and a process for appointing responsible bodies (moderators) for each dataset.

To enhance its usefulness in terms of the adjustment of search input to the needs and desires of the user, the tool should include the possibility to "mutate" an existing node from a representative starting point toward an idealized version of itself (see Figure 75). For example, if a user loves detective stories as well as medieval settings, they could start by mutating the node representing the "Sherlock Holmes" books (Doyle, 1992) in a book-related dataset to be set in medieval times instead of the late Victorian era. As such, books like "The Name of the Rose" (Eco, 1983) should then appear as similar.

For users that have no reference in mind but want to construct their dream novel from scratch in the hopes of finding similar results, a function could be added that allows creating a node within the existing schema but with user-generated input (phantom node). This node does not exist in the original dataset but will be treated by the algorithm of the tool as if it existed, pulling similar results from the existing files. This function would be very beneficial as it would not only afford users to find results that are closest to their current needs but also afford the dataset owners information on user consumption intention through data mining of user interests in latent space.

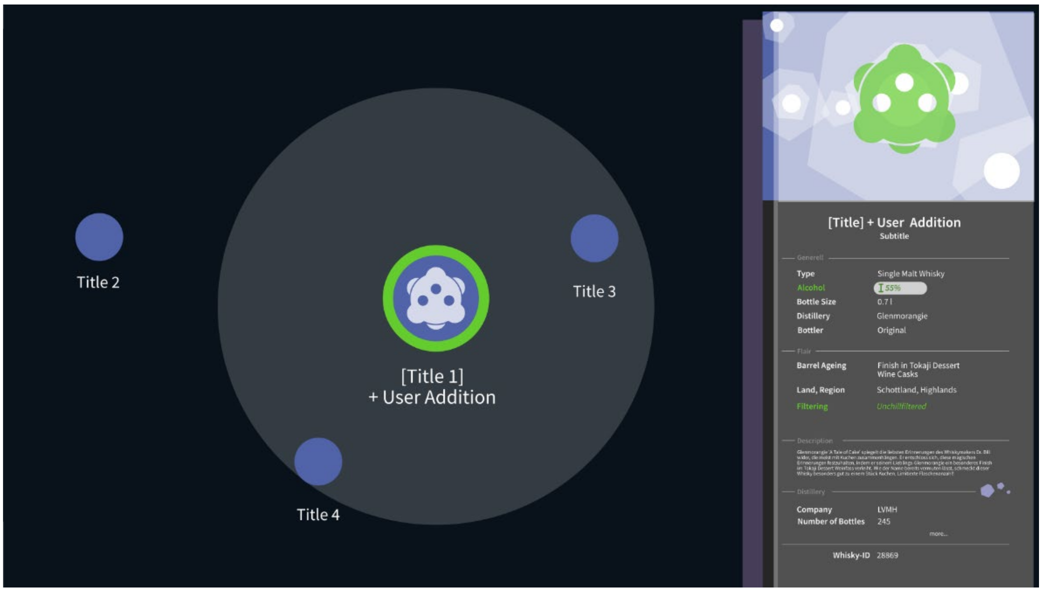


Figure 75 – Example Design for the Mutation Mode

Such data mining is currently oftentimes conducted through machine learning algorithms in social media (Ding et al., 2015) and is considered valuable for product recommendations and targeted advertising. It can also be used for learning purposes and to understand users' pain points and thus afford product iteration and innovation (Chen et al., 2021). By collecting and mining user queries that don't yield initial results, Kubun could be used to identify meaningful gaps in existing datasets. Furthermore, by analyzing parameter-based change requests in the mutation mode, it would be possible to extract user needs currently not reflected within the dataset. Given this function, the tool would gain the potential to directly serve users while equally providing content creators with relevant information regarding the market potentials of their product or media.

In terms of future avenues of research, I see several points of contact within our research. While the tool itself was developed with game design elements in mind, its core functionality as a database navigation tool can be applied to many contexts. In this regard, it would be interesting to assess what types of datasets can particularly benefit from the new interaction approach and identify what types of functionality are yet missing. Also, building on new developments in the area of machine learning, further research could be aimed at the development and assessment of methods by which further metadata can be derived and integrated automatically and how entries in the overarching ontological cosmos can be meaningfully connected. Given the positive feedback in terms of usability enhancement of the new browsing functions integrated into Kubun, it would be interesting to assess these features for usability enhancement of related application contexts (websites and tools that feature large amounts of similar data).

In terms of the evaluation of specific game design elements, particularly the outcome of the second in-depth experiment (assessing the single and combined effects of a repeat and look-up design element) highlights the potential variations in outcome given a combined effect. As such, given that almost no design decision is made in an isolated context, particularly systematic game design element assessment could be enhanced by including reports on interaction effects with related functionalities.

7.3. The End

In this dissertation, I started with the goal of developing a functional ontology for game design elements. I tackled this goal through different avenues, informing research through foundational literature on related domains of research, namely game design, human-computer interaction, and cognitive science. I built a foundational dataset by extracting game design elements from literature and expanded it through different explorative methods that allowed us to enrich the data but also led us to be confronted with the limitations of my research approach. By further conducting in-depth research on specific design elements, I was able to enhance my understanding of the user's perspective while getting confronted with further challenges regarding the complexity of the content matter. The overall process of this work allowed us to gain a better understanding of the research goal I originally set out to achieve (*to design and develop a digital, interactive ontology that contributes to the current research efforts by extending them with user-friendly affordances for easier navigation*). While the process of designing a functional game design element ontology opened more questions and problems to be tackled than I believe I could manage to answer, the resulting artifact transcended my initial expectations in terms of repurposing potential.

The main contribution of this work lies in the digital ontology itself and the insights I could derive through the process of developing a useful and useable application for the navigation of complex, multidimensional, ontological data via an iterative, explorative process building on research methods from IS. By exploring different avenues of research and employing suitable methodologies, I arrive at an ontological artifact featuring extensive data with relevant metadata to be explored and implemented by researchers and practitioners equally. I reinforce this perspective of researchers' needs for the information provided by a game design element ontology through a series of in-depth studies where I evaluate game design elements in terms of their outcomes on different parameters. Through evaluating needs that arose through the process of the in-depth research, I arrived at new requirements to be implemented towards the underlying logic and structure of the artifact.

From the perspective of a designer venturing into the domain of IS research, despite certain difficulties in terms of transfer – particularly regarding the integration of different ways of thinking and differences in addressing problems, the outcome of this interdisciplinary process has shown to benefit both IS research and design practice. First, in terms of benefit for both disciplines, the artifact in its current manifestation could not have come into existence without combining established research methodologies with best practices from design (Norman, 2013) and human-computer interaction (Hassenzahl, 2010). Second, in terms of contribution to the IS community, the difference in the research approaches of these disciplines allowed for the exploration of new methodologies like the experimental keyword matching presented in Chapter 3.3.4. Furthermore, in terms of contribution to the domain of design, the rigorous approach I chose for the in-depth evaluation of specific design decisions as presented in Chapters 5 and 6 can serve as a demonstrator for the benefits of integrating systematic research into the design and development process. Such processes can afford the detection of unexpected outcomes, for example, the finding that the perfect reward element did not incentivize perfect playing behavior (see Chapter 5.5) and can further be used to quantify differences in training and testing medium, as I found with the differences in knowledge transfer to real-life (see Chapter 6).

The ontology that was developed through the overall process, as well as the tool that was created to navigate it, contribute to research and practice as a method and tool that can serve as an effective starting point for finding relevant design enhancements, facilitate research processes by identifying connections and metadata relevant to the users' problems and offer interesting avenues for future research. With this, I want to close with the following quote from Aristotle: "For the things we have to learn before we can do them, we learn by doing them." (Aristotle, n. d.)

Part V

Appendix

Appendix – A

Supplementary Material Chapter 3

A.1 Supplementary Information Literature Review Gamification

A.1.1 Literature Review – Relevant Studies

Table 26 – Overview Gamification Studies with Domain, Country, Sample Size, and Demographic Data

Ref. ID	Title	Author	Journal	Year	Country	Domain	Duration	Sample Size /Demographics
10.1016/j.chb.2015.04.018	"Working out for likes": An empirical study on social influence in exercise gamification	Juho Hamari, Jonna Koivisto	Computers in Human Behavior 50 (2015) 333–347	2015	Finland	fitness	n.a.	200
10.1002/cae.21992	A comparative study on gamification of the flipped classroom in engineering education to enhance the effects of learning	Jo Jaechoon; Jun Heeyeon, Lim Heuiseok	Computer Applications in Engineering Education, 26(5):1626-1640, 2018	2018	Korea	education	Seven weeks	30
10.1016/j.compedu.2014.08.019	Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance	Michael D. Hanus, Jesse Fox	Computers & Education 80 (2015) 152-161	2015	USA	education	Four surveys in a 16-week semester	80 participants; 57 males, 23 females
10.1145/2047196.2047248	Calibration Games: Making Calibration Tasks Enjoyable by Adding Motivating Game Elements	David R. Flatla, Carl Gutwin, Lennart E. Nacke, Scott Bateman, Regan L. Mandryk	UIST'11, October 16–19, 2011, Santa Barbara, CA, USA	2011	Canada	consumer motivation	one session??	36 (3 studies, 12 per study)
10.1016/j.chb.2015.07.045	Challenging games help students learn: An empirical study on engagement, flow, and immersion in game-based learning	Juho Hamari, David J. Shernoff b, Elizabeth Rowe, Brianno Coller, Jodi Asbell-Clarke, Teon Edwards	Computers in Human Behavior 54 (2016) 170-179	2015	USA	education	15-week	173
ISSN 1479-439X	Climbing Up the Leaderboard: An Empirical Study of Applying Gamification Techniques to a Computer Programming Class	Panagiotis Fotaris	Electronic Journal of e-Learning, 2016, Vol.14(2), p.94-110	2016	UK/Greece	education	Two years	106

10.3991/ijet.v10i1.4247	Cogent: A Case Study of Meaningful Gamification in Education with Virtual Currency	Y. Chen, T. Burton, V. Mihaela, D.M. Whittinghill	International Journal of Emerging Technologies in Learning (iJET), 01 February 2015, Vol.10(1), pp.39-45	2015	USA	education	Two days	32
10.1016/j.chb.2014.03.007	Demographic differences in perceived benefits from gamification	Jonna Koivisto, Juho Hamari	Computers in Human Behavior 35 (2014) 179–188	2014	Finland	health	*undefined*	195
10.3389/fpubh.2014.00042	Development and implementation of a smartphone application to promote physical activity and reduce screen-time in adolescent boys	David R. Lubans, Jordan J. Smith, Geoff Skinner, Philip J. Morgan	Front. Public Health, 20 May 2014	2014	Australia	health	20 weeks	361
10.1016/j.chb.2015.03.036	Do badges increase user activity? A field experiment on the effects of gamification	Juho Hamari	Computers in Human Behavior 71 (2017) 469-478	2015	Finland	economy	Two years each group 1 year: gamified/non-gamified condition	non-gamified condition: 1410 gamified condition: 1579
10.13140/RG.2.1.4783.5686	Does Gamification Work for Boys and Girls? An Exploratory Study with a Virtual Learning Environment	Lais Zagatti Pedro, Bruno Genova Prates, Aparecida Maria Zem-Lopes, Julita Vassileva	SAC'15, April 13-17, 2015	2015	Brazil	education	One session	16
ISBN: 978-1-4673-2936-1	Encouraging User Behaviour with Achievements: An Empirical Study	Scott Grant, Buddy Betts	Proceeding MSR '13 Proceedings of the 10th Working Conference on Mining Software Repositories Pages 65-68	2013	USA	online participation	Four months	1295620
10.1016/j.chb.2010.08.005	Encouraging user participation in a course recommender system: An impact on user behavior	Rosta Farzan, Peter Brusilovsky	Computers in Human Behavior 27 (2011) 276–284	2011	USA	online participation	Three years	171
10.1016/j.compedu.2015.10.010	Engaging Asian students through game mechanics: Findings from two experiment studies	Khe Foon Hew, Biyun Huang, Kai Wah Samuel Chu, Dickson K.W. Chiu	Computers & Education 92-93 (2016) 221-236	2016	China	education	18 days	64
10.1016/j.compedu.2018.01.009	Enhancing student learning experience with technology-mediated gamification: An empirical study	Crystal Han-Huei Tsay; Alexander Kofinas; Jing Luo	Computers&Education, Vol. 121, June 2018, pp. 1-17	2018	UK	education	Two academic terms (24 weeks)	136
10.1145/1658866.1658873	Evaluation of a Pervasive Game for Domestic Energy Engagement Among Teenagers	Anton Gustafsson, Cecilia Katzeff	ACM Comput. Entertain. 7, 4, Article 54 (December 2009)	2009	Sweden	sustainability	ten days	Six households
10.1145/2856767.2856790	Expense Control: A Gamified, Semi-Automated, Crowd-Based Approach For Receipt Capturing	Maximilian Altmeyer, Pascal Lessel, Antonio Kruger	IUI '16 Proceedings of the 21st International Conference on Intelligent User Interfaces Pages 31-42	2016	Germany	online participation	Three weeks	12

10.2196/games.5678	Game On? Smoking Cessation Through the Gamification of mHealth: A Longitudinal Qualitative Study	Abdulrahman Abdulla El-Hilly, Sheeraz Syed Iqbal, Maroof Ahmed, Yusuf Sherwani, Mohammed Muntasir, Sarim Siddiqui, Zaid Al-Fagih, Omar Usmani, Andreas B Eisingerich	Jmir Serious Games, 2016 Jul-Dec, Vol.4(2)	2016	UK	health	Five weeks	16 smokers in 2 cohorts
10.1007/978-3-319-39399-5_14	Gamification Aspects in the Context of Electronic Government and Education: A Case Study	Fernando Timoteo Fernandes, Plinio Thomaz Aquino Junior	F.F.-H. Nah and C.-H. Tan (Eds.): HCIBGO 2016, Part II, LNCS 9752, pp. 140–150, 2016.	2016	Switzerland	education	15 days	26
ISSN: 1535-0975	Gamification from Player Type Perspective: A Case Study	Selay Arkün Kocadere, Seyma Caglar	Educational Technology and Society, 2018, Vol.21(3), pp.12-22	2017	Turkey	education	One course	41
10.1016/j.procir.2014.07.056	Gamification in factory management education – a case study with Lego Mindstorms	Bastian C. Müller, Carsten Reise, Günther Seliger	Procedia Cirp 26 (2015): 121-126.		Germany /Vietnam	working environment	One term	n.a.
10.24251/HICSS.2018.149	Gamification of Older Adults' Physical Activity: An Eight-Week Study	Kappen, Dennis; Mirza-Babaei, Pejman; Nacke, Lennart	Proceedings of the 51st Hawaii International Conference on System Sciences, 2018, pp. 1207-1216	2018	Canada	fitness	Eight weeks	30 adults (over 50 y.o.)
10.1016/j.chb.2016.04.035	Gamification: A framework for designing software in e-banking	Luís Filipe Rodrigues, Carlos J. Costa, Abílio Oliveira	Computers in Human Behavior 62 (2016) 620-634	2016	Portugal	economy	one trial	53
0031-9120/16/055007	Gamification: using elements of video games to improve engagement in an undergraduate physics class	J A Rose, J M O'Mear, T C Gerhardt, M Williams	Physics Education 51.5 (2016): 055007	2016	Canada	education	Two semesters	ca. 800
	Gamified Goals: an Empirical Study of Online Weight-Loss Challenges	Behnaz Bojd; Young Tan, Xiaoling Song; Xiangbin Yan	Thirty-Ninth International Conference on Information Systems, San Francisco, 2018	2018	USA/China	fitness	Eight months	4208
10.1016/j.procs.2017.08.017	Gamified Learning: A role-playing approach to increase student in-class motivation	Alexandru Topîrceanu	Procedia Computer Science 112, (2017), 41-50	2017	Romania	education	3 years	190
10.1016/j.chb.2015.08.025	How gamification motivates visits and engagement for online academic dissemination - An empirical study	Ming-Shiou Kuo, Tsung-Yen Chuang	Computers in Human Behavior 55 (2016) 16-27	2016	Taiwan	online participation	Ten months	73
10.1016/j.chb.2016.12.033	How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction	Michael Sailer, Jan Ulrich Hense, Sarah Katharina Mayr, Heinz Mandl	Computers in Human Behavior 69 (2017) 371e380	2017	Germany	psychology	n.a.	419
10.3390/ijerph15092027	Improving Sexual Health Education Programs for Adolescent Students through Game-Based Learning and Gamification	Hussein Haruna, Xiao Hu, Samuel Kai Wah Chu, Robin R. Mellecker, Goodluck Gabriel, Patrick Siril Ndekao	Open Access Int. J. Environ. Res. Public Health 2018, 15(9), 2027;	2018	China, Tanzania	education	five-week	120 lower secondary school

								students (11–15 y.o.)
https://hdl.handle.net/10652/4332	Improving Students' Performace through Gamification: A User Study	Natalia Nehring, Nilufar Baghaei, Nilufar Baghaei, Simon Dacey, Simon Dacey	CSEDU 2018	2018	New Zealand	education	11 weeks	180
DiVA.org:umu-122624	Internal motivation in Gamification - An intervention study in an exercise context	David Nedergård	Informatik Student Paper Master (INFSPM) ; 2016.18	2016	Sweden	fitness	30 days	10
10.1007/978-3-319-71940-5_9	Investigating Motivation in Gamification - Results from an Experimental Pilot Study	Peter Bußwolder, Andreas Gebhardt	GALA 2017, LNCS 10653, pp. 95–104	2017	Germany	surveys, sociology	30 min	55
10.1016/j.procs.2012.10.059	iThink : A game-based approach towards improving collaboration and participation in requirement elicitation	Joao Fernandes, Diogo Duarte, Claudia Ribeiro, Carla Farinha, Joao Madeiras Pereira, Miguel Mira da Silva	Procedia Computer Science 15 (2012) 66 – 77	2012	Portugal	working environment	One day	17
10.17083/ijsg.v4i4.192	Metrics Feedback Cycle: measuring and improving user engagement in gamified eLearning systems	Adam Atkins, Vanissa Wanick, Gary Wills	International Journal of Serious Games Volume 4, Issue 4, December 2017	2017	UK	education	one session (about one h)	36
10.2501 /IJMR-54-5-613-633	Myths and realities of respondent engagement in online surveys	Theo Downes-Le Guin, Reg Baker, Joanne Mechling, and Erica Ruyle	International Journal of Market Research Vol. 54 Issue 5; 613-633	2012	USA	surveys, sociology	Eight days	1007
10.1186/s41239-018-0107-0	Online learning readiness and attitudes towards gaming in gamified online learning – a mixed-methods case study	Bovermann, K., Weidlich, J. & Bastiaens	T. Int J Educ Technol High Educ (2018) 15: 27.	2018	Germany	education	Three weeks	62
10.3991/ijet.v13i02.7467	Perceptions of Students for Gamification Approach: Kahoot as a Case Study	Huseyin Bicen; Senay Kocakoyun	International Journal of Emerging Technologies in Learning	2018	North Cyprus/ Turkey	education	One term	65
10.1016/j.addbeh.2016.11.024	PNF 2.0? Initial Evidence that Gamification Can Increase the Efficacy of Brief, Web-based Personalized Normative Feedback Alcohol Interventions	Sarah C. Boyle, Andrew M. Earle, Joseph W. LaBrie, and Daniel J. Smith	Addictive behaviors, April 2017, Vol.67, pp.8-17	2017	USA	health	Two sessions with two weeks in between	237
10.2436/20.3008.01.148	Points, badges, and news. A study of the introduction of gamification into journalism practice	Raul Ferrer Conill	Comunicació: Revista de Recerca i d'Anàlisi, Vol. 33, no 2, p. 45-63	2016	Sweden	online participation	80 hours	20000
Corpus ID: 19007509	Quick Quiz: A Gamified Approach for Enhancing Learning	Christopher Cheong, Christopher Cheong, France Cheong, France Cheong, Justin Filippou, Justin Filippou	Proceedings of Pacific Asia Conference on Information Systems, June 18-22, 2013	2013	Korea	education	Four weeks, one questionnaire at the end	76

10.1007/s11606-018-4552-1	Social Incentives and Gamification to Promote Weight Loss: The LOSE IT Randomized, Controlled Trial	Gregory W. Kurtzman, Susan C. Day, Dylan S. Small, Marta Lynch, Jingsan Zhu, Wenli Wang, Charles A. L. Rareshide, Mitesh S. Patel	Journal of General Internal Medicine, 33(10):1669-1675, 2018	2018	USA	health	36 weeks	198 participants with BMI>30, mean age: 41.4years, 85% women
10.1145/2488388.2488398	Steering user behavior with badges	Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, Jure Leskovec	Proceedings of the 22nd international conference on World Wide Web, May 13-17, 2013	2013	Brazil	online participation	1.5 years July 31st, 2008 to December 31st, 2010	several million users on Stack Overflow
10.1186/s41039-018-0078-8	Students' perception of Kahoot!'s influence on teaching and learning	Sherlock Licorish, Helen Owen, Ben Daniel, Jade George	Research and Practice in Technology Enhanced Learning, 2018, Vol.13(1), pp.1-23	2018	New Zealand	education	part of the third-year course	14 students(university) interviewed from 48 participated
10.1007/s12525-014-0179-1	The application and impact of gamification funware on trip planning and experiences: the case of TripAdvisor's funware	Marianna Sigala	Electron Markets (2015) 25:189–209	2015	Switzerland	online participation	one quick online session	463
10.2196/jmir.3510	The Effect of Social Support Features and Gamification on a Web-Based Intervention for Rheumatoid Arthritis Patients: Randomized Controlled Trial	Ahmed Allam, Zlatina Kostova, Kent Nakamoto, and Peter Johannes Schulz	Journal of medical Internet research 17.1 (2015): e3510.	2015	Switzerland, Italy	health	Four months	157
10.1145/2470654.2470763	The Effect of Virtual Achievements on Student Engagement	Paul Denny	In Proceedings of CHI 2013: Changing Perspectives, April 27–May 2, 2013, Paris, France, pp. 763-772	2013	New Zealand	education	questionnaire; badge achievements log	>1000
10.1016/j.jbusres.2018.07.056	The effects of gamified customer benefits and characteristics on behavioral engagement and purchase: Evidence from mobile exercise application uses	Seongsoo Jang, Philip Jame Kitchen, Jinwon Kim	Journal of Business Research 92, November 2018, pp. 250-259	2018	UK/USA/France	fitness	Three years	5072
http://hdl.handle.net/10214/9197	The Gamification of Physics Education: A Controlled Study of the Effect on Motivation of First-Year Life Science Students	Jordan Rose	-	2015	Canada	education	Two terms	591
10.1007/978-3-319-07293-7_38	The Global Leadership of Virtual Teams in Avatar-Based Virtual Environments	Paul Hayes Jr.	Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2014, Vol.8527, pp.390-400	2014	USA	working environment	one session	34-44

10.1109/ICEED.2016.7856058	The impacts of infusing game elements and gamification in learning	Mageswaran Sanmugam, Norasykin Mohd Zaid, Zaleha Abdullah, Baharuddin Aris	2016 IEEE 8th International Conference on Engineering Education (ICEED)	2016	Malaysia	education	Eight weeks	28
10.1016/j.chb.2015.08.048	Towards understanding the effects of individual gamification elements on intrinsic motivation and performance	Elisa D. Mekler, Florian Brühlmann, Alexandre N.Tuch, Klaus Opwis	Computers in Human Behavior Volume 71, June 2017, Pages 525-534	2015	Switzerland	online participation	one session	273
10.1186/s41239-018-0114-1	Turning a traditional teaching setting into a feedback-rich environment	Arturo González	International Journal of Educational Technology in Higher Education (2018) 15:32	2018	Ireland	education	12 weeks	49
10.1145/3170427.3188608	Understanding Fitness App Usage Over Time: Moving Beyond the Need For Competence	Vanessa R. Lerch, Sharon T. Steinemann, Klaus Opwis	Proceeding CHI EA '18 Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems Paper No. LBW065	2018	Switzerland	fitness	one session	47
10.1109/ISC2.2015.7366196	Using gamification for sustainable transport education: results from an empirical study	Lisa-Maria Putz, Horst Treiblmaier, Sarah Pfoser	Proceedings of 7th Transport Research Arena TRA 2018, April 16-19, 2018, Vienna, Austria	2018	Austria	sustainability	One day	384
10.1145/2556420.2556502	Using gamification to inspire new citizen science volunteers	Anne Bowser, Derek Hansen, Yurong He, Carol Boston, Matthew Reid, Logan Gunnell, Jennifer Preece	Proceeding Gamification '13 Proceedings of the First International Conference on Gameful Design, Research, and Applications; Pages 18-25	2013	USA	surveys, sociology	Five weeks	71
10.2196/games.8902	Using Mobile Health Gamification to Facilitate Cognitive Behavioral Therapy Skills Practice in Child Anxiety Treatment: Open Clinical Trial	Gede Pramana, Bambang Parmanto, James Lomas, Oliver Lindhiem, Philip C Kendall, Jennifer Silk	JMIR serious games, 10 May 2018, Vol.6(2), pp. e 9	2018	USA	health	Eight weeks	35
Corpus ID: 17318743	When the experiment is over: Deploying an incentive system to all the users	Rosta Farzan, Joan M. DiMicco, David R. Millen, Beth Brownholtz, Werner Geyer, Casey Dugan	presented at Symposium on persuasive Technology, 2008	2008	USA	working environment	Six weeks	421

A.1.2 Correlation Matrix Gamification Elements and Application Domains

Table 27 – Correlation Matrix Gamification Elements and Application Domains

	Points	Badges	Leader boards	Teams		Virtual Goods/ Ownership	Quests/ Challenges	Gifting/ Sharing	Social Feedback					Random Encourage ment	Immediate Feedback	Levels		Avatar	Performance Graphs		Narrative	Time Con straints	Rules and Barriers
				Compe-titive	Collabo-rative				Likes	Private Chat	Groups	Linked Social Networks	Public Comm ents			Public Score	Levels		Misc	Progress Bars			
Education	12	14	15	3	2	1	6	1	0	0	0	0	0	1	7	2	6	1	3	2	1	3	1
Fitness	5	5	4	0	0	0	4	0	1	0	0	1	3	0	0	0	2	0	3	0	0	0	0
Health	4	3	1	2	0	0	2	0	2	1	1	1	1	1	2	0	2	0	0	0	0	0	0
Economy	1	2	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0
Psychology	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0
Sustainability	0	1	2	2	1	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	1	1	0
Surveys, Sociology	1	2	2	1	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1	1
Online Participation	4	4	5	0	2	0	1	1	2	0	0	1	3	0	0	1	2	1	0	2	0	0	0
Working Environment	3	1	1	0	3	0	1	0	2	0	0	0	2	0	0	0	0	1	2	0	1	0	0
Sum	31	33	32	8	9	1	16	2	7	1	1	3	10	2	10	3	14	5	10	5	4	5	2

A.2 Supplementary Research Materials Closed Card Sorting

A.2.1 Design and Development of Card Sorting Tool "Tako"

We started the design process by devising a list of requirements (following the method suggested by Robertson and Robertson (2013)). In terms of requirements specific to our setup, we arrived at the following central features:

1) the creation of *presets* for our dataset: The recommended number for one sorting session is set between 50- a maximum of 100 cards – our dataset consists of 591 entries. As such, the preset feature should allow specifying the specific sections of the underlying dataset that will be used in the final sorting.

2) the creation of *open, closed, and hybrid modes*. Behind each of these are a set of different user interactions. While the open mode should allow participants to create and delete folders during the sorting, the closed mode should only allow the experimenter to create an initial set of folders that can't be changed by the participants. The hybrid mode should afford both functionalities.

3) the creation of *subcategories*. A full classification might need a hierarchical structure instead of only one overarching layer of categories. For this, the hierarchical depth should be afforded by the tool.

4) the creation of *card designs* that include a title/name, a description, and an image. To accommodate for better outcomes of sorts, experimenters should be able to design the elements that will be sorted. Therein, the placement of additional information on the sorting items (like pictures and descriptions) should be accommodated by the tool.

5) the functionality to add *descriptions* to the categories before the sorting sessions (for closed and hybrid mode) and for users to add category *descriptions* during the sorting sessions (for hybrid and open mode). This allows for better communication and understanding between the experimenter and the experimentee.

6) the functionality to *clone cards* to afford multi-label classification. It might happen that a card fits several categories. When aiming for card labels rather than one hierarchical outcome, experimenters should be given the possibility to turn on a function where cards can be cloned to then be sorted in different folders.

7) the functionality of *transforming a card into a folder* in hybrid and open mode. To accommodate potential hierarchical diversities, as is the case with our dataset, a function should be given where the experimenter can allow for cards to be transformed into a folder, thus allowing for the identification of underlying hierarchical structures within a dataset.

Interaction Design

We started by researching the tools that were mentioned in the literature on card-sorting (Chaparro et al., 2008); however, several of the tools were either set offline or transformed into proprietary software. None of the tools we found in an additional search included all functionalities we had listed for our categorization of game design elements. As we were also concerned with matters of data protection and further processing of the experiment data, we concluded that it would be worth the effort to build our card sorting tool.

We separated the interaction design of the tool into three general parts: preparation, conduct, and processing. Preparation concerns the upload and preparation of the datasets that serve as the foundation for the categorization process and the preparation of presets (subsamples of the full datasets) for when datasets are too large to be processed in one card sorting session. We first included a process for uploading datasets so that they could be processed by the backend as well as edited in the frontend, allowing for troubleshooting and editing. We designed the tool in a way that different datasets can be stored and organized in a list view. Next, we implemented functionality to prepare presets based on the datasets. Ideally, a card sort consists of datasets of between 30 to 100 cards (Chaparro et al., 2008). However, as is the case for us, datasets can consist of higher numbers of entries. In this case, it is necessary to prepare presets consisting of sections of the original datasets. We also implemented a feature to write and display instructional texts that can be shown to participants at the beginning of a card sort. Finally, we included the selection of card sorting types: closed (where a structure is predefined by the experimenter), open (where users build their structure), and hybrid (where a structure is predefined, but users can change the structure by adding folders). By going through all these steps, a full experimental design for a card sort can be prepared.

In terms of conduct, the interaction design of the card sorting screen was based on classic hierarchical folder structures, as can be found in standard explorers across most operating systems. The (folder) structure is given to the left – representing the categories the cards are to be sorted into, and the cards are listed to the right. The layout is dynamic; however, the default uses a three-lane layout with a downwards scrolling canvas. The card sorting process uses a drag & drop interaction. These choices were made to build on known interaction patterns and to follow rules of keeping the cognitive load at manageable levels (Oviatt, 2006). We also added a search function to increase the ease with which sorted as well as unsorted items can be retrieved (see Figure 76).

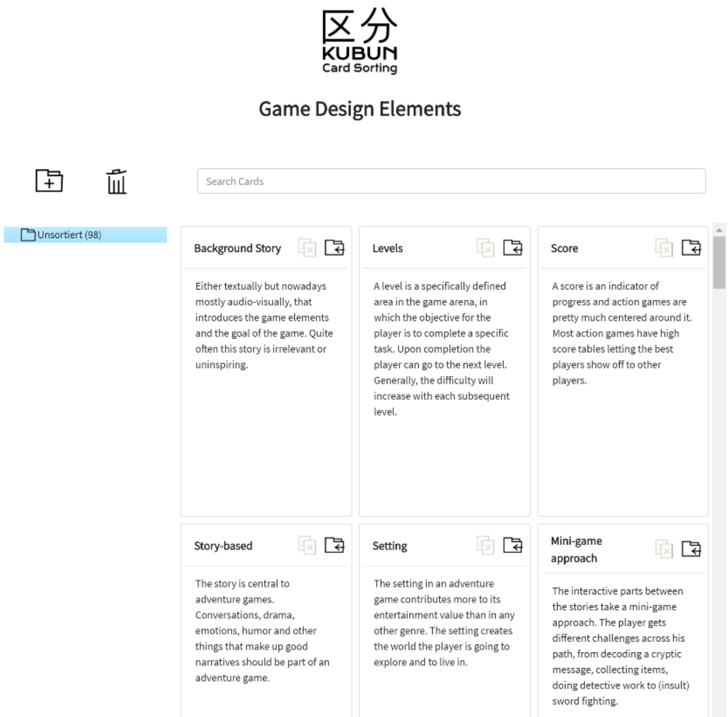


Figure 76 – Screenshot Sorting Page Card Sorting Tool Kubun (now Tako)

In an *open sort*, participants can create a new folder by clicking on the -icon in the upper left corner of the category list. If a folder is already selected, the new folder will be created as a subfolder of the selected folder. Users can rename folders by double-clicking on the folder's label. Empty folders can be deleted by selecting them and then clicking the -icon in the upper left corner. However, if a folder already contains cards, a prompt asks the user to remove the cards before commencing with the deletion. In a *close sort*, neither of these actions are available, and in a *hybrid sort*, these actions can only be performed on folders that were created by the users. The "unsorted" folder follows special rules that can never be deleted. Apart from being drag-and-droppable into any folder, users can conduct two additional actions with the cards. By clicking on the -icon, the card will be transformed into a folder. This action will create a new folder that has the element's name as the label. The corresponding card will automatically be sorted into this folder. Furthermore, users can clone a card by clicking on the -icon. The card is then duplicated into the ever-present "Unsorted" folder. The overall progress is saved in real-time, which is why the tool does not have a save or submit button (see Figure 76). Finally, in terms of processing the experimental data after a sorting process, we added functionality to store each sort in a backend with related metadata – the name of the sort, folder structure, and related card references, as well as the related preset. The list of sorts can be accessed via the main interface.

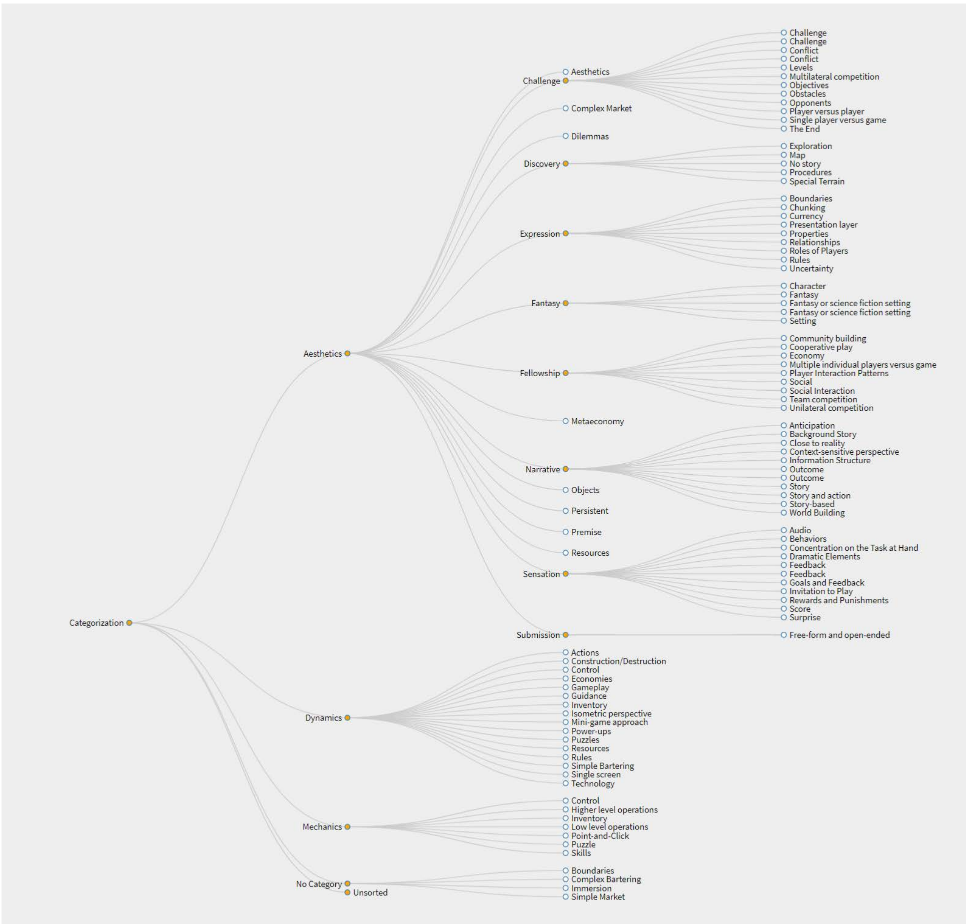


Figure 77 – Screenshot Sort Visualization Card Sorting Tool Kubun (now Tako)

We added two options two examine the sorts: First, the sort can be reopened within the original sorting interface. This way, experimenters can navigate through folder structures the way the participants were able to. Second, we included a tree graph-based visualization to ease qualitative analyses of the sorts

(Figure 77). We included this functionality to better afford an overview of the whole sorted set since all categories and elements can be viewed on a single screen (however - element descriptions are not included in this mode of presentation). To help with navigation, we added functionality to hide subcategories and elements by clicking on a category's node. The functions described above were all implemented in the prototype.

We titled our card sorting tool "Kubun Card Sorting." Kubun (区分) is a word originating in the Japanese language and means "categorization," "segmentation," or "sorting." Later, due to the follow-up project usurping the original name of "Kubun" (see Chapter 4.3.3), we renamed the redesigned version of the card-sorting tool "Tako" (Japanese: たこ), which means "octopus." We chose this name in reference to the tentacular structure of the hierarchies post-sorting with their strands of categories and subcategories originating from one central head node and culminating into one overall structure like an octopus (or, as in our case, an n-pus).

A.2.2 Closed Card Sorting Framework Exploration

Table 28 – Overview of the Card Distributions and Fitness of the Explorative Sorts for Framework Selection

Framework	Viewpoint	No of Categories	Cards in "No Category" (fitness)	Distribution
"MDA" – Hunicke et al. (2004)	Game Analysis	3	3 / 98	(44|14|37)
"Game frame: Using Games as a Strategy for Success" – Dignan (2011)	Game Analysis	10	29 / 98	(7|6|12|8|5|5|2|15|7|3)
"The Art of Game Design: A Book of Lenses" – Schell (2014)	Game Analysis	4	6 / 98	(12|66|10|4)
" It is always a lot of fun! " – Poels et al. (2007)	Game Experience	9	15 / 98	(7|12|5|8|13|2|13|10|13)
"From Game Design Elements to Gamefulness: Defining "Gamification" – Deterding et al. (2011)	Gamification	5	4 / 98	(0|30|28|5|31)
"VAC" – Robinson and Bellotti (2013)	Gamification Analysis	7	2 / 98	(1|11|15|12|6|41|10)
"Five-Feature Model" – King et al. (2010)	Psychology	5	27 / 98	(26|17|8|9|11)
"Player Motivation" – Yee (2006)	Playing Motivations	3	19 / 98	(42|29|8)
"Octalysis" – Chou (2016)	Playing Motivations	8	29 / 98	(13|7|5|10|12|0|8|14)
"Hexad" – Tondello et al. (2016)	Player Types	6	57 / 98	(11|0|17|1|1|11)
"The Trojan Player Typology" – Kahn et al. (2015)	Player Types	6	20 / 98	(12|8|16|16|12|14)
Combined Sort: - "Intrinsic Motivation and Self-Determination in Human Behavior" – Deci and Ryan (1985) - "Gifts Differing: Understanding Personality Type" – Myers and Myers (2010) - "Who Am I?" – Reiss (2001) - "Explorations in Personality" – (Henry A. Murray, 2008)	Psychology	20	45 / 98	(0|9|8|1|2|5|1|1|1|0|4|1|3|0|7|0|7|3|0|0)
"Game Design Essentials" – Mitchell (2012)	Competences	10	8	(1|8|8|9|3|3|39|7|2|10)

A.2.3 Closed Card Sorting Frameworks

MDA (Mechanics, Dynamics, Aesthetics) Framework by Hunicke et al. (2004)

Categorical Structure:

Aesthetics with Subcategories: **Challenge, Discovery, Expression, Fantasy, Fellowship, Narrative, Sensation Submission**
Dynamics
Mechanics

The MDA framework is an established gamification framework that was originally developed at the Game Developers Conference, San Jose 2001-2004, with over 3200 citations as of May 2022 on Google Scholar. Its focus is on offering a perspective from a game developer's as well as a player's perspective by highlighting the overarching processes of what can be perceived by the player, what is preconstructed by the designer, and what happens through the playing interaction. It derives its name from its three overarching categories: Mechanics, Dynamics and Aesthetics. The latter is subdivided into a substructure that lists the types of experiences players seek within the game. This substructure is connected to the Aesthetics component, as it reflects the player-facing facets of games. On the opposite side of this model, the mechanics refer to the components game designers construct to afford the aforementioned experiences. Finally, in between the two constructs, the Dynamics reflect the states that emerge during the playing process between the preset Mechanics of the game and the experienced Aesthetics of the players.

VAC framework (Taxonomy of Gamification Elements for Varying Anticipated Commitment) by Robinson and Bellotti (2013)

Categorical Structure:

Extrinsic Incentives
Feedback and Status Information
General Framing
General Rules and Performance Framing
Intrinsic Incentives
Resources and Constraints
Social Features

The VAC framework is a structure created by Robinson and Bellotti (2013) with the aim of creating a useful resource for practitioners with little experience or expertise in the domain of game design that aims to use gamification. Their framework originates from an online survey on benefits enrollment, which showed that employees wouldn't use available benefits to their full extent. Their structure consists of six major categories of gamification elements: Intrinsic incentives ((such as experiencing flow) and extrinsic incentives (such as gaining virtual value), feedback and status information (such as what is currently happening, what players must do next, what they have done already and what other players are doing and have done others), general framing (such as providing context and motivation for participation, e. g. through back-stories), general rules and performance framing (explanations of what is expected to orient the user towards what constitutes 'good' performance in the gamified context), resources and constraints (the bounds within which the users operate during participation) and social features (affording user interaction with others, outside as well as within the gamified experience).

MfP Framework (Motivations for Play) by Yee (2006)

Achievement with Subcategories: **Advancement, Competition, Mechanics**
Social with Subcategories: **Relationship, Socializing, Teamwork**
Immersion with Subcategories: **Customization, Discovery, Escapism, Role-Playing**

The framework of playing motivations by Yee (2006) is a structure comprising of three overarching categories, with each featuring three to four subcategories. Building on the player types of Bartle (1996), Yee developed a 40-item questionnaire using a 5-point fully-labeled construct-specific scale that they promoted on online portals catering to MMORPG players. In total, over 3000 MMORPG players participated in the online surveys, from which he derived ten components via a round of conducting a first principal component analysis that he formed into the final structure in a second PCA. The Achievement component consists of three subcategories that reflect achievement-oriented motivations: *Advancement* (gain power, progress rapidly, and accumulate in-game symbols of wealth or status), *Mechanics* (interest in analyzing the underlying rules and system in order to optimize character performance), and Competition (desire to challenge and compete with others). The Social component consists of three categories that reflect facets of social motivations: *Socializing* (helping and chatting with other players), *Relationship* (forming long-term, meaningful relationships with others), and *Teamwork* (being part of a group effort). Finally, the Immersion component consists of four categories reflecting components related to escapism and fantasy: Discovery (finding and knowing things that most other players don't), Role-Playing (creating a persona with a background story and interacting with other players to create an improvised story), Customization (customizing the appearance of their character) and Escapism (avoid thinking about real-life problems).

Octalysis Framework by Chou (2016)

Development & Accomplishment
Empowerment of Creativity & Feedback
Epic Meaning & Calling
Loss & Avoidance
Ownership & Possession
Scarcity & Impatience
Social Influence & Relatedness
Unpredictability & Curiosity

The Octalysis Framework developed by Chou (2016) is a gamification design framework built around eight core drives for human motivation. They are mapped on an octagon that further incorporates two additional dimensions. First, the left/right layout is used to differentiate between left-brain-oriented functions such as creativity, self-expression, and social aspects and right-brain-oriented functions such as logic, calculations, and ownership. The top /bottom layout is further used to differentiate between "White Hat Gamification" (drives that make the user feel good) and "Black Hat Gamification" (drives that build on fears and rejective features). While consisting of 5 levels of analysis, the first level suffices for the purpose of labeling gamification elements. The eight core drives summarize the following motifs: *Development and Accomplishment* (sense of growth towards a goal and accomplishing it, challenge-based), *Empowerment of Creativity and Feedback* (creative processes, repeatedly figuring new things out, trying different combinations leading to an end result), *Epic Meaning and Calling* (doing something greater than oneself, being chosen to take action), *Loss and Avoidance* (avoid something negative from happening), *Ownership and Possession* (need to own or control something), *Scarcity and Impatience* (wanting something because it is extremely rare, exclusive, or immediately unattainable), *Social Influence and Relatedness* (mentorship, social acceptance, companionship, competition, envy) and *Unpredictability and Curiosity* (not knowing what is going to happen next).

A.2.4 Participant Attribution

Table 29 – Participant Attribution Matrix

Group	Participant	Set		Framework			
		1	2	MDA	Octalysis	MfP	VAC
1	1	X		X	X		
	2	X			X	X	
	3	X				X	X
	4	X		X			X
	5	X		X		X	
	6	X			X		X
2	7		X	X	X		
	8		X		X	X	
	9		X			X	X
	10		X	X			X
	11		X	X		X	
	12		X		X		X
3	13	X	X	XX			
	14	X	X		XX		
	15	X	X			XX	
	16	X	X				XX

A.2.5 Participant Overview

Table 30 – Overview Participants' Demographic Information, Playing Expertise, and Player Types – Closed Card Sort

Part ID*	Gender	Age	Education	Major	Expertise	Playing Exp.**	Preferred Devices	Preferred Game Type***	Player Type
1	m	25	Abitur	Mechanical Engineering	Hobby	>16 hpw >10 y	PC, Console	sp&mp, on&off	Free Spirit
2	m	25	Bachelor	Infonomics/ Computer Science	Hobby	4-6 hpw >10 y	PC	sp&mp, on&off	Philanthropist
3	m	22	Abitur	Mathematics	Hobby	10-12 hpw >10 y	PC, Mobile Device	sp&mp, on&off	Philanthropist
4	m	26	Bachelor	Infonomics/ Management	Developme nt	<1 hpw >10 y	PC	sp&mp, on&off	Philanthropist
5	f	24	Abitur	Infonomics	None	-	-	-	Socializer
6	m	26	Abitur	-	Hobby	>16 hpw >10 y	PC	sp&mp, on&off	Socializer/ Disruptor
7	m	28	Master	Electrical Engineering	Hobby	>16 hpw >10 y	PC, Console, Mobile Device	sp&mp, on	Philanthropist
8	m	25	Atcc	-	None	4-6 hpw >10 y	Mobile Device	mp, on	Free Spirit
9	f	24	Abitur	Ecosystem Management	None	- 7-8 y	PC, Console, Mobile Device	mp, off	Philanthropist
10	f	23	Abitur	Infonomics	Research	7-9 hpw >10 y	PC, Console, Mobile Device	sp&mp, on	Player

11	m	26	Abitur	Infonomics	Hobby	<1 hpw >10 y	PC	sp&mp, on&off	Philanthropist
12	m	22	Abitur	Infonomics	Hobby	4-6 hpw 9-10 y	PC, Console, Mobile Device	sp&mp, on	Socializer
13	m	30	Diploma	Mechanical Engineering	Hobby	7-9 hpw >10 y	PC, Console, Mobile Device	sp&mp, on&off	Socializer
14	f	26	Bachelor	Infonomics	None	-	-	-	Socializer
15	f	23	Bachelor	Infonomics	Hobby	>1 hpw 3-4 y	PC Console, Mobile Device	sp&mp, on&off	Achiever
16	m	25	Abitur	Infonomics	Hobby	1-3 hpw >10 y	PC Console, Mobile Device	sp&mp, on	Philanthropist

* Participant ID
**hours per week (hpw), year (y)
**Single-player (sp), multi-player (mp), online gaming (on), offline gaming (off)

Table 31 – Overview Hexad Player Type Distribution – Closed Card Sort

Part ID*	Major	Philantropist	Socialiser	Free Spirit	Achiever	Disruptor	Player
1	Mechanical engineering	14	17	19	18	16	17
2	Infonomics/ Computer Science	21	16	19	19	14	12
3	Mathematics	23	13	19	19	8	20
4	Infonomics/ Management	20	16	18	15	14	17
5	Infonomics	22	26	18	3	15	16
6	-	19	21	17	18	21	13
7	Electrical Engineering	22	20	15	18	9	15
8	-	17	15	20	14	14	19
9	Ecosystem management	21	19	17	14	13	16
10	Infonomics	18	17	16	11	17	20
11	Infonomics	21	14	20	19	14	12
12	Infonomics	17	19	18	17	13	17
13	Mechanical engineering	16	19	16	18	16	14
14	Infonomics	19	22	14	19	7	20
15	Infonomics	18	15	18	21	12	16
16	Infonomics	20	18	18	12	18	16

* Participant ID

A.2.6 Closed Card Sort Additional Data

Table 32 – Overview Sorting Outcome Closed Card Sort Fitness and Sort Duration

Participant	Group	Sort 1	Set	No Category (fitness)	Sort Duration	Sort 2	Set	No Category (fitness)	Sort Duration
2	1	MDA	1	-	20 min	Yee	1	29	19 min
10	1	MDA	1	9	36 min	Robinson	1	7	27 min
7	1	Octalysis	1	9	21 min	MDA	1	2	20 min
12	1	Octalysis	1	18	24 min	Yee	1	20	19 min
8	1	VAC	1	68	24 min	Yee	1	44	19 min
9	1	VAC	1	1	39 min	Octalysis	1	18	42 min
6	2	VAC	2	24	23 min	Octalysis	2	49	35 min
4	2	Yee	2	0	36 min	Octalysis	2	ID	20 min
11	2	Yee	2	4	18 min	Robinson	2	2	52 min
5	2	MDA	2	1	17 min	Robinson	2	6	10 min
16	2	MDA	2	0	30 min	Octalysis	2	24	23 min

14	2	MDA	2	3	27 min	Yee	2	3	18 min
1	3	MDA	1	2	50 min	MDA	2	4	109 min
13	3	Yee	1	0	21 min	Yee	2	-	9 min
15	3	Octalysis	1	10	34 min	Octalysis	2	13	21 min
3	3	Robinson	1	7	29 min	Robinson	2	4	15 min

MDA Framework by Hunicke et al. (2004)

Table 33 – Overview Category Use per Set and Total – MDA Framework

			Aesthetics											
Part*	Sort	Set	Challenge	Discovery	Expression	Fantasy	Fellowship	Narrative	Sensation	Submission	Total	Dynamics	Mechanics	No Category
1	1	1	6	5	10	3	4	11	14	2	57	12	13	2
2	1	1	-	-	-	-	-	-	-	-	-	-	-	-
10	1	1	10	7	4	4	6	3	4	0	38	16	21	9
7	2	1	9	11	6	6	6	7	6	6	57	8	17	2
Average Set1			**9.9%**	**9.1%**	**7.9%**	**5.2%**	**6.3%**	**8.3%**	**9.5%**	**3.2%**	**60.3%**	**14.3%**	**20.2%**	**5.2%**
5	1	2	5	2	1	3	7	1	1	0	30	29	38	1
14	1	2	13	10	10	6	8	5	7	8	67	14	14	3
16	1	2	8	2	3	5	12	8	4	0	49	24	25	0
1	2	2	12	5	9	5	9	11	11	1	63	16	7	4
Average Set2			**9.7%**	**4.8%**	**5.9%**	**4.8%**	**9.2%**	**6.4%**	**5.9%**	**2.3%**	**53.3%**	**21.2%**	**21.4%**	**2.0%**
Total Average			**9.8%**	**7.0%**	**6.9%**	**5.0%**	**7.8%**	**7.4%**	**7.7%**	**2.7%**	**56.8%**	**17.7%**	**20.8%**	**3.6%**

*Participant

Table 34 – Overview Card Commonality per Category per Set and Total – MDA Framework

		Aesthetics											
In Common	Set	Challenge	Discovery	Expression	Fantasy	Fellowship	Narrative	Sensation	Submission	Avg.	Dynamics	Mechanics	No Category
4	**1**	-	-	-	-	-	-	-	-	-	-	-	-
4	**2**	0%	0%	0%	20%	19%	0%	0%	0%	5%	0%	0%	0%
Average		-	-	-	-	-	-	-	-	-	-	-	-
Delta		-	-	-	-	-	-	-	-	-	-	-	-
3	**1**	11%	0%	0%	0%	20%	0%	0%	0%	4%	3%	6%	10%
3	**2**	16%	6%	0%	10%	25%	11%	0%	0%	9%	0%	9%	0%
Average		14%	3%	0%	5%	23%	6%	0%	0%	6%	2%	8%	5%
Delta		5%	6%	0%	10%	5%	11%	0%	0%	3%	3%	3%	10%
2	**1**	17%	21%	21%	30%	20%	5%	4%	0%	15%	10%	34%	10%
2	**2**	20%	0%	10%	10%	19%	17%	10%	0%	11%	38%	28%	14%
Average		19%	11%	16%	20%	20%	11%	7%	0%	13%	24%	31%	12%
Delta		3%	21%	11%	20%	1%	12%	6%	0%	7%	28%	6%	4%
1	**1**	72%	79%	79%	70%	60%	95%	96%	100%	81%	87%	60%	80%
1	**2**	64%	94%	90%	60%	38%	72%	90%	100%	76%	62%	64%	86%
Average		68%	87%	85%	65%	49%	84%	93%	100%	79%	75%	62%	83%
Delta		8%	15%	11%	10%	22%	23%	6%	0%	8%	25%	4%	6%

"VAC" Framework by Robinson and Bellotti (2013)

Table 35 – Overview Category Use per Set and Total – VAC Framework

Participant	So	Se	Extrinsic Incent.*	Feedback & Stat. Inf.*	General Framing*	Gen. Rules & Perf. Fr.*	Intr. Incent.*	Resourc. & Constr.*	Social Feat.*	No Category
3	**1**	**1**	13	5	16	14	11	11	7	7
8	**1**	**1**	4	0	3	2	3	2	2	68
9	**1**	**1**	23	5	16	20	7	5	7	1
10	**2**	**1**	10	7	20	10	12	12	6	7
Average Set1			**14.9%**	**5.1%**	**16.4%**	**13.7%**	**9.8%**	**8.9%**	**6.5%**	**24.7%**
6	**1**	**2**	0	24	22	14	2	11	11	24
3	**2**	**2**	11	6	25	17	12	13	10	4
5	**2**	**2**	7	6	11	39	15	8	6	6

11	2	2	2	4	22	38	8	15	10	2
Average Set2			**5.1%**	**10.2%**	**20.4%**	**27.6%**	**9.4%**	**12.0%**	**9.4%**	**9.4%**
Overall Average			**10.0%**	**7.6%**	**18.4%**	**20.6%**	**9.6%**	**10.5%**	**8.0%**	**16.9%**

* Extrinsic Incentives, Feedback and Status Information, General Framing, General Rules and Performance Framing, Intrinsic Incentives, Resources and Constraints, Social Features

Table 36 – Overview Card Commonality per Category per Set and Total – VAC Framework

In Common	Set	Extr. Incent.*	Feedback & Stat. Inf. *	General Framing*	Gen. Rules & Perf. Fr. *	Intr. Incent. *	Resourc. & Constr. *	Social Feat. *	No Category
4	1	0%	0%	3%	0%	0%	0%	8%	1%
4	2	0%	10%	0%	3%	0%	0%	18%	0%
Average		0%	5%	2%	2%	0%	0%	13%	1%
Delta		0%	10%	3%	3%	0%	0%	10%	1%
3	1	11%	0%	5%	9%	14%	5%	8%	6%
3	2	0%	0%	14%	4%	4%	26%	18%	0%
Average		6%	0%	10%	7%	9%	16%	13%	3%
Delta		11%	0%	9%	5%	10%	21%	10%	6%
2	1	20%	13%	26%	21%	29%	33%	42%	6%
2	2	0%	14%	33%	31%	25%	22%	29%	24%
Average		10%	14%	30%	26%	27%	28%	36%	15%
Delta		20%	1%	7%	10%	4%	11%	13%	18%
1	1	69%	87%	66%	70%	57%	62%	42%	87%
1	2	100%	76%	53%	63%	71%	52%	35%	76%
Average		85%	82%	60%	67%	64%	57%	39%	82%
Delta		31%	11%	13%	7%	14%	10%	7%	11%

* Extrinsic Incentives, Feedback and Status Information, General Framing, General Rules and Performance Framing, Intrinsic Incentives, Resources, and Constraints, Social Features

"MfP" Framework by Yee (2006)

Table 37 – Overview Category Use per Set and Total – MfP Framework

			Achievement				Immersion					Social				
Part*	So	Se	Advancement	Competition	Mechanics	Total	Customization	Discovery	Escapism	Role-Playing	Total	Relationship	Socializing	Teamwork	Total	No Cat.
13	1	1	9	9	23	41	9	2	5	16	32	5	2	4	11	0
2	2	1	19	3	5	27	3	7	3	13	29	0	1	1	2	29
8	2	1	16	2	7	25	2	2	7	3	14	0	1	0	1	44
12	2	1	4	7	5	22	14	2	1	2	19	5	12	1	23	20
Average Set1			**14.3%**	**6.3%**	**11.9%**	**34.2%**	**8.3%**	**3.9%**	**4.8%**	**10.1%**	**28.0%**	**3.0%**	**4.8%**	**1.8%**	**11.0%**	**27.7%**
4	1	2	21	14	18	53	3	12	8	12	36	2	3	4	9	0
11	1	2	11	9	34	56	0	6	1	7	32	1	1	0	6	4
13	2	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	2	2	13	17	24	54	6	10	2	3	22	7	10	2	19	3
Average Set2			15.3%	13.6%	25.9%	55.4%	3.1%	9.5%	3.7%	7.5%	30.6%	3.4%	4.8%	2.0%	11.6%	2.4%
Overall Average			**14.8%**	**9.9%**	**18.9%**	**44.8%**	**5.7%**	**6.7%**	**4.3%**	**8.8%**	**29.3%**	**3.2%**	**4.8%**	**1.9%**	**11.3%**	**15.0%**

*Participant

Table 38 – Overview Card Commonality per Category per Set and Total – MfP Framework

In Common	Set	Achievement				Immersion					Social				No Cat.
		Advancement	Competition	Mechanics	Avg.	Customization	Discovery	Escapism	Role-Playing	Avg.	Relationship	Socializing	Teamwork	Avg.	
4	1	0%	3%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
4	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Average		-	-	-	-	-	-	-	-	-	-	-	-	-	-
Delta		-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	1	10%	0%	4%	5%	5%	11%	0%	8%	6%	0%	7%	0%	2%	7%
3	2	3%	3%	18%	8%	0%	4%	0%	0%	1%	0%	0%	0%	0%	0%
Average		7%	2%	11%	6%	3%	8%	0%	4%	4%	0%	4%	0%	1%	4%
Delta		7%	3%	14%	6%	5%	7%	0%	8%	4%	0%	7%	0%	2%	7%

2	1	26%	11%	36%	24%	37%	22%	0%	25%	21%	0%	0%	20%	7%	42%
2	2	13%	31%	36%	27%	0%	8%	9%	29%	12%	0%	17%	0%	6%	0%
Average		20%	21%	36%	26%	19%	15%	5%	27%	16%	0%	9%	10%	6%	21%
Delta		13%	20%	0%	8%	37%	14%	9%	4%	11%	0%	17%	20%	9%	42%
1	1	61%	89%	61%	70%	58%	67%	100%	67%	73%	100%	93%	80%	91%	52%
1	2	84%	66%	45%	65%	100%	88%	91%	71%	88%	100%	83%	100%	94%	100%
Average		73%	78%	53%	68%	79%	78%	96%	69%	80%	100%	88%	90%	93%	76%
Delta		23%	23%	16%	15%	42%	21%	9%	4%	13%	0%	10%	20%	7%	48%

Octalysis Framework by Chou (2016)

Table 39 – Overview Category Use per Set and Total – Octalysis Framework

Part*	**So**	**Se**	**Dev. & Acc****	**Emp & Feedb. ****	**Ep.M. & Call. ****	**Loss & Av. ****	**Own. & Pos. ****	**Scarc. & Imp****	**Soc.I. & Rel. ****	**Unpr. & Cur****	**No Cat.**
7	1	1	17	9	11	6	10	7	8	7	9
12	1	1	13	10	6	8	3	5	15	6	18
15	1	1	18	3	16	11	6	2	8	10	10
9	2	1	10	14	4	6	8	4	14	6	18
Average Set1			**17.3%**	**10.7%**	**11.0%**	**9.2%**	**8.0%**	**5.4%**	**13.4%**	**8.6%**	**16.4%**
4	2	2	-	-	-	-	-	-	-	-	-
6	2	2	15	2	8	5	1	11	7	0	49
15	2	2	18	6	20	0	16	1	14	10	13
16	2	2	10	3	30	6	2	4	9	10	24
Average Set2			**14.6%**	**3.7%**	**19.7%**	**3.7%**	**6.5%**	**5.4%**	**10.2%**	**6.8%**	**29.3%**
Overall Average			**15.9%**	7.2%	**15.4%**	**6.5%**	7.2%	**5.4%**	**11.8%**	**7.7%**	**22.8%**

*Participant

** Development & Accomplishment, Empowerment of Creativity & Feedback, Epic Meaning & Calling, Loss & Avoidance, Ownership & Possession, Scarcity & Impatience, Social Influence & Relatedness, Unpredictability & Curiosity

Table 40 – Overview Card Commonality per Category per Set and Total – Octalysis Framework

In Common	**Set**	**Dev. & Acc***	**Emp & Feedb. ***	**Ep.M. & Call. ***	**Loss & Av. ****	**Own. & Pos. ***	**Scarc. & Imp***	**Soc.I. & Rel. ***	**Unpr. & Cur***	**No Cat.**
4	1	0%	0%	0%	5%	0%	0%	4%	5%	0%
4	2	-	-	-	-	-	-	-	-	-
Average		-	-	-	-	-	-	-	-	-
Delta		-	-	-	-	-	-	-	-	-
3	1	16%	7%	4%	5%	5%	0%	12%	5%	15%
3	2	7%	0%	8%	0%	0%	0%	33%	0%	8%
Average		12%	4%	6%	3%	3%	0%	23%	3%	12%
Delta		9%	7%	4%	5%	5%	0%	21%	5%	7%
2	1	21%	19%	30%	30%	32%	29%	44%	20%	36%
2	2	30%	10%	37%	0%	19%	7%	33%	18%	21%
Average		26%	15%	34%	15%	26%	18%	39%	19%	29%
Delta		9%	9%	7%	30%	13%	22%	11%	2%	15%
1	1	63%	74%	67%	60%	63%	71%	40%	70%	48%
1	2	63%	90%	55%	100%	81%	93%	33%	82%	71%
Average		63%	82%	61%	80%	72%	82%	37%	76%	60%
Delta		0%	16%	12%	40%	18%	22%	7%	12%	23%

*Development & Accomplishment, Empowerment of Creativity & Feedback, Epic Meaning & Calling, Loss & Avoidance, Ownership & Possession, Scarcity & Impatience, Social Influence & Relatedness, Unpredictability & Curiosity

A.3 Supplementary Research Materials Open Card Sorting

A.3.1 Rationale Participant Selection

The GameLab Karlsruhe is an interdisciplinary institution consisting of members of different departments spanning two universities (Karlsruhe Institute of Technology (KIT) and Karlsruhe University of Arts and Design (HfG)). It covers courses on game design history and philosophy, seminars on active game development, courses on programming, drawing, 2D animation, 3D modeling and animation, and a discussion forum on current topics in the world of gaming. Its goal is to offer a space for interdisciplinary discussion and creation and provide the necessary education for game designers (active as well as in training). As can be verified by the majors of our participants, the sample group of our experts consists of a representative sample of typical GameLab members. We acknowledge that the overall number of participants in the experiment is too low to make any conclusions on a quantitative level – we consider the gained insights as pointers for our next steps of research. However, in terms of a sample, this group of experts is representative of our peers and the future target audience for using the final game design element ontology.

A.3.2 Participant Data

Table 41 – Overview Participants' Demographic Information, Playing Expertise, and Player Types – Open Card Sort

Part ID*	Gender	Age	Education	Expertise	Playing Exp**	Preferred Devices	Preferred Game Type***	Player Type
1	f	27	Bachelor	Research	7-9 hpw >10 y	PC/Console/ Mobile Device	sp	Achiever
2	f	23	Abitur	Hobby	4-6 hpw >10 y	Console/Mobile Device	sp, on&off	Free Spirit
3	m	24	Abitur	Development	>16 hpw >10 y	PC/Console/ Mobile Device	sp&mp, on&off	Philanthropist, Player
4	m	31	Bachelor	Development/ Research	4-6 hpw >10 y	PC/Console/ Mobile Device	sp&mp, on&off	Free Spirit
5	n. a.	25	Abitur	Development	7-9 hpw >10 y	PC/Console	sp&mp, on&off	Free Spirit
6	f	23	Abitur	Development	13-16 hpw >10 y	PC/Console/ Mobile Device	sp&mp, on&off	Free Spirit
7	m	26	Bachelor	Hobby	10-12 hpw >10 y	PC	sp&mp, on&off	Free Spirit
8	m	27	Bachelor	Development/ Research	4-6 hpw >10 y	PC/Console/ Mobile Device	sp&mp, on&off	Philanthropist

* Participant ID
**hours per week (hpw), year (y)
**Single-player (sp), multi-player (mp), online gaming (on), offline gaming (off)

Table 42 – Overview Hexad Player Type Distribution – Open Card Sort

ID	Major	Philanthropist	Socialiser	Free Spirit	Achiever	Disruptor	Player
1	Science of Art/ German Studies	24%	20%	16%	**32%**	-16%	24%
2	Product Design	18%	12%	**26%**	21%	3%	21%
3	Communication/ Graphic Design	**24%**	15%	21%	21%	-6%	**24%**
4	Computer Science	11%	14%	**28%**	25%	11%	11%
5	Computer Science	32%	12%	**40%**	0%	24%	-8%
6	Media Art	-17%	-17%	**56%**	39%	17%	22%

7	Computer Science	27%	-5%	**36%**	18%	14%	9%
8	Product Design	**28%**	24%	16%	16%	20%	-4%

Table 43 – Comparison of Player Type Distribution in the Open Card Sort vs. Average on Gamified.uk

Player Types	Average on Gamified.uk	Experiment Average
Philanthropist	26%	18%
Free Spirit	22%	30%
Achiever	18%	22%
Socialiser	16%	9%
Player	14%	12%
Disruptor	3%	8%

A.3.3 Usability Evaluation & Suggestions

As we wanted to gain feedback on the usability of our tool, we analyzed the notes the subjects made during the pretest and gathered further feedback during the debriefing session. In terms of overall usability, all functions worked as implemented and were understood without further explanations. We discovered two implementation issues during the experiment: duplicated cards could not be deleted, and lengthy folder labels were unreadable as they were overlapped by the cards on the main area of the screen. All other usability issues revolved mainly around the topic of missing features: First, it was criticized that only empty categories could be deleted. Accordingly, it was suggested to allow category deletion in any case and automatically place elements of the deleted folder back into the unsorted folder. It was further suggested to keep the folder structure on the left floating on the spot during the scrolling of the list of cards. It was suggested that instead of sorting cards in a static order (in this experiment, the cards were always organized according to their ID, resulting in cards with high IDs always ending up at the bottom of the list), the tool should afford sorting options (sort cards alphabetically, by ID, by last interaction). Another suggestion was to add a visual indication on the card if it was duplicated. Generally, there was a demand for more audio-visual feedback (like confirmation dialogues) was requested, especially for actions like card duplication. Finally, it was recommended that a feature should be added where categories, as well as the overall viewpoint, can be given detailed descriptions.

A.4 Supplementary Research Materials Keyword Matching

A.4.1 Explorative, Qualitative Adjacent-Node-Analysis

In an additional explorative analysis, we semantically clustered the connected nodes for each of Bartle's four items based on similar underlying motifs and gave each cluster a representative category name. When comparing the original descriptions of each of the four nodes on a semantic level to the emerging clusters from the connected nodes, each player type is well represented in terms of their original description through the sum of the nodes they connect to (see Figure 78):

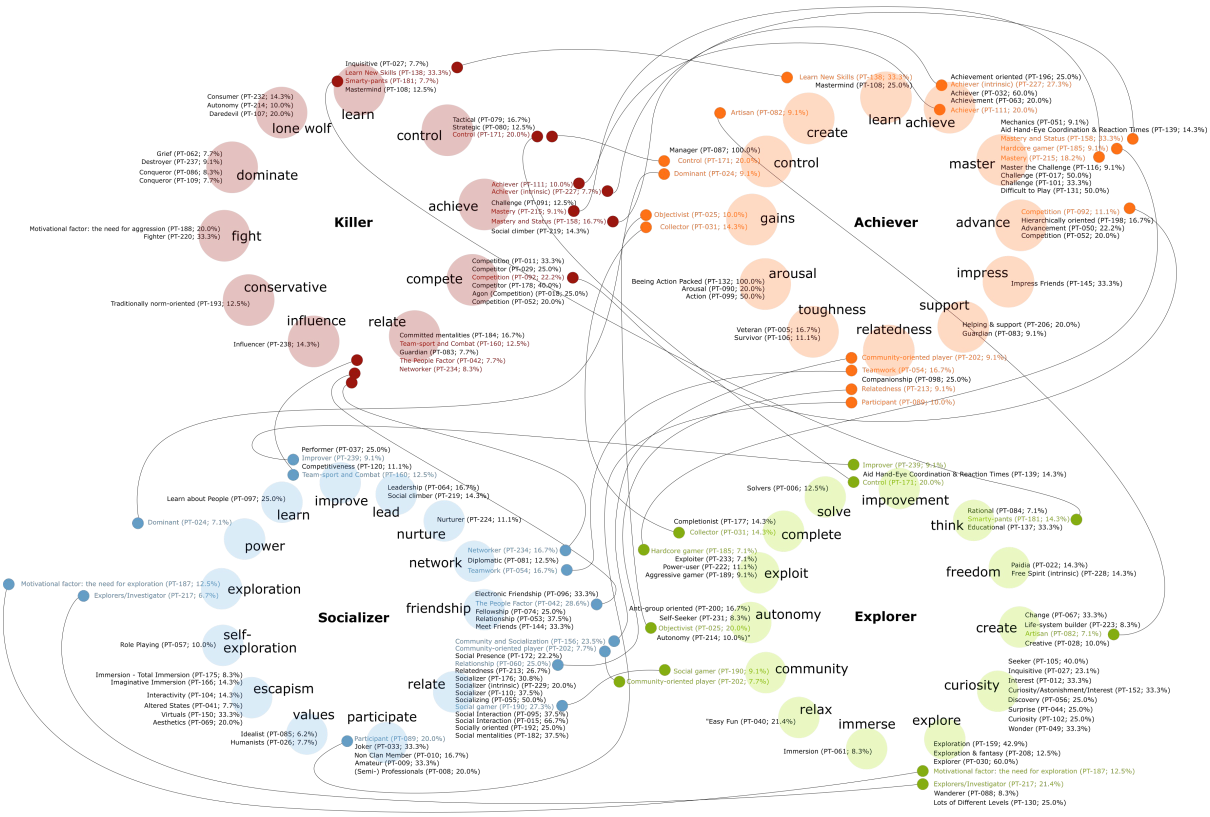


Figure 78 – Exemplary Cluster Analysis of Nodes Connected to Bartle's Player Types

Achiever

Bartle's Description: "Achievers are interested in doing things to the game, ie. in ACTING on the WORLD. It's the fact that the game environment is a fully-fledged world in which they can immerse themselves that they find compelling; its being shared with other people merely adds a little authenticity, and perhaps a competitive element. The point of playing is to master the game, and make it do what you want it to do; there's nothing intrinsically worthwhile in rooting out irrelevant details that will never be of use, or in idling away your life with gossip. Achievers are proud of their formal status in the game's built-in level hierarchy, and of how short a time they took to reach it."

Table 44 – Cluster Overview for Bartle's Player Type Achiever

Category	Cat. Size	Cat. Rel.	IDs Connected to Achiever	ID	Re. Avg. Connection Strength
Achieve	5	30.46%	Achievement oriented	PT-196	25.00%
			Achiever (intrinsic)	PT-227	27.30%
			Achiever	PT-032	60.00%
			Achiever	PT-111	20.00%
			Achievement	PT-063	20.00%
Advance	4	17.50%	Competition	PT-092	11.10%
			Competition	PT-052	20.00%
			Hierarchically oriented	PT-198	16.70%
			Advancement	PT-050	22.20%
Arousal	3	56.67%	Beeing Action Packed	PT-132	100.00%
			Arousal	PT-090	20.00%
			Action	PT-099	50.00%
Control	3	43.03%	Dominant	PT-024	9.10%
			Control	PT-171	20.00%
			Manager	PT-087	100.00%
Create	1	9.10%	Artisan	PT-082	9.10%
Gains	2	12.15%	Collector	PT-031	14.30%
			Objectivist	PT-025	10.00%
Impress	1	33.30%	Impress Friends	PT-145	33.30%
Learn	2	29.15%	Learn New Skills	PT-138	33.30%
			Mastermind	PT-108	25.00%
Master	9	25.16%	Hardcore gamer	PT-185	9.10%
			Mastery and Status	PT-158	33.30%
			Challenge	PT-101	33.30%
			Aid Hand-Eye Coordination & Reaction Times	PT-139	14.30%
			Mastery	PT-215	18.20%
			Challenge	PT-017	50.00%
			Difficult to Play	PT-131	50.00%
			Master the Challenge	PT-116	9.10%
			Mechanics	PT-051	9.10%
Relatedness	5	13.98%	Community-oriented player	PT-202	9.10%
			Teamwork	PT-054	16.70%
			Participant	PT-089	10.00%
			Relatedness	PT-213	9.10%
			Companionship	PT-098	25.00%
Support	2	14.55%	Helping & support	PT-206	20.00%
			Guardian	PT-083	9.10%
Toughness	2	13.90%	Survivor	PT-106	11.10%
			Veteran	PT-005	16.70%

Explorer

Bartle's Description: "Explorers are interested in having the game surprise them, ie. in INTERACTING with the WORLD. It's the sense of wonder which the virtual world imbues that they crave for; other players add depth to the game, but they aren't essential components of it, except perhaps as sources of new areas to visit. Scoring points all the time is a worthless occupation, because it defies the very open-endedness that makes a world live and breathe. Most accomplished explorers could easily rack up sufficient points to reach the top, but such one-dimensional behaviour is the sign of a limited intellect.

Explorers are proud of their knowledge of the game's finer points, especially if new players treat them as founts of all knowledge."

Table 45 – Cluster Overview for Bartle's Player Type Explorer

Category	Category Size	Category Relevance	IDs Connected to Explorer	ID	Re. Avg. Connection Strength
Autonomy	4	13.75%	Anti-group oriented	PT-200	16.70%
			Self-Seeker	PT-231	8.30%
			Objectivist	PT-025	20.00%
			Autonomy	PT-214	10.00%
Community	2	8.40%	Community-oriented player	PT-202	7.70%
			Social gamer	PT-190	9.10%
Complete	2	14.30%	Collector	PT-031	14.30%
			Completionist	PT-177	14.30%
Create	4	14.68%	Change	PT-067	33.30%
			Life-system builder	PT-223	8.30%
			Artisan	PT-082	7.10%
			Creative	PT-028	10.00%
Curiosity	8	29.75%	Interest	PT-012	33.30%
			Curiosity/Astonishment/Interest	PT-152	33.30%
			Inquisitive	PT-027	23.10%
			Discovery	PT-056	25.00%
			Surprise	PT-044	25.00%
			Curiosity	PT-102	25.00%
			Seeker	PT-105	40.00%
			Wonder	PT-049	33.30%
Exploit	4	8.60%	Hardcore gamer	PT-185	7.10%
			Exploiter	PT-233	7.10%
			Power-user	PT-222	11.10%
			Aggressive gamer	PT-189	9.10%
Explore	7	26.09%	Motivational factor: the need for exploration	PT-187	12.50%
			Explorers/Investigator	PT-217	21.40%
			Explorer	PT-030	60.00%
			Exploration	PT-159	42.90%
			Exploration & fantasy	PT-208	12.50%
			Lots of Different Levels	PT-130	25.00%
			Wanderer	PT-088	8.30%
Freedom	2	14.30%	Free Spirit (intrinsic)	PT-228	14.30%
			Paidia	PT-022	14.30%
Immerse	1	8.30%	Immersion	PT-061	8.30%
Improvement	3	14.47%	Improver	PT-239	9.10%
			Aid Hand-Eye Coordination & Reaction Times	PT-139	14.30%
			Control	PT-171	20.00%
Relax	1	21.40%	Easy Fun	PT-040	21.40%
Solve	1	12.50%	Solvers	PT-006	12.50%
Think	3	18.23%	Smarty-pants	PT-181	14.30%

			Rational	PT-084	7.10%
			Educational	PT-137	33.30%

Socializer

Bartle's Description: "Socialisers are interested in INTERACTING with other PLAYERS. This usually means talking, but it can extend to more exotic behaviour. Finding out about people and getting to know them is far more worthy than treating them as fodder to be bossed around. The game world is just a setting; it's the characters that make it so compelling."

Table 46 – Cluster Overview for Bartle's Player Type Socializer

Category	Category Size	Category Relevance	IDs Connected to Socializer	ID	Re. Avg. Connection Strength
Escapism	6	16.32%	Altered States	PT-041	7.70%
			Immersion - Total Immersion	PT-175	8.30%
			Interactivity	PT-104	14.30%
			Aesthetics	PT-069	20.00%
			Imaginative Immersion	PT-166	14.30%
			Virtuals	PT-150	33.30%
Exploration	2	9.60%	Explorers/Investigator	PT-217	6.70%
			Motivational factor: the need for exploration	PT-187	12.50%
Friendship	6	30.45%	Relationship	PT-060	25.00%
			Fellowship	PT-074	25.00%
			Meet Friends	PT-144	33.30%
			Relationship	PT-053	37.50%
			Electronic Friendship	PT-096	33.30%
			The People Factor	PT-042	28.60%
Improve	4	14.43%	Improver	PT-239	9.10%
			Competitiveness	PT-120	11.10%
			Team-sport and Combat	PT-160	12.50%
			Performer	PT-037	25.00%
Lead	2	15.50%	Social climber	PT-219	14.30%
			Leadership	PT-064	16.70%
	1	25.00%	Learn about People	PT-097	25.00%
Network	3	15.30%	Teamwork	PT-054	16.70%
			Networker	PT-234	16.70%
			Diplomatic	PT-081	12.50%
Nurture	1	11.10%	Nurturer	PT-224	11.10%
Participate	5	24.66%	Participant	PT-089	20.00%
			Joker	PT-033	33.30%
			(Semi-) Professionals	PT-008	20.00%
			Non Clan Member	PT-010	16.70%
			Amateur	PT-009	33.30%
Power	1	7.10%	Dominant	PT-024	7.10%

Relate	13	31.72%	Social mentalities	PT-182	37.50%
			Social Presence	PT-172	22.20%
			Community-oriented player	PT-202	7.70%
			Relatedness	PT-213	26.70%
			Social Interaction	PT-095	37.50%
			Socializer	PT-110	37.50%
			Social gamer	PT-190	27.30%
			Community and Socialization	PT-156	23.50%
			Socially oriented	PT-192	25.00%
			Socializing	PT-055	50.00%
			Social Interaction	PT-015	66.70%
			Socializer	PT-176	30.80%
			Socializer (intrinsic)	PT-229	20.00%
Self-exploration	1	10.00%	Role Playing	PT-057	10.00%
Values	2	6.95%	Idealist	PT-085	6.20%
			Humanists	PT-026	7.70%

Killer

Bartle's Description: "Killers are interested in doing things to people, ie. in ACTING on other PLAYERS. Normally, this is not with the consent of these "other players" (even if, objectively, the interference in their play might appear "helpful"), but killers don't care; they wish only to demonstrate their superiority over fellow humans, preferably in a world which serves to legitimise actions that could mean imprisonment in real life. Accumulated knowledge is useless unless it can be applied; even when it is applied, there's no fun unless it can affect a real person instead of an emotionless, computerised entity. Killers are proud of their reputation and of their oft-practiced fighting skills."

Table 47 – Cluster Overview for Bartle's Player Type Killer

Category	Category Size	Category Relevance	Connected IDs to Killer	ID	Re. Avg. Connection Strength
			Mastery and Status	PT-158	16.70%
			Social climber	PT-219	14.30%
			Achiever (intrinsic)	PT-227	7.70%
			Mastery	PT-215	9.10%
			Achiever	PT-111	10.00%
Achieve	6	11.72%	Challenge	PT-091	12.50%
			Competitor	PT-029	25.00%
			Competition	PT-092	22.20%
			Competitor	PT-178	40.00%
			Agôn (Competition)	PT-018	25.00%
			Competition	PT-052	20.00%
Compete	6	27.58%	Competition	PT-011	33.30%
Conservative	1	12.50%	Traditionally norm-oriented	PT-193	12.50%
Control	3	16.40%	Strategic	PT-080	12.50%

			Control	PT-171	20.00%
			Tactical	PT-079	16.70%
			Grief	PT-062	7.70%
			Destroyer	PT-237	9.10%
			Conqueror	PT-109	7.70%
			Conqueror	PT-086	8.30%
Dominate	5	8.38%	Griefer	PT-236	9.10%
			Motivational factor: the need for aggression	PT-188	20.00%
Fight	2	26.65%	Fighter	PT-220	33.30%
Influence	1	14.30%	Influencer	PT-238	14.30%
			Smarty-pants	PT-181	7.70%
			Learn New Skills	PT-138	33.30%
			Mastermind	PT-108	12.50%
Learn	4	15.30%	Inquisitive	PT-027	7.70%
			Daredevil	PT-107	20.00%
			Consumer	PT-232	14.30%
Lone Wolf	3	14.77%	Autonomy	PT-214	10.00%
			Committed mentalities	PT-184	16.70%
			Networker	PT-234	8.30%
			The People Factor	PT-042	7.70%
			Team-sport and Combat	PT-160	12.50%
Relate	5	10.58%	Guardian	PT-083	7.70%

A.4.2 In-Depth View of Graph

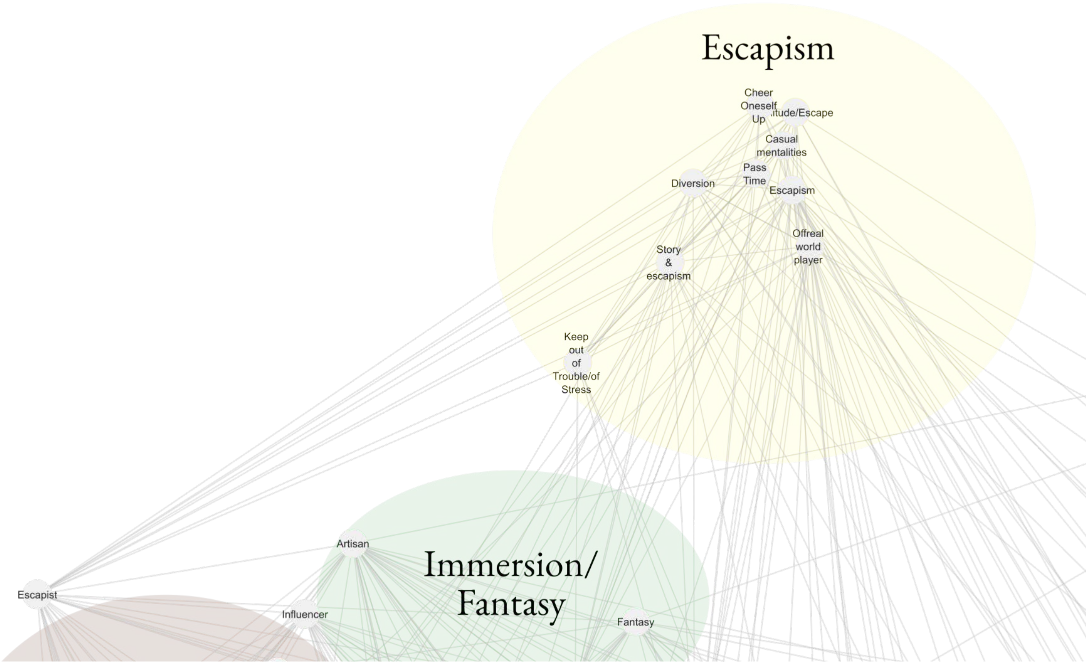


Figure 79 – In-Depth View Escapism Cluster

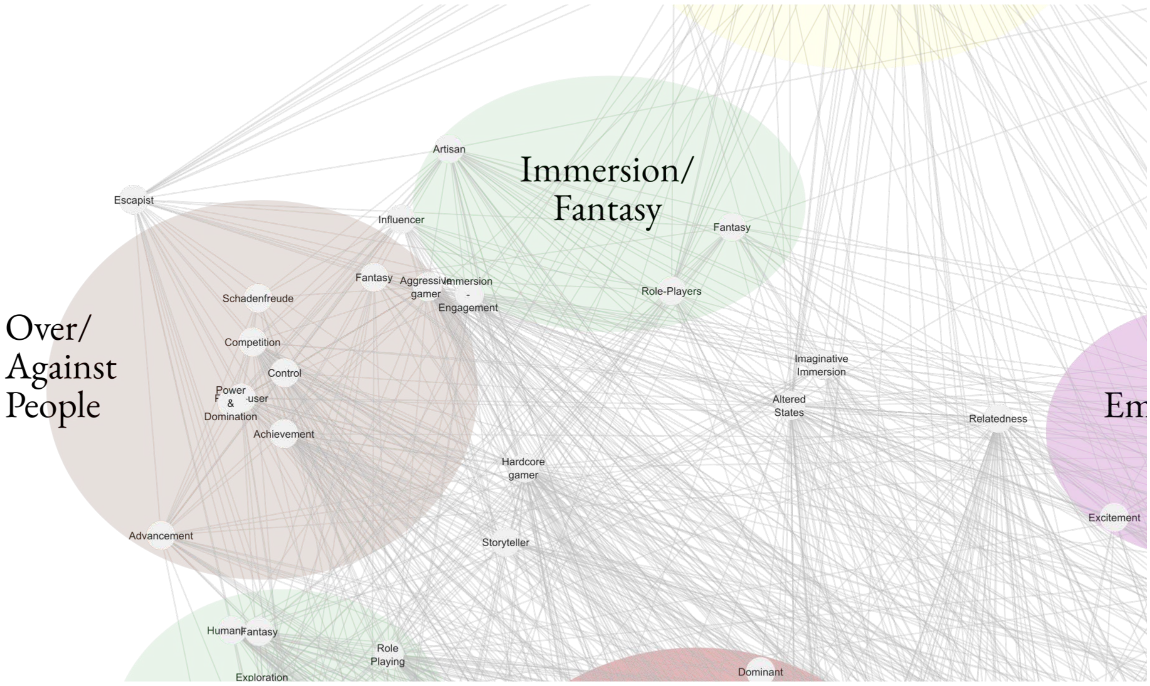


Figure 80 – In-Depth View Over/Against People and Immersion/Fantasy Cluster

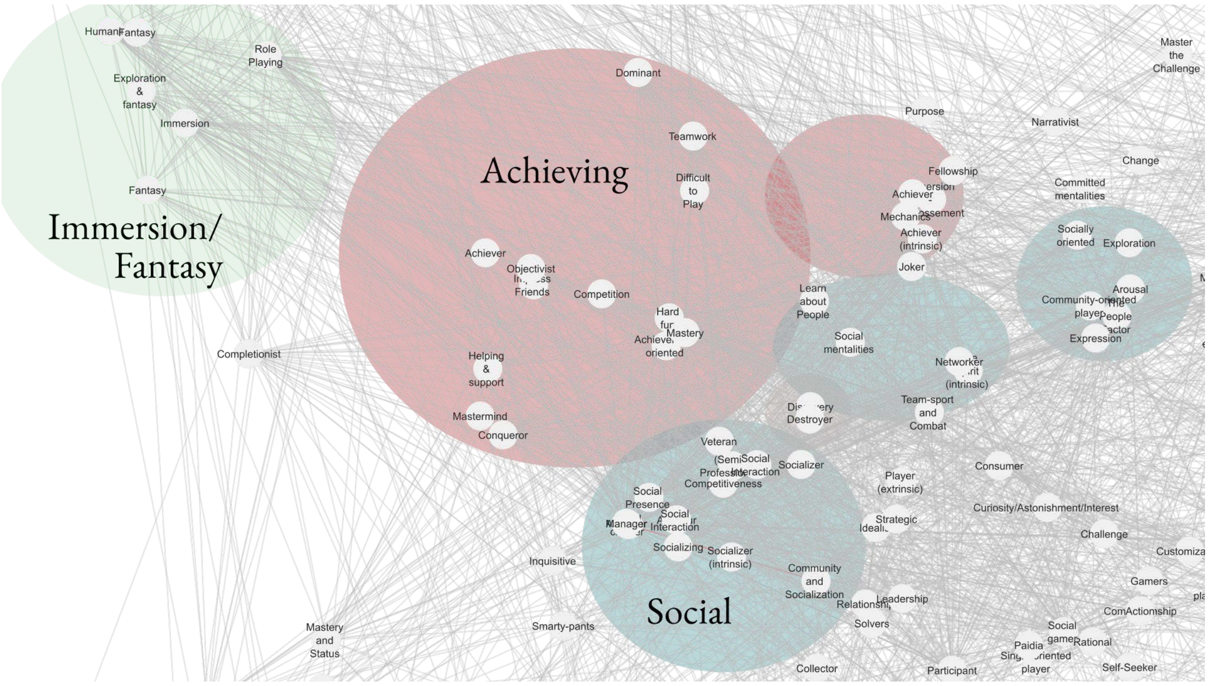


Figure 81 – In-Depth View Achieving, Social and Immersion/Fantasy Cluster

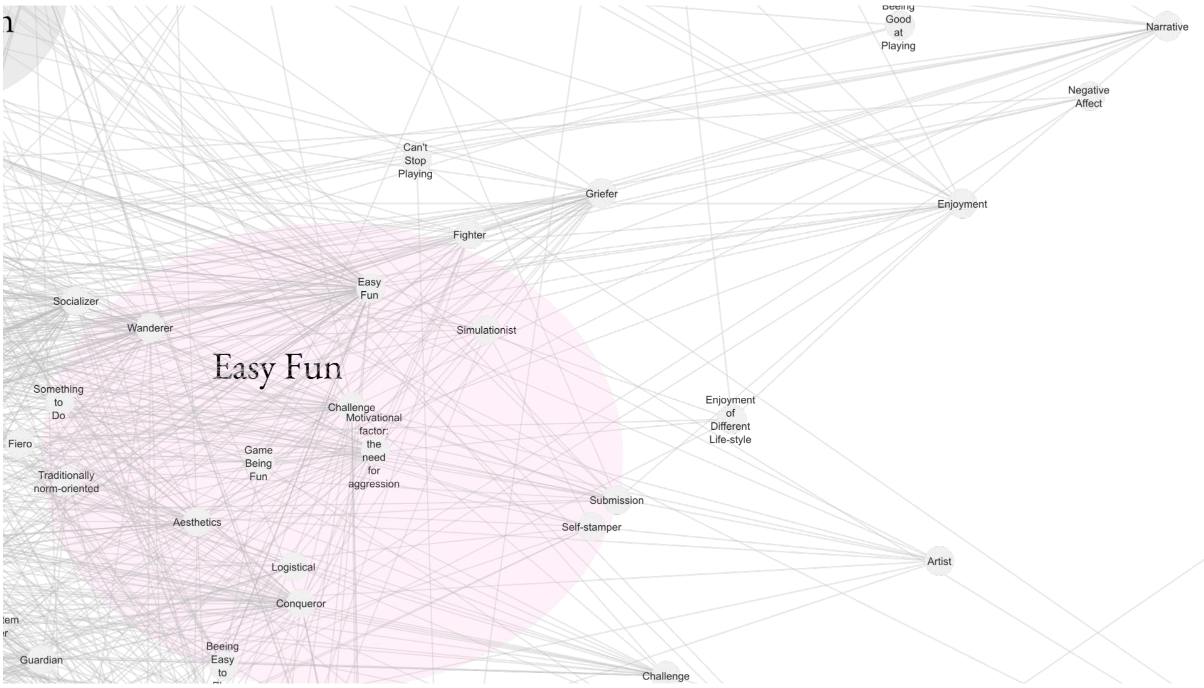


Figure 82 – In-Depth View Easy Fun Cluster

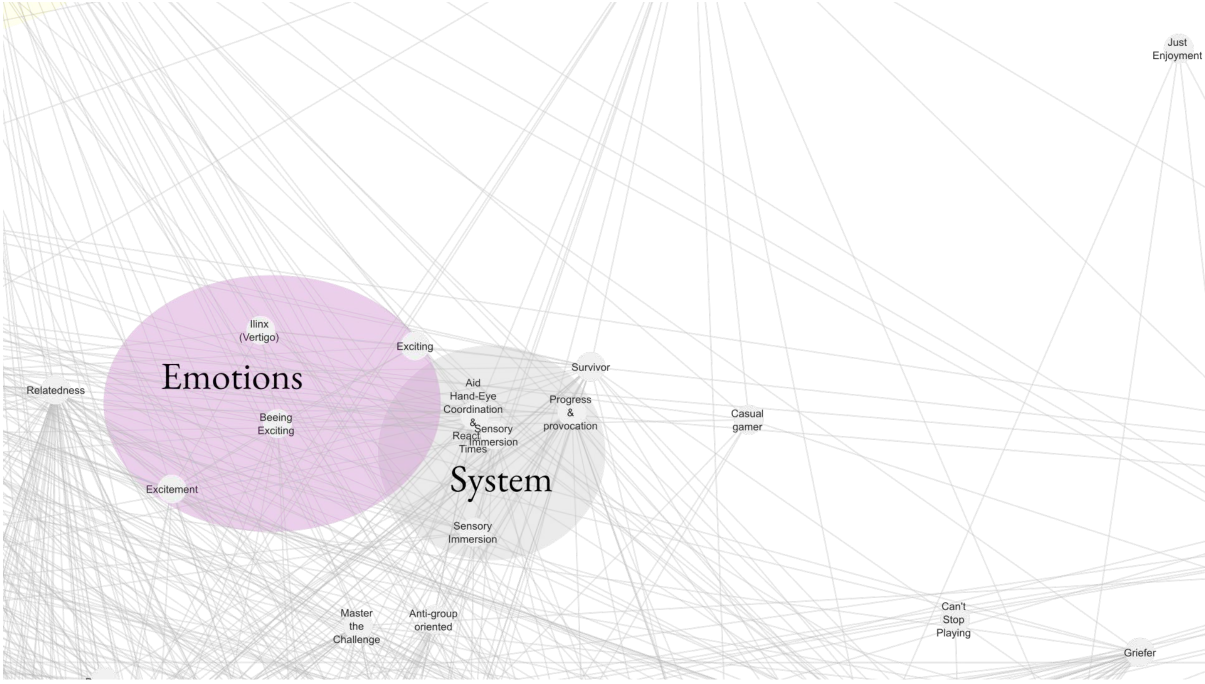


Figure 83 – In-Depth View Emotions and System Cluster

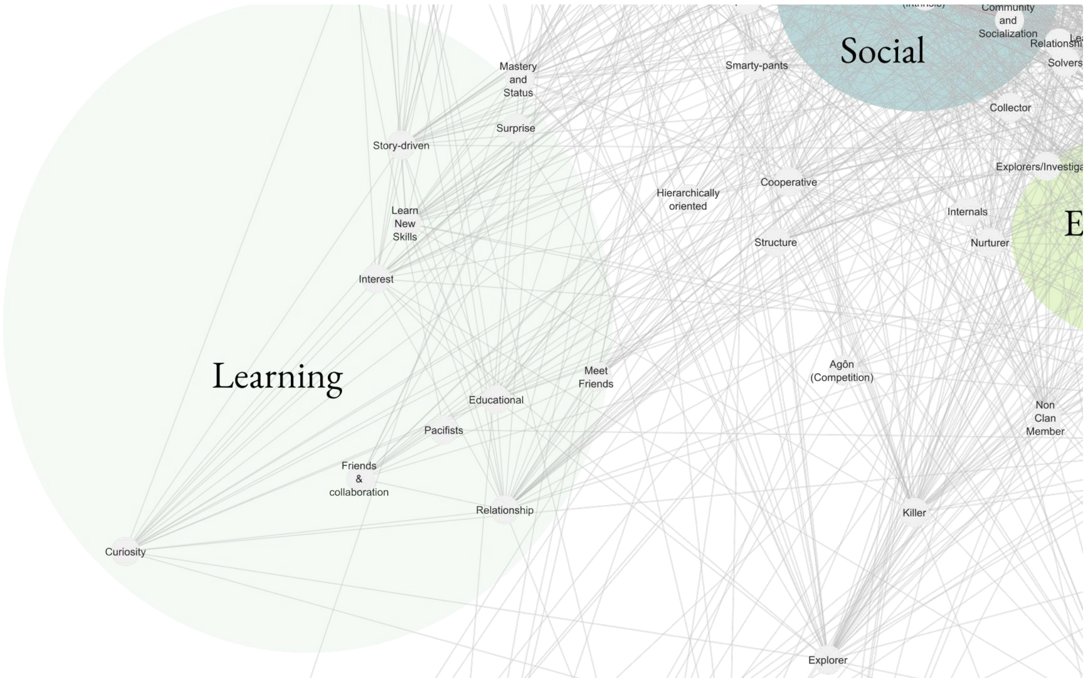


Figure 84 – In-Depth View Learning Cluster

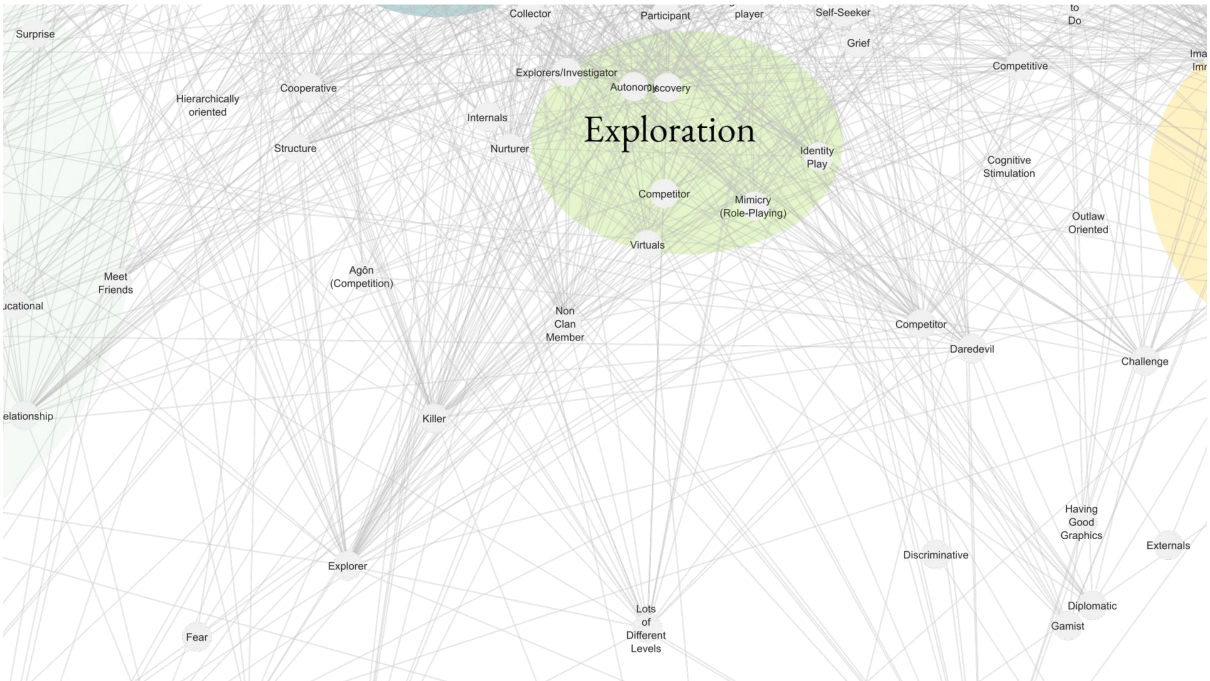


Figure 85 – In-Depth View Exploration Cluster

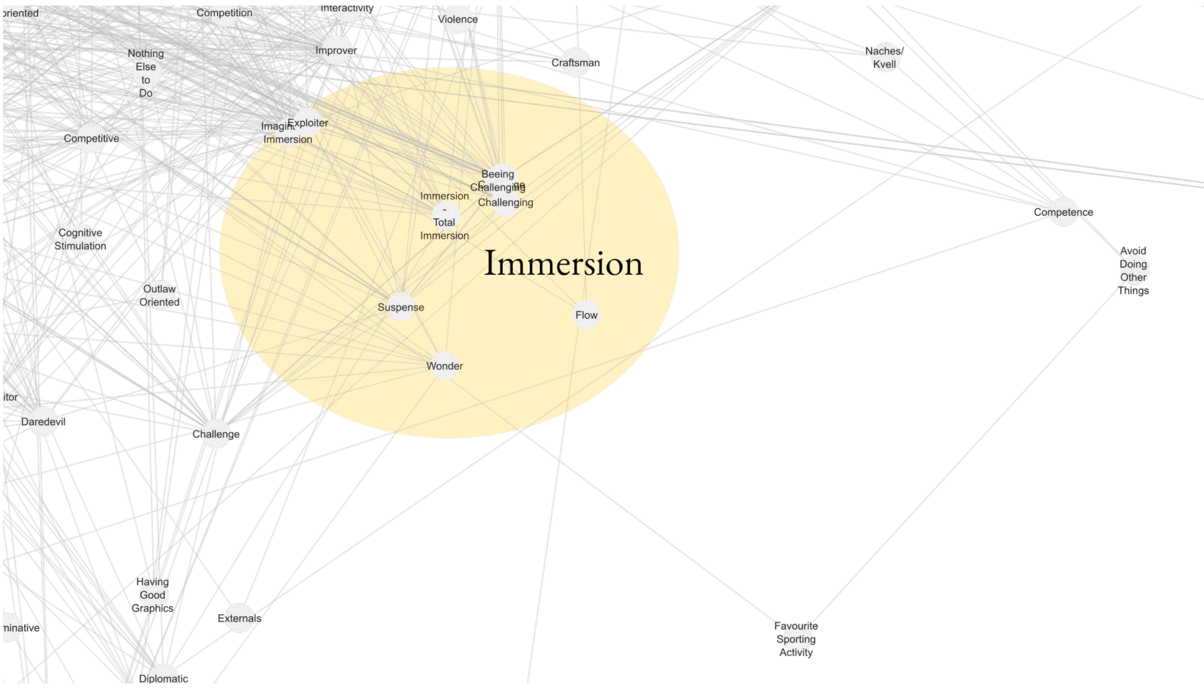


Figure 86 – In-Depth View Immersion Cluster

A.4.3 Additional Data Keyword Matching

Table 48 – List of Nodes Playing Motivations Graph Not Connected to Human Needs

ID	Name Node	Author	Year	Title	Keywords
PT-237	**Destroyer**	Andrzej C. Marczewski	2015	Even Ninja Monkeys Like to Play: Gamification, Game Thinking and Motivational Design	system, desolation, waste, dislike, loophole, destruction, hacking, destroyer, ruin, destroy, break
PT-186	**Casual gamer**	Gabriel Jacobs; Barry Ip	2003	Matching games to gamers with quality function deployment	handling, impatience, gaming, intolerance, amusement, facile, casual, casualness, pleasure
PT-021	**Ilinx (Vertigo)**	Roger Caillois	1961	Man, Play, and Games	exhilaration, frenzy, excitement, perception, vertigo, ilinx, alter
PT-058	**Customization**	Nick Yee	2006	Motivations of Play in Online Games	immersion, accessory, style, customization, color, appearance, scheme
PT-038	**Craftsman**	Tracy Fullerton	2008	Game Design Workshop: A Playcentreic Approach to Creating Innovative Games	building, craft, crafting, craftsman, engineering, puzzle
PT-198	**Hierarchically oriented**	Leo Sang-Min Whang; Geunyoung Chang	2009	Lifestyles of Virtual World Residents: Living in the On-Line Game "Lineage"	system, hierarchy, strictness, command, rank, military
PT-209	**Story & escapism**	Zackariasson, P., Wahlin, N., & Wilson, T. L.	2010	Virtual Identities and Market Segmentation in Marketing in and Through Massively Multiplayer Online Games (MMOGs)	lore, escapism, story, escape, color, scheme
PT-076	**Expression**	Robin Hunicke; Marc LeBlanc; Robert Zubek	2004	MDA: A Formal Approach to Game Design and Game Research	discovery, expression, self, aesthetic, self-discovery
PT-216	**Purpose**	Andrzej C. Marczewski	2013	The intrinsic motivation ramp	mean, philanthropy, reason, purpose, altruism
PT-235	**Disruptor**	Andrzej C. Marczewski	2015	Even Ninja Monkeys Like to Play: Gamification, Game Thinking and Motivational Design	disruption, disrupt, disruptor, chance, trouble
PT-019	**Alea (Chance)**	Roger Caillois	1961	Man, Play, and Games	alea, uncertainty, chance, dice

PT-044	**Surprise**	Nicole Lazzaro	2004	Why We Play Games: Four Keys to More Emotion without Story	relief, brief emotion, sudden, surprise
PT-070	**Sensation**	Robin Hunicke; Marc LeBlanc; Robert Zubek	2004	MDA: A Formal Approach to Game Design and Game Research	sense, aesthetic, sensation, pleasure
PT-074	**Fellowship**	Robin Hunicke; Marc LeBlanc; Robert Zubek	2004	MDA: A Formal Approach to Game Design and Game Research	aesthetic, fellowship, framework, socialize
PT-127	**Good Sound Effects**	Mark D. Griffiths; Nigel Hunt	1995	Computer Game Playing in Adolescence: Prevalence and Demographic Indicators	music, sfx, sound, sound-effect
PT-154	**Enjoyment of Different Life-style**	Alexander E. Voiskounsky; Olga V. Mitina ; Anastasiya A. Avetisova	2005	Psychological research of MUD gamers -Communicative patterns and flow experience of MUD players	different, virtuality, enjoyment, life-style
PT-165	**Flow**	Karolien Poels; Yvonne De Kort ; Wijnand A. Ijsselsteijn	2007	" It is always a lot of fun! " Exploring Dimensions of Digital Game Experience using Focus Group Methodology	concentration, detachment, flow, absorption
PT-183	**Casual mentalities**	Kirsi Pauliina Kallio; Frans MÃ¤yrÃ¤ ; Kirsikka Kaipainen	2011	At least nine ways to play: approaching gamer mentalities	escapism, occupation, time, casual
PT-192	**Socially oriented**	Leo Sang-Min Whang; Geunyoung Chang	2009	Lifestyles of Virtual World Residents: Living in the On-Line Game "Lineage"	other, relation, orientation, socialize
PT-199	**Discriminative**	Leo Sang-Min Whang; Geunyoung Chang	2009	Lifestyles of Virtual World Residents: Living in the On-Line Game "Lineage"	discrimination, age, gender, superficiality
PT-072	**Narrative**	Robin Hunicke; Marc LeBlanc; Robert Zubek	2004	MDA: A Formal Approach to Game Design and Game Research	aesthetic, drama, narrative
PT-075	**Discovery**	Robin Hunicke; Marc LeBlanc; Robert Zubek	2004	MDA: A Formal Approach to Game Design and Game Research	uncharte, discovery, aesthetic
PT-077	**Submission**	Robin Hunicke; Marc LeBlanc; Robert Zubek	2004	MDA: A Formal Approach to Game Design and Game Research	pastime, aesthetic, submission
PT-112	**Pass Time**	Carol A. Phillips; Susan Rolls; Andrew Rouse; Mark D. Griffiths	1995	Home video game playing in schoolchildren: a study of incidence and patterns of play	pastime, escapism, boredom
PT-114	**Cheer Oneself Up**	Carol A. Phillips; Susan Rolls; Andrew Rouse; Mark D. Griffiths	1995	Home video game playing in schoolchildren: a study of incidence and patterns of play	escapism, uplift, cheer
PT-147	**Can't Stop Playing**	Mark D. Griffiths; Nigel Hunt	1998	Dependence on computer games by adolescents	flow, ceaseless, addiction
PT-151	**Skeptics**	Sonja Utz	2000	Social information processing in MUDs: The development of friendships in virtual worlds	skepticism, disinterest, disassociate

PT-167	**Sensory Immersion**	Karolien Poels; Yvonne De Kort; Wijnand A. Ijsselsteijn	2007	" It is always a lot of fun! " Exploring Dimensions of Digital Game Experience using Focus Group Methodology	sensation, immersion, presence
PT-194	**Outlaw Oriented**	Leo Sang-Min Whang; Geunyoung Chang	2009	Lifestyles of Virtual World Residents: Living in the On-Line Game "Lineage"	outlaw, orientation, separation
PT-212	**Narrativist**	Ron Edwards	2004	System Does Matter	narrativist, story, narrative
PT-016	**Excitement**	Jeroen Jansz; Martin Tanis	2007	Appeal of playing First Person Shooter Games	emotion, excitement
PT-123	**Being Exciting**	Mark D. Griffiths; Nigel Hunt	1995	Computer Game Playing in Adolescence: Prevalence and Demographic Indicators	emotion, excitement
PT-142	**Nothing Else to Do**	Mark D. Griffiths; Nigel Hunt	1998	Dependence on computer games by adolescents	boredom, aimlessness
PT-155	**Recreational Refreshment**	Alexander E. Voiskounsky; Olga V. Mitina; Anastasiya A. Avetisova	2005	Psychological research of MUD gamers - Communicative patterns and flow experience of MUD players	refreshment, recreationality
PT-125	**Violence**	Mark D. Griffiths; Nigel Hunt	1995	Computer Game Playing in Adolescence: Prevalence and Demographic Indicators	violence
PT-135	**Exciting**	Mark D. Griffiths; Nigel Hunt	1998	Dependence on computer games by adolescents	excitement

Table 49 – List of Keywords That Are Not Directly Relatable to Playing Motivations

Keyword	Not a PM	Author	Node	Year	Theory
question	Gamelike/-related System	Henry A. Murray	Cognizance	1938	Explorations in Personality
language	Gamelike/-related System	Manfred Max-Neef, Antonio Elizalde, Martin Hopenhayn	Identity	1987	Human scale development: An option for the future
expectancy	Motivation	Victor H. Vroom	Expectancy	1964	Work and Motivation
motivator	Motivation	Frederick Herzberg	Motivator	1959	The Motivation to Work
drive	Motivation	Isabel Briggs Myers, Peter B. Myers	Turbulent	1980	Gifts Differing: Understanding Personality Type
willingness	Motivation	Edwin Locke & Gary Latham	Persistence	1990	A theory of goal setting & task performance
motivational	Motivation	Victor H. Vroom	Force	1964	Work and Motivation
habit	Motivation	Isabel Briggs Myers, Peter B. Myers	Observant	1980	Gifts Differing: Understanding Personality Type
objectivity	Neutral Parameter	Isabel Briggs Myers, Peter B. Myers	Thinking	1980	Gifts Differing: Understanding Personality Type
portability	Neutral Parameter	Victor H. Vroom	Instrumentality	1964	Work and Motivation

capacity	Neutral Parameter	Manfred Max-Neef, Antonio Elizalde, Martin Hopenhayn	Understanding	1987	Human scale development: An option for the future
correlation	Neutral Parameter	Victor H. Vroom	Instrumentality	1964	Work and Motivation
dissatisfaction	Neutral Parameter	Frederick Herzberg	Hygiene	1959	The Motivation to Work
high	Neutral Parameter	Victor H. Vroom	Force	1964	Work and Motivation
secondary	Neutral Parameter	Victor H. Vroom	Secondary outcome	1964	Work and Motivation
primary	Neutral Parameter	Victor H. Vroom	Primary outcome	1964	Work and Motivation
satisfaction	Neutral Parameter	- Frederick Herzberg - Victor H. Vroom	- Hygiene - Valence	1959 1964	- The Motivation to Work - Work and Motivation
extroverted	Personality	Isabel Briggs Myers, Peter B. Myers	Extroverted	1980	Gifts Differing: Understanding Personality Type
agreeable	Personality	Gordon W. Allport, Henry S. Odbert	Agreeable/ Disagreeable	1936	Trait-names: A psycho-lexical study
pragmatic	Personality	Isabel Briggs Myers, Peter B. Myers	Observant	1980	Gifts Differing: Understanding Personality Type
psychological	Personality	Abraham Maslow	Psychological	1943	A Theory of Human Motivation
realistic	Personality	Isabel Briggs Myers, Peter B. Myers	Observant	1980	Gifts Differing: Understanding Personality Type
vulnerable	Personality	Steven Reiss	Acceptance	2001	Who am I
even-tempered	Personality	Isabel Briggs Myers, Peter B. Myers	Assertive	1980	Gifts Differing: Understanding Personality Type
extraversion	Personality	- Isabel Briggs Myers, Peter B. Myers - Gordon W. Allport, Henry S. Odbert	- Extroverted - Introversion/ Extraversion	1980 1936	- Gifts Differing: Understanding Personality Type - Trait-names: A psycho-lexical study
confident	Personality	Isabel Briggs Myers, Peter B. Myers	Assertive	1980	Gifts Differing: Understanding Personality Type
spontaneous	Personality	Isabel Briggs Myers, Peter B. Myers	Prospecting	1980	Gifts Differing: Understanding Personality Type
ignorance	Personality	Henry A. Murray	Rejection	1938	Explorations in Personality
neuroticism	Personality	Gordon W. Allport, Henry S. Odbert	Neuroticism	1936	Trait-names: A psycho-lexical study
boldness	Personality	Manfred Max-Neef, Antonio Elizalde, Martin Hopenhayn	Creation	1987	Human scale development: An option for the future
insecure	Personality	Steven Reiss	Acceptance	2001	Who am I
down-to-earth	Personality	Isabel Briggs Myers, Peter B.Myers	Observant	1980	Gifts Differing: Understanding Personality Type
shy	Personality	- Steven Reiss - Isabel Briggs Myers, Peter B.Myers - Gordon W. Allport, Henry S. Odbert	- Acceptance - Introverted - Introversion/ Extraversion	2001 1980 1936	- Who am I - Gifts Differing: Understanding Personality Type - Trait-names: A psycho-lexical study
temper	Personality	Isabel Briggs Myers, Peter B. Myers	Assertive	1980	Gifts Differing: Understanding Personality Type

receptiveness	Personality	Manfred Max-Neef, Antonio Elizalde, Martin Hopenhayn	Participation	1987	Human scale development: An option for the future
intuitive	Personality	Isabel Briggs Myers, Peter B. Myers	Intuitive	1980	Gifts Differing: Understanding Personality Type
conservative	Personality	Gordon W. Allport, Henry S. Odbert	Openness to Experience	1936	Trait-names: A psycho-lexical study
narrow-minded	Personality	Gordon W. Allport, Henry S. Odbert	Openness to Experience	1936	Trait-names: A psycho-lexical study
introversion	Personality	Isabel Briggs Myers, Peter B.Myers Gordon W. Allport, Henry S. Odbert	Introverted Introversion/ Extraversion	1980 1936	- Gifts Differing: Understanding Personality Type - Trait-names: A psycho-lexical study
open-mindedness	Personality	Manfred Max-Neef, Antonio Elizalde, Martin Hopenhayn	Freedom	1987	Human scale development: An option for the future
neurotic	Personality	Gordon W. Allport, Henry S. Odbert	Neuroticism	1936	Trait-names: A psycho-lexical study
sex	Physical Need	Steven Reiss	Romance	2001	Who am I
body	Physical Need	Steven Reiss	Physical activity	2001	Who am I
water	Physical Need	Abraham Maslow	Psychological	1943	A Theory of Human Motivation
physical	Physical Need	Steven Reiss	Physical activity	2001	Who am I
eat	Physical Need	Steven Reiss Abraham Maslow	Eating Psychological	2001 1943	Who am I A Theory of Human Motivation
cold	Physical Need	Isabel Briggs Myers, Peter B. Myers	Thinking	1980	Gifts Differing: Understanding Personality Type
air	Physical Need	Abraham Maslow	Psychological	1943	A Theory of Human Motivation
dining	Physical Need	Steven Reiss	Eating	2001	Who am I
existence	Physical Need	Clayton Alderfer	Existence needs	1972	Existence, Relatedness, and Growth; Human Needs in Organizational Settings
hygiene	Physical Need	Frederick Herzberg	Hygiene	1959	The Motivation to Work
animalistic	Physical Need	Frederick Herzberg	Animalistic needs	1959	The Motivation to Work
need	Physical Need	- Abraham Maslow - Clayton Alderfer - Clayton Alderfer	- Psychological - Existence needs - Growth needs	1943 1972 1972	- A Theory of Human Motivation - Existence, Relatedness, and Growth; Human Needs in Organizational Settings - Existence, Relatedness, and Growth; Human Needs in Organizational Settings

A.4.4 Suggestions for New Player Types derived from Human Needs

Table 50 – Suggestions for New Player Types Derived From Unconnected Human Needs Keywords

Keywords	Derived Playing Motivation	Derived Player Type
steal	allurement of prohibition	**Scoundrel**
luxury, royalty, celebrity, extravagance, prestige	the high end of the hierarchy	**Aristocrat**
desirability, attraction, self-image	attention seeking	**Narcissos**
principles, duty, patriotism, idealism, justice, values, honor	intrinsic values	**Hero**
humorous, amuse, cheerful, happiness, laugh	cheer seeking	**Happy Fool**
analyze, retention, intelligence, cognizance	cognitive satisfaction	**Brainee**
culture, language, questions	philosophic/ aesthetic	**Connaisseur**
stability, traditions, conservative	conservative values	**Conservativist**
imaginative, inventiveness	innovative creation	**Creator/Inventor**
truth-seeking, inspect, observant, observation	seeking and solving	**Detective**
produce	observing growth	**Gardener**
thrilling, hatred, shocking, passion, sensuality, love, turbulent	strong emotion	**Hedonist**
self-reliant, tranquility, solitary, observant, observation	pleasant solitude	**Lone Wolf**
improvise, spontaneity	flexible mind	**JackofallTrades**
explain	love of lecturing	**Teacher**
consistency, arrange, tidy, closure, perfectionistic, clean, thoroughly	aiming for a perfect state	**OCDist**
spirituality, transcendence, self-transcendence, ritual, religion, clarity, belief	aiming for insight	**Recreational Self-Seeker**
resistance, refuse	righteous disobedience	**Freedom Fighter**
obey, serve, masochism, abasement, surrender, assertive	self-abandonment	**Submittor**
self-dramatization	dramatic expression	**Histrionist**
code	logically constructing	**Coder**
behave, well-behaved	changing behavior in others	**Trainer**
Judge, justice, decisive, choose	meaningful choices	**Judge**
confession, relax, relaxed, stress-avoiding	unburdening	**Reliever**
body, health, exercise, fitness, well-being	encouraged fitness	**Tailwinder**

Appendix – B

Supplementary Material Chapter 4

B.1 Design Evolution of the Digital Artifact

B.1.1 Design Iterations

After having assessed and listed the requirements, we started by sketching the first designs for the overall interaction and transforming them into a screenplay to create a mock-up video showcasing the envisioned interaction (the video can be found in the Supplementary Materials: https://gonku.de/sup-mat-phd-gho/Kubun-Kubun_Demo_Movie.mp4). Figures 87 and 88 show Mockups of this stage of the design.

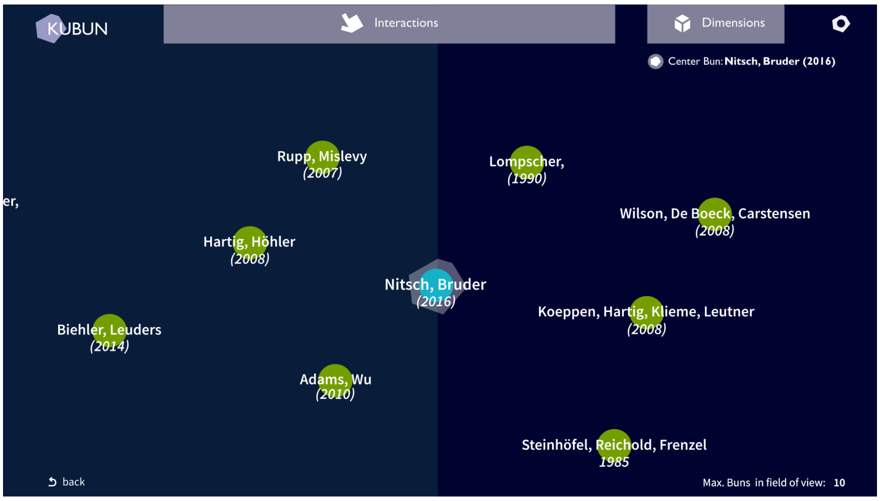


Figure 87 – Mockup Result Overview

Figure 88 –Mockup Detail View for Single Node

This initial design featured a map-based visualization as an overview map with a separate node view representing the modules and an interaction panel separate from the information. It, however, already incorporated the centered search and the weighting interaction.

Based on this mock-up interaction, we changed the design to incorporate the interaction features into the information modules via a side panel, rendering the node-detail view obsolete (see Figure 89).

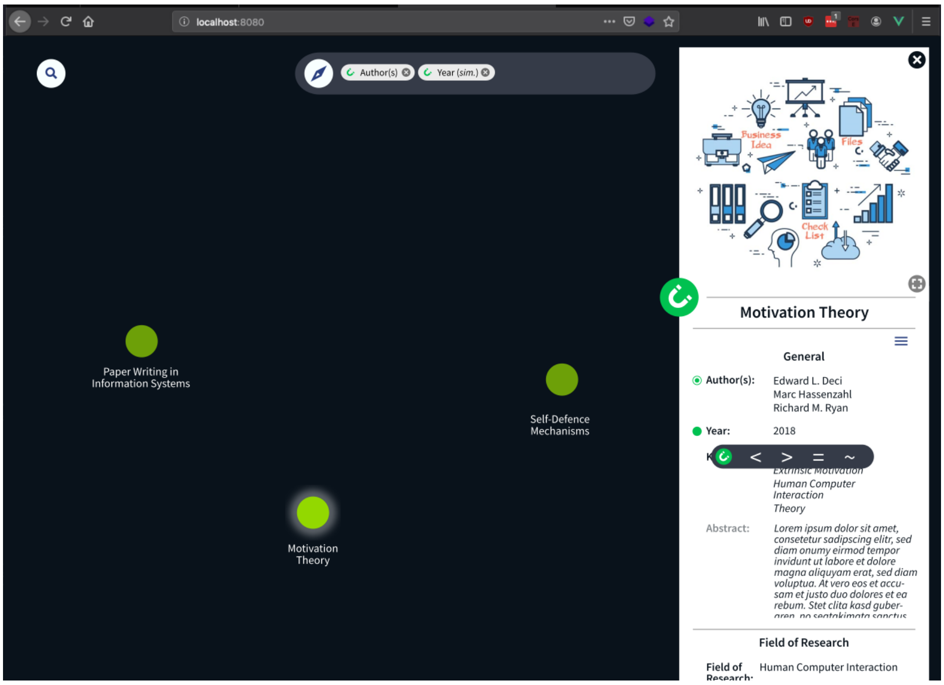


Figure 89 – Mockup Sidepanel-based Interaction

We additionally implemented features to afford quick access to information and flow. First, we added a menu accessible via right-click on a node. The menu allows setting a new node as a center, thus rearranging the result space. It further allows to set a node as a favorite to remember nodes for future reference and finally. While we designed a "hide" function, it was not implemented due to limitations in terms of time and resources (see Figure 90). Another feature we implemented was a panel displaying short information that automatically opens when users hover over a node (see Figure 91). These features afford a streamlined browsing process, as users can quickly assess their interest via the hover panel and recenter on a node they prefer to their current center. Thus, they can traverse the graph seamlessly, recentering each new node that catches their interest.

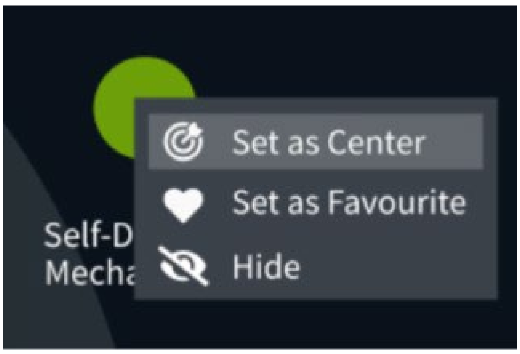


Figure 90 – Set Node as new Center, Set Node as Favourite, Hide Node

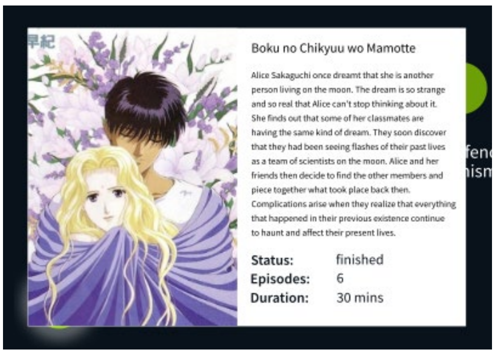


Figure 91 – Hover Panel

B.1.2 Representation in the Field Study

The full interactions with the different social media posts are uploaded in Supplementary Material: https://gonku.de/sup-mat-phd-gho/Kubun-Social-Media-Feedback.pdf.

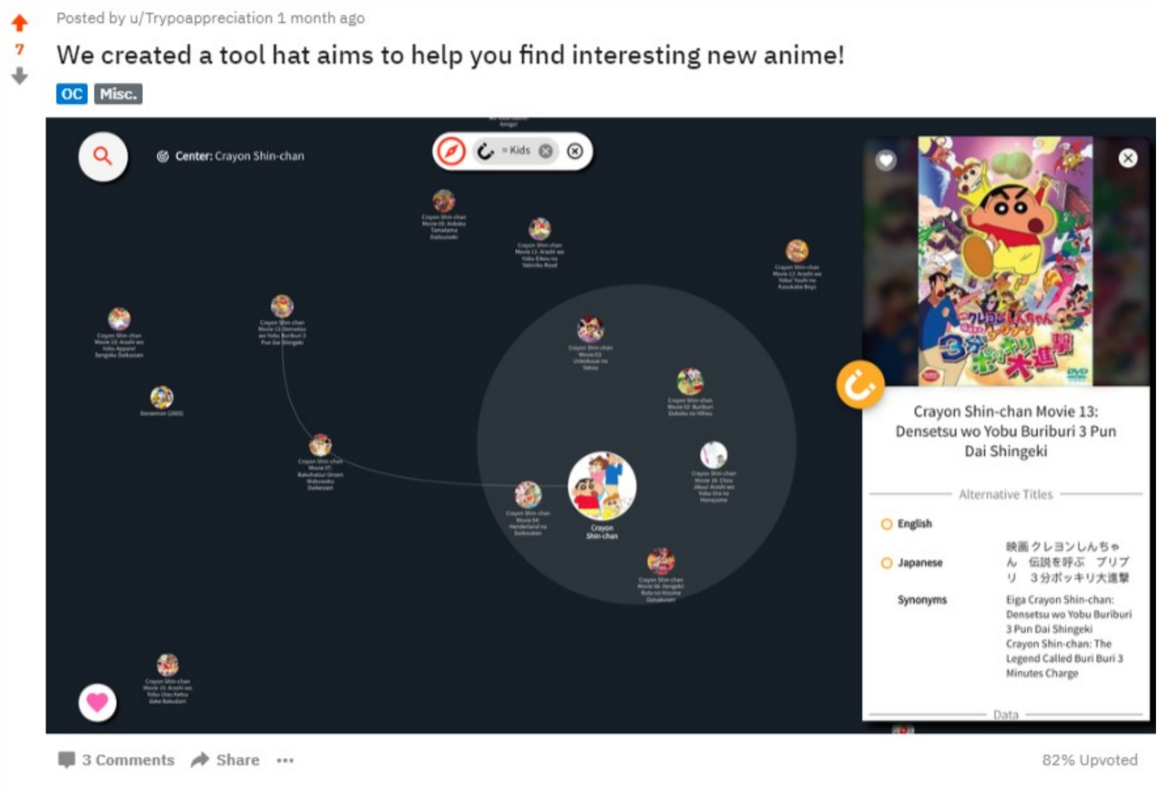


Figure 92 – Screenshot of the Post for the Field Study in the Subreddit r/dataisbeautiful

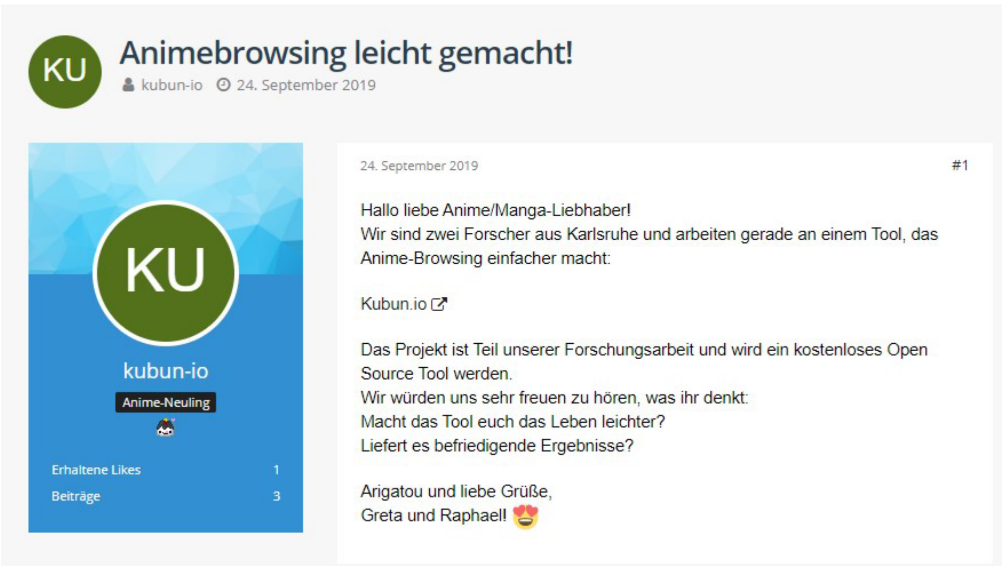


Figure 93 – Screenshot of the Post for the Field Study in the German Anime Forum anime-community.de

Appendix – C

Supplementary Material Chapters 5 & 6

C.1 Supplementary Research Materials Perfect Reward Experiment

C.1.1 Control and Additional Variables

Table 51 – Operationalization of Control and Additional Variables (Experiment Perfect Reward)

Controls	English	Tested with
Age	"How old are you?"	Integer value
Gender	"What is your gender or the gender you identify with? Please press "no answer" if none of the available options represents your gender."	Female/Male
Preference for Intuition and Deliberation (PID-19)	"Please answer all the following questions about your life in general. Your answers should correspond to the way you generally make decisions. Choose the number that best represents your opinion. (1) means that you very much disagree; (5) means that you very much agree."	Likert (five-point): 1-5
Short Version of the Big Five Inventory (BFI-10)	"How well do the following statements describe your personality? Choose the number that best represents your opinion. (1) means that you disagree strongly; (5) means that you agree strongly. I see myself as someone who..."	Likert (five-point): 1-5
The Decision Making Tendency (DmTI-29)-Factors A	"Please answer all the following questions about your life in general. Your answers should correspond to the way you generally make decisions. Part 1 Choose the number that best represents your opinion. (1) means that you very much disagree; (5) means that you very much agree."	Likert (five-point): 1-5
The Decision Making Tendency (DmTI-29)-Factors B	"Please answer all the following questions about your life in general. Your answers should correspond to the way you generally make decisions. Part 2 Choose the number that best represents your opinion. (1) means that you very much disagree; (5) means that you very much agree."	Likert (five-point): 1-5
Self-Evaluation of Playing Behavior	"Finally we'd like to ask you to answer some last questions about the game and your perception of it. Choose the number that best represents your opinion. (1) means that you very much disagree; (5) means that you very much agree."	Likert (five-point): 1-5
	"I constantly tried to get better at the game."	Likert (five-point): 1-5
	"It was important to me not to worsen in terms of performance."	Likert (five-point): 1-5
	"The final result was very important to me."	Likert (five-point): 1-5
	"The Highscore at the end of the game was more important to me than the whole game."	Likert (five-point): 1-5
	"Constant accomplishments weren't important to me."	Likert (five-point): 1-5
	"In the end it is the result that counts."	Likert (five-point): 1-5

Self-Evaluation of Playing Enjoyment	"Playing the game is exciting."	Likert (five-point): 1-5
	"Playing the game gives me a lot of pleasure."	Likert (five-point): 1-5
	"I enjoyed playing the game."	Likert (five-point): 1-5
Self-Evaluation of Usage and Behavior regarding the Learning Enhancing Elements	"The "Trashdex" (Mülldex) played an important role to me."	Likert (five-point): 1-5
	"And why? Please elaborate."	Free text
	"Based on what principle did you decide if you wanted to repeat or not?"	Free text
	"If you played more than the mandatory waves, please elaborate why."	Free text
Evaluation of Design Decisions in the Game	"How did you feel about the speed of the game in general?"	Free text
	"How did you feel about the increase in speed?"	Free text
	"I was satisfied with the visual design of the game."	Free text
	"The feedback of the monsters could have been left out of the game."	Free text
General Feedback	"The experiment is now over. Please stay seated until you get called. We would like to thank you very much for participating! Please feel free to give us some final feedback on the experiment in general and/or the app in particular."	Free text

C.1.2 Additional Analyses

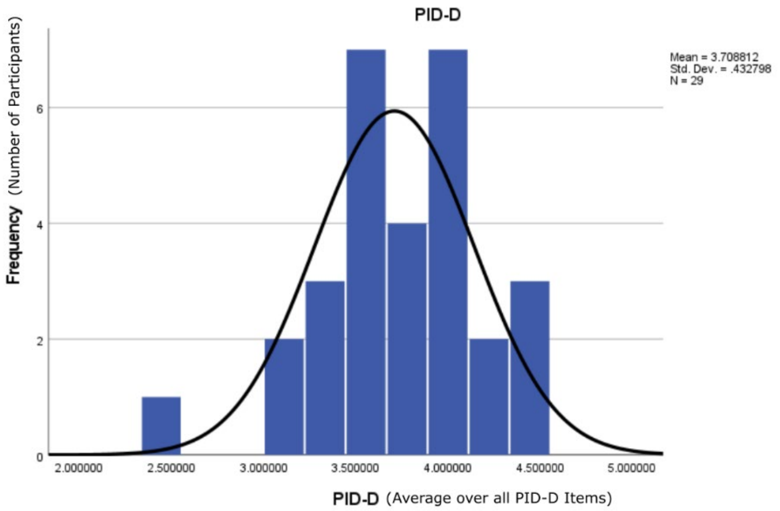


Figure 94 – Distribution of Participants Deliberation

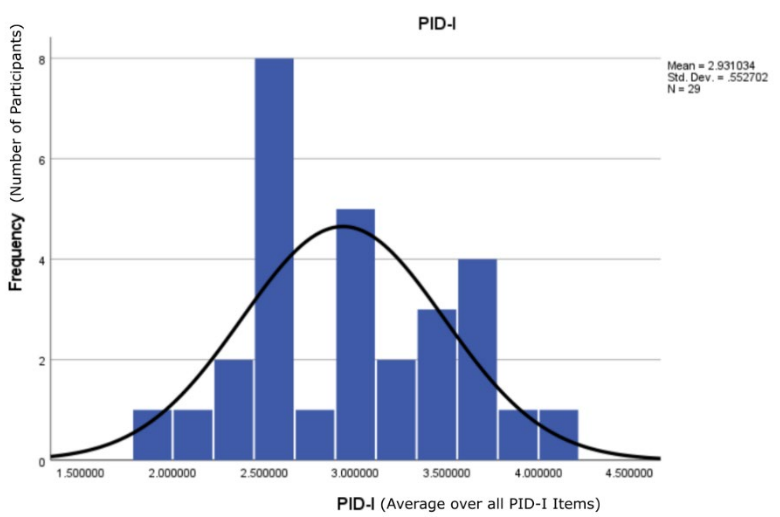


Figure 95 – Distribution of Participants' Scores on Intuition

C.2 Supplementary Research Materials Repetition and Look-Up Experiment

C.2.1 Experimental Procedure – Detailed

P1*:* After closing the registration for participation in the experiment, we randomly assigned participants to one of the five experimental groups in a between-subject design. We sent instructions via email for participants to fill out the first survey and download the app (as an apk-file). The game within was locked by a server to prevent premature play and could only be accessed once the first phase officially started. In the survey, we assessed demographic information (age, gender, how long they had been living in Germany, how long they had been living in the city in which the experiment was conducted), participants' game motivation (how much they were involved in and how they felt about these games in general) and their general waste sorting motivation (how they felt about municipal waste sorting). We also included several controls checking language proficiency and conscientiousness in answering the questions. To ensure absolute anonymity in the datasets when linking the game data to the survey entries, each app showed a unique code that participants had to report in each respective survey. For this phase, we set a 48-hour timeframe followed by a pause of 24 hours that allowed for troubleshooting.

P2: In the second phase, we sent the next set of instructions as well as another survey link via email. We instructed the participants on the four game-based treatments to open the application, to play it through to the end, and then complete the survey. In contrast, we told the control group with the non-interactive materials to attentively read through the teaching materials provided through the link for 25 minutes (this time was derived from the average playtime of the experimental version of the game during the pre-tests) and then complete the survey. The last part of the survey was the same for all treatments: we measured the perceived usability of the application—or the materials in the case of the non-game material treatment—with the system usability scale (Brooke, 1996) as well as self-stated perceived growth in competency and growth in motivation. To adapt the 30 minutes of focused attention to the survey and training, we gave participants a four-day timeframe—including a weekend—to finish the task. We scheduled the final sessions 10-12 days after the deadline for the second phase, depending on the day of the assigned session.

P3: The experiment took place in a laboratory in 19 experimental sessions. Each participant was seated in a cabin where they were guided through the first part of the experiment with the final survey. We first asked participants about their perceived growth in competency and growth in motivation, and there was a final control question on any prior knowledge about the project. Next, we tested the learning outcome in three different performance measures. First, the participants completed a multiple-choice test in which they had to match all 108 trained waste items. Second, we asked all participants to take their phones and start the game application, where they had to sort all 108 items in a special version of the game. Here, each item appeared only once in one big game wave without the two additional design elements. Third, we called the participants into a separate room, where we asked them to sort a selection of real-life waste items. The design of the experimental procedure was pre-tested with seven participants.

C.2.2 Non-Game Materials

Four Bins - Examples

Residual waste		Recycleables	Biowaste*	Paper | Cardboard
Ring binder, plastic	Nylon tights	**Wood, untreated, like:**	Balcony plants	**Recycled paper like:**
Ashes - packed	Camera lenses	Wooden boards, Fruit crate	Banana peels	Ring binder - cardboard
Baking/grease-proof paper	Paper, very soiled or imbued		Food waste bin liners	Envelopes, with and without viewing panel
Eye glasses, broken	Paper towels & tissues, soiled	**Recyclables, like:**	Bread	Brochures
Sanitary pads	Parchment Paper	CDs	Eggshells	Books
Photographic slides	Sticky plaster	Bucket - emptied	Fish offal	Egg boxes
Floppy disks	Paintbrush	Plastic crockery	Offal	Wrapping paper, uncoated
Extractor fan filter	Porcelain	Bottles, canisters	Vegetable peel	Notebooks
Bicycle saddle	Dolls	Plastic film, plastic bags	Hair	Cardboard boxes
Pelts / Skins	Cleaning rags	Childrens toys	Burlap	Catalogues
Binoculars	Eraser	Mixing bowl	Coffee grounds	Magazines
Heat-proof glass	Razor blades	Styrofoam (sundries in transparent bags)	Cheese residues	Paper - loose
Lighter, empty	Roller skates		Bones	Papertowels - if only slightly moist
Photographic film	Soot - packed	**Metals, like:**	Dead parts of plants	Paper packaging
Felt-tip markers, dried out	Shoes - unusable	Antenna	Nutshells	Cardboard
Photographs	Napkins - used	Baking Dish, metal	Fruit waste	Posters
Fountain pen, empty	Skateboard	Sheet metal	Orange peel	Brochures
Doormat	Mirror glass	Cans - emptied	Seeds	Writing paper
Garden hose	Syringes, safely packed	Electric cable	Cut flowers	Packaging from paper, cardboard, carton
Gift wrap, coated	Vacuum cleaner bag	Crown cap	Left-overs, raw (no soups and sauces)	Advertisements - printed
Light bulbs	Tampons	Brass keys	Tea bags	Journals
Rubber materials	Wallpaper leftovers	Pans, pots	Potted plants	Newspapers
Suspenders	Electric torch, without batteries	Tool parts	Rotting food, wihout packaging	
Inline skates	Pieces of carpet, chopped up		Sausage and processed meat leftovers	**Recycled paper can also be donated to the wastepaper collection.**
Cassettes (audio | video)	Thermal paper	**Aluminium, like:**	Lemon peel	
Chewing gum	Thermos flask	Cling film		**Tipps:** Remove the tape from your shipping boxes, to fold them in a space-saving manner. Carton packaging can equally be folded flatly. Avoid unnecessary paper waste e.g. by using an advertising ban sticker on your mailbox.
Sweepings	Animal bedding	Yoghurt pot lids	**Biowaste that does not emerge from households has to be disposed of commercially.**	
Ceramics, no sanitary ceramics	Clocks - no batteries	Chocolate foil		
Candle stubs	Wound dressing	Tubes, no contaminants, emptied	*if not available: self-composting	
Sticky tape	Packaging, strongly soiled			
Sticky labels	Hot-water bottle - gummi	**Composite packaging like:**		
Carbon paper	Absorbent cotton	Blister packs		
Ball pen refill	Wicker basket	Milk carton - emptied		
Pleather	Diapers	Paper bag with synthetic padding		
Cuddly toys	Cigarette ends	Juice cartons - emptied		
Leatherbags, -belts	Ignition plugs	Vacuum packaging		
Left-over linoleum				
Airbed		**Packaging, scraped clean, from:**		
Rags		Wood, plastic, metal		
Crayons, solvent-free				
Left-over bits of food				

Some waste types are collected via alternative waste disposal facilities like recycled glass-, used textiles-, organic waste containers, composting facilities, recycling stations, bulk waste collections as well as contaminant collection facilities.

Electrical and electronic appliances up to 50 centimeters edge length can be dropped free of charge at all recycling stations.

Large appliances can also be dropped free of charge at all recycling stations. Large electrical appliances (ovens, stoves, cooling- and freezing appliances, dryers, washing machines) will be picked up on demand. Small equipment can be placed adjacently free of charge.

Figure 96 – Flyer on Waste Categories in Karlsruhe

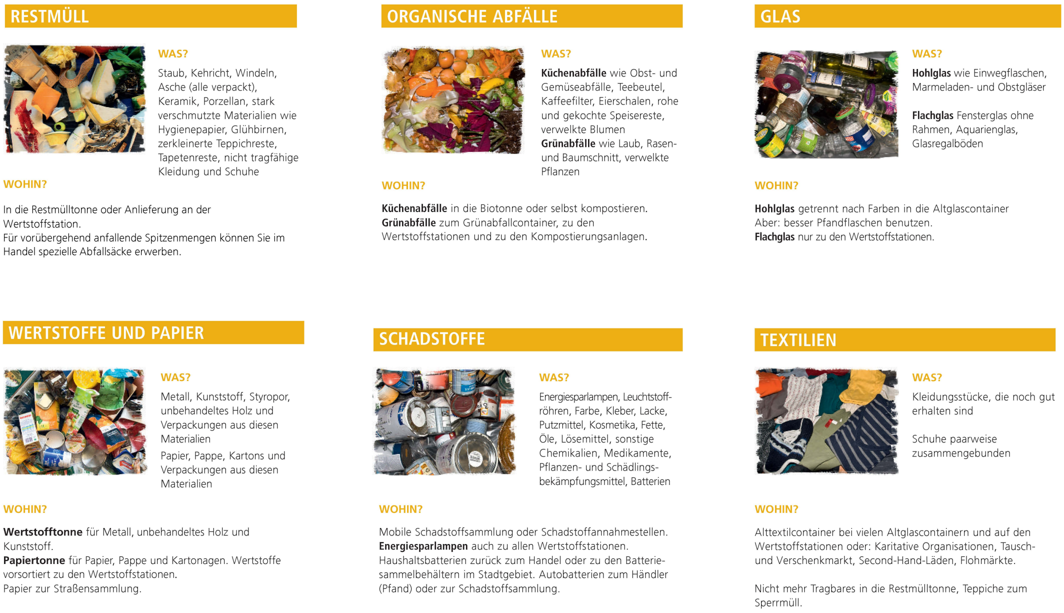


Figure 97 – Flyer on General Waste Sorting in Karlsruhe

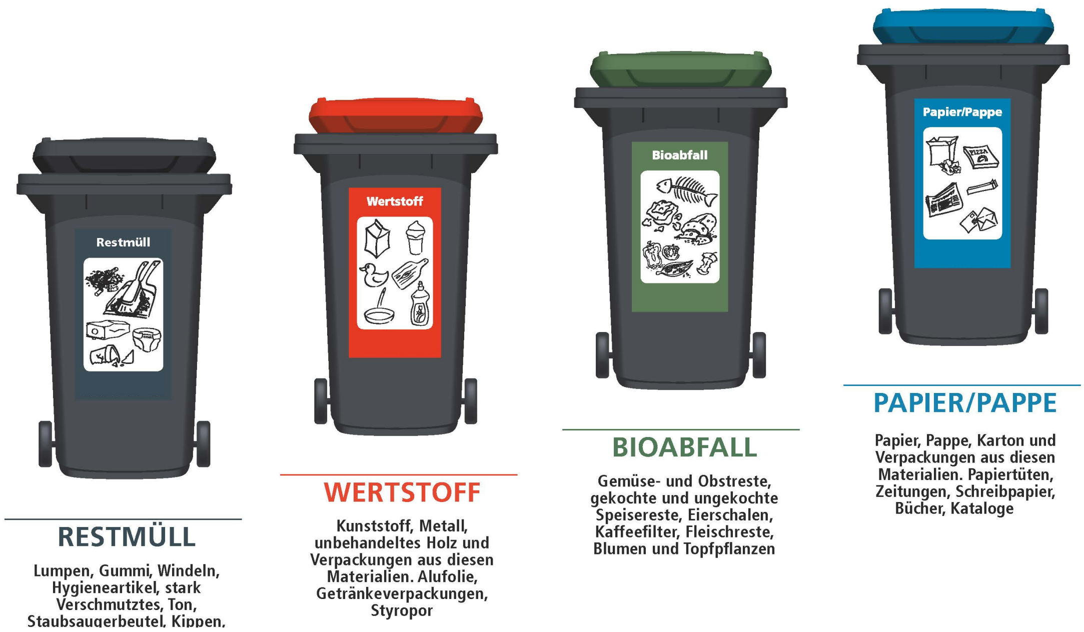


Figure 98 – Flyer on Bins and Representative Waste Items in Karlsruhe

C.2.3 Additional Literature Overviews

Table 52 – Literature Comparison Between Errorful (EF) and Errorless (EL) Learning

Authors	Context	Subjects	Conclusion
Baddeley and Wilson (1994)	Clinical	16 people with brain injuries and memory impairment, and 16 young and older controls each	EL is better than EF
Clare et al. (1999)	Clinical	One participant with Alzheimer's disease	EL is effective and useful for memory problems
Clare and Jones (2008)	Clinical	Six participants with early-stage DAT	EL is effective and useful for memory problems
Donaghey et al. (2010)	Clinical	30 people with an amputated limb, randomly assigned to either the experiment or control group	EL is better than EF
Dunn and Clare (2007)	Clinical	10 people with different conditions	No difference
Evans et al. (2000)	Clinical	Phase 1: 18 people with brain injuries and memory impairment. Phase 2: 16 people with brain injuries and memory impairment. Phase 3: 34 people with brain injuries and memory impairment	Mixed results but overall better performance with EL
Hunkin et al. (1998)	Clinical	Eight people with memory impairment	EL is better than EF
K. Ivancic and Hesketh (2000)	Driving Education	Experiment 1: 44 people in two equal groups Experiment 2: 32 people in two equal groups	EF is better than EL
Johnson (2004)	Learning Strategies	Evidence aggregation of different studies	EF is better than EL
Jones and Eayrs (1992)	Teaching Strategies	Literature synopsis	Inconclusive
Kessels and Haan (2003)	Natural Ageing	18 elderly and 16 young controls	EL is better than EF
Kessels et al. (2007)	Clinical	10 people with Korsakoff Syndrome	No difference
Ohlsson (1996)	Learning Strategies	Tests on the evaluation of own performance errors—more theoretical	Inconclusive
Prather (1971)	Airforce Education	96 people	EF and EL are similarly effective
Page et al. (2006)	Clinical	Experiment 1: 23 people with memory impairment and 20 controls Experiment 2: 20 people with memory impairment	EL is better than EF
Tailby and Haslam (2003)	Clinical	24 people in three groups of eight, each with different severity of memory impairment	EL is better than EF

C.2.4 Control and Additional Variables

Table 53 – Operationalization of Control and Additional Variables (Experiment Repeat & Look-Up)

Controls	English	German	Tested with
Age	"Please tell us your age"	*„Bitte teile uns Dein Alter mit."*	Integer value
Gender	"Which gender do you identify with?"	*„Welchem Geschlecht ordnest Du Dich zu?"*	Male/female/other *Männlich/ Weiblich/ Sonstiges:*
Living in	How long have you been living in	*„Wie lange wohnst Du schon in*	Integer value

Germany	Germany? Please answer with number of full years.	*Deutschland? Bitte antworte in ganzen Jahren."*	
Living in XX City	How long have you been living in XX? Please answer with number of full years.	*„Wie lange wohnst Du schon in XX? (Bitte antworte in ganzen Jahren)"*	Integer value
Game motivation (medium acceptance)	Please tell us about your attitude towards games.	*„Bitte teile uns Deine Einstellung gegenüber Gaming mit."*	(sub-headline)
	I play videogames (computer games, smartphone games, console games, ...) in my free time.	*„Ich spiele in meiner Freizeit Videospiele (Computerspiele, Handygames, Konsolenspiele,...)."*	Likert (five-point): Strongly disagree, rather disagree, neither agree nor disagree, rather agree, strongly agree *Stimme gar nicht zu, stimme eher nicht zu, teils-teils, stimme eher zu, stimme voll und ganz zu*
	I am prejudiced towards grown-ups who play videogames. (r)	*„Ich habe Vorurteile gegenüber erwachsenen Menschen, die Videospiele spielen."(r)*	
	I wish videogames were more accepted in society.	*„Ich wünschte, Videospiele würden eine höhere Akzeptanz in der Gesellschaft genießen."*	
	I think videogames are a waste of time. (r)	*„Ich denke, dass Videospiele eine Form der Zeitverschwendung sind."(r)*	
	Videogames are my hobby.	*„Videospiele sind mein Hobby."*	
	I feel that too much attention is spent on videogames. (r)	*"Ich finde, dass man Videospielen zu viel Aufmerksamkeit schenkt."(r)*	
General waste sorting motivation (general interest in the topic)	What is your attitude towards waste sorting at home? Please answer honestly.	*„Wie ist Deine Einstellung zu Mülltrennung? Bitte antworte ehrlich."*	Likert (five-point) Fully applicable, rather applicable, partly applicable, rather not applicable, not applicable *trifft voll zu, trifft eher zu, teils-teils, trifft eher nicht zu, trifft nicht zu*
	I have never given any thought to waste sorting.	*„Ich habe mir noch nie über Mülltrennung Gedanken gemacht."*	
	Waste sorting at home is very important to me.	*„Mir ist Mülltrennung im Haushalt sehr wichtig."*	
Waste sorting motivation and competency	Please let us know to what extent you agree with the following statements.	*„Bitte teile uns mit, inwiefern Du den folgenden Aussagen zustimmst"*	Likert (five-point) Strongly disagree, rather disagree, neither agree nor disagree, rather agree, strongly agree *Stimme gar nicht zu, stimme eher nicht zu, teils-teils, stimme eher zu, stimme voll und ganz zu*
Waste sorting motivation: last two weeks	Since part 2 of the experiment, have you been more motivated to correctly sort your waste?	*Warst Du seit Teil 2 des Experimentes motivierter, Deinen Müll korrekt zu trennen?*	
Waste sorting motivation: from now on	Since part 2 of the experiment, have you felt more skilled at correctly sort your waste?	*Hast Du Dich seit Teil 2 des Experimentes kompetenter darin gefühlt, Deinen Müll richtig zu trennen?*	
Waste sorting competency: last two weeks	After participating in this experiment, do you feel more motivated to correctly sort your waste from now on?	*Bist Du nach Abschluss dieses Experiments motivierter, ab jetzt Deinen Müll korrekt zu trennen?*	
Waste sorting competency: from now on	After participating in this experiment, do you feel more skilled at correctly sort your waste from now on?	*Fühlst Du Dich nach Abschluss dieses Experiments kompetenter darin, Deinen Müll ab jetzt richtig zu trennen?*	
SUS	See Brooke (1996).	See Brooke (1996)	Likert-based five-point See Brooke (1996)

(r) refers to the questions being reverse-coded

Table 54 – Control Variables - Descriptive Statistics (Experiment Repeat & Look-Up)

	Mean	Std. Dev.	Min	Max	Scale/Type of Measure
Age	22.72	3.01	17	41	Age in years (integer values)
Living in Germany	20.73	5.90	0	30	Number of years (integer values)
Living in XX City	4.28	5.46	0	28	Number of years (integer values)
Gaming motivation	3.12	.90	1.17	5	Likert five-point (six items, three reverse-coded)
General waste sorting motivation	4.23	.80	1.5	5	Likert five-point (two items)
SUS	78.79	12.93	30	100	SUS score: map answers (Likert five-point) from 0 (lowest) to 4 (highest), add the values of all 10 items and multiply by 2.5

Table 55 – Control Variables – Descriptive Statistics per Treatment (Experiment Repeat & Look-Up)

	Non-Game Material		Repeat Element		Look-up Element		Combined		Core Gameplay	
	mean (min/max) / percent for gender	std. dev.	mean (min/max)	std. dev.	mean (min/max)	std. dev.	mean (min/max)	std. dev.	mean (min/max)	std. dev.
Age	23.28 (19/30)	3.28	22.6 (18/41)	3.76	22.42 (17/28)	2.46	23.34 (18/32)	2.83	22.09 (19/30)	2.47
Gender (male)	71.8%		65.2%		75.6%		63.4%		54.5%	
Gender (female)	28.2%		32.6%		24.4%		36.6%		45.4%	
Gender (diverse)			2.2%							
Living in Germany	21.85 (3/30)	5.46	20.22 (2/28)	5.33	21.13 (1/28)	5.48	21.34 (3/28)	5.64	19.30 (0/30)	7.27
Living in XX City	4.08 (0/28)	5.60	4.82 (0/23)	6.13	3.77 (0/22)	4.00	4.46 (0/27)	5.74	4.26 (0/27)	5.77
Gaming motivation	3.25 (1.17/5)	1.05	3.04 (1.17/5)	.89	3.21 (2/4.83)	.80	3.03 (1.33/4.67)	.88	3.07 (1.17/4.67)	.88
General waste sorting motivation	4.13 (1.5/5)	.92	4.23 (2/5)	.74	4.13 (2/5)	.84	4.44 (2.5/5)	.64	4.23 (2/5)	.84
SUS	76.73 (32.5/95)	14.13	78.91 (47.5/95)	11.91	75.44 (45/97.5)	13.86	81.59 (42.5/100)	10.81	81.31 (30/100)	13.13

C.2.5 Additional Analyses

Table 56 – Effect of the Game in Comparison with the Non-Game Material Group with Control Variables (Experiment Repeat & Look-Up)

	In-Game Performance		Multiple-Choice Test		Real-Life Sorting	
Reference category: Non-game material	coef. (bootstr. std. error) [conf. interval]	p (two-tailed)	coef. (bootstr. std. error) [conf. interval]	p (two-tailed)	coef. (bootstr. std. error) [conf. interval]	p (two-tailed)
Game (all 4 game treatments	.045 (.016) [.019, .072]	.005**	.090 (.019) [.058, .121]	.000**	.068 (.031) [.018, .119]	.025*
Control Variables						
Age	.000 (.003) [-.005, .005]	.961	-.001 (.003) [-.007, .005]	.714	-.007 (.004) [-.015, .000]	.060
Gender	-.020 (.014) [-.048, .007]	.145	-.017 (.016) [-.049, .015]	.306	-.049 (.027) [-.102, .005]	.073
Living in Germany	.005 (.001) [.002, .007]	.000**	.005 (.002) [.002, .008]	.002*	.003 (.003) [-.002, .009]	.229
Living in XX City	.001 (.001) [-.001, .003]	.341	.002 (.001) [-.001, .004]	.170	.002 (.002) [-.002, .005]	.406
Gaming motivation	.009 (.007) [-.005, .023]	.210	.009 (.008) [-.007, .025]	.290	.010 (.013) [-.016, .036]	.446
General waste sorting motivation	.020 (008) [.004, .035]	.012*	.016 (.009) [-.002, .033]	.085	.027 (.016) [-.005, .059]	.100
SUS	.001 (.000) [-.001, .002]	.064	.001 (.001) [-.000, .002]	.262	.001 (.001) [-.001, .003]	.181
Constant	.426 (.066) [.318, .534]	.000**	.361 (.079) [231, .492]	.000**	.570 (.108) [.392, .748]	.000**
N	213		213		213	
R^2	.193		.212		.091	
Adj. R^2	.161		.181		.055	

For the treatment groups, we used an alpha-error level of 10% (*p<0.1, ** p<0.01).
For the other controls that did not have directed hypotheses, we set the alpha-error level to 5% (* p<0.05, ** p<0.01).
Male was coded as 0, female as 1, and diverse as 2.

Table 57 – Effects of the Design Elements in Comparison with the Non-Game Material Group with Control Variables (Experiment Repeat & Look-Up)

	In-Game Performance		Multiple-Choice Test		Real-Life Sorting	
Reference category: Non-game material	coef. (bootstr. std. error) [conf. interval]	p (two-tailed)	coef. (bootstr. std. error) [conf. interval]	p (two-tailed)	coef. (bootstr. std. error) [conf. interval]	p (two-tailed)
Repeat element	.033 (.020) [.001, .066]	.094*	.086 (.023) [.048, .124]	.000**	.073 (.038) [.010, .135]	.056*
Look-up element	.044 (.021) [.009, .078]	.037*	.090 (.023) [.052, .127]	.000**	.072 (.037) [.012, .132]	.050*
Combined	.076 (.019)	.000**	.117 (.023)	.000**	.056 (.040)	.163

	[.044, .107]		[.079, .154]		[-.010, .123]	
Core gameplay	.029 (.020) [-.004, .062]	.144	.065 (.023) [.027, .104]	.005**	.071 (.035) [.013, .129]	.045*
Control Variables						
Age	-.000 (.002) [-.005, .004]	.841	-.002 (.003) [-.007, .004]	.542	-.007 (.004) [-.015, .001]	.074
Gender	-.019 (.014) [-.046, .008]	.175	-.014 (.016) [-.046, .018]	.389	-.049 (.028) [-.103, .005]	.078
Living in Germany	.004 (.001) [.002, .007]	.000**	.005 (.002) [.002, .008]	.002**	.003 (.003) [-.002, .009]	.239
Living in XX City	.001 (.001) [-.001, .003]	.265	.002 (.001) [-.001, .004]	.140	.002 (.002) [-.002, .005]	.433
Gaming motivation	.009 (.007) [-.005, .023]	.195	.009 (.008) [-.007, .026]	.267	.010 (.013) [-.016, .036]	.461
General waste sorting motivation	.018 (.008) [.003, .033]	.019*	.014 (.009) [-.003, .032]	.114	.028 (.017) [-.005, .060]	.095
SUS	.001 (.000) [-.000, .002]	.064	.001 (.001) [-.000, .002]	.255	.001 (.001) [-.001, .003]	.170
Constant	.449 (067) [.318, .580]	.000**	.382 (.079) [.228, .536]	.000**	.561 (108) [.347, .774]	.000**
N	213		213		213	
R^2	.219		.233		.092	
Adj. R^2	.176		.191		.042	

For the treatment groups, we used an alpha-error level of 10% (*p<0.1, ** p<0.01).
For the other controls that did not have directed hypotheses, we set the alpha-error level to 5% (* p<0.05, ** p<0.01).

Table 58 – Effects of the Design Elements in Comparison with the Core Gameplay (Experiment Repeat & Look-Up)

	In-Game Performance		**Multiple-Choice Test**		**Real-Life Sorting**	
Reference category: Core gameplay	coef. (bootstr. std. error) [conf. interval]	p (two-tailed)	coef. (bootstr. std. error) [conf. interval]	p (two-tailed)	coef. (bootstr. std. error) [conf. interval]	p (two-tailed)
Repeat element	.004 (.019) [-.027, .035]	.831	.021 (.022) [-.015, .056]	.337	.002 (.033) [-.052, .056]	.958
Look-up element	.015 (.021) [-.019, .049]	.470	.024 (.021) [-.011, .059]	.256	.001 (.033) [-.023, .055]	.978
Combined	.047 (.019) [.016, .077]	.012*	.052 (.022) [.017, .086]	.015*	-.015 (.036) [-.073, .044]	.681
Non-game material	-.029 (.020) [-.062, .004]	.144	-.065 (.023) [-.104, -.027]	.005*	-.071 (.035) [-.129, -.013]	.045*
Control Variables						
Age	-.000 (.002) [-.005, .004]	.841	-.002 (.003) [-.007, .004]	.542	-.007 (.004) [-.015, .001]	.074
Gender	-.019 (.014) [-.046, .008]	.175	-.014 (.016) [-.046, .018]	.389	-.049 (.028) [-.103, .005]	.078
Living in Germany	.004 (.001) [.002, .007]	.000**	.005 (.002) [.002, .008]	.002**	.003 (.003) [.002, .009]	.239
Living in XX City	.001 (.001) [-.001, .003]	.265	.002 (.001) [-.001, .004]	.140	.002 (.002) [-.002, .005]	.433

Gaming motivation	.009 (.007) [-.005, .023]	.195	.009 (.008) [-.007, .026]	.267	.010 (.013) [-.016, .036]	.461
General waste sorting motivation	.018 (.008) [.003, .033]	.019*	.014 (.009) [-.003, .032]	.114	.028 (.017) [-.005, .060]	.095
SUS	.001 (.000) [.000, .002]	.064	.001 (.001) [-.000, .002]	.255	.001 (.001) [-.001, .003]	.170
Constant	.478 (.065) [.350, .606]	.000**	.447 (.078) [.294, .601]	.000**	.632 (.107) [.421, .842]	.000**
N	213		213		213	
R^2	.219		.233		.092	
Adj. R^2	.176		.191		.042	

For the treatment groups, we used an alpha-error level of 10% (*p<0.1, ** p<0.01).
For the other controls that did not have directed hypotheses, we set the alpha-error level to 5% (* p<0.05, ** p<0.01).

Bibliography

(BMUV), B. für U. N. nukleare S. und V. (2021). *Umwelt im Unterricht - Aktuelle Bildungsmaterialien*. https://www.umwelt-im-unterricht.de/unterrichtsvorschlaege/wie-laesst-sich-abfall-vermeiden/

Abdel-Shafy, H. I., & Mansour, M. S. M. (2018). Solid waste issue: Sources, composition, disposal, recycling, and valorization. *Egyptian Journal of Petroleum*, *27*(4), 1275–1290. https://doi.org/10.1016/j.ejpe.2018.07.003

Adams, E. (2014). *Fundamentals of game design*. Pearson Education.

Adams, E. (Ernest W. ., & Dormans, J. (2012). *Game mechanics : advanced game design*. New Riders.

Adams, S. A. (2010). Use of "serious health games" in health care: a review. *Information Technology in Health Care: Socio-Technical Approaches 2010*, 160–166.

Advanced Research Consortium at Texas A&M U. (2015). *BigDIVA*. https://bigdiva.org/

Ahmadian, M. J. (2012). Task repetition in ELT. *English Language Teaching Journal*, *66*(3), 380–382. https://doi.org/10.1093/elt/ccs020

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, *16*(3), 183–198. https://doi.org/10.1016/j.learninstruc.2006.03.001

Alderfer, C. (1972). ERG Theory. *New York: Free Press.* http://changingminds.org/explanations/theories/erg_theory.htm

Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help Seeking and Help Design in Interactive Learning Environments. *Review of Educational Research*, *73*(3), 277–320. https://doi.org/10.3102/00346543073003277

Alexander, C. (1977). *A pattern language: towns, buildings, construction*. Oxford university press.

Allam, A., Kostova, Z., Nakamoto, K., & Schulz, P. J. (2015). The effect of social support features and gamification on a Web-based intervention for rheumatoid arthritis patients: Randomized controlled trial. *Journal of Medical Internet Research*, *17(1)*(1), e14. https://doi.org/10.2196/jmir.3510

Amanita Design. (2009). *Machinarium*. www.amanita-design.net

Amazon.com, I. (1994). *Amazon.com*. www.amazon.com

Amt für Abfallwirtschaft (AfA) Karlsruhe. (n.d.). *Amt für Abfallwirtschaft (AfA) Karlsruhe*. https://www.karlsruhe.de/b4/buergerdienste/abfall.de

Anime Kultur Verein. (2014). *Animetreff.de/*. https://www.animetreff.de/

Animexx e.V. (2000). *animexx*. https://www.animexx.de/

Arcade Analytics LTD. (2019). *Arcade Analytics*. https://arcadeanalytics.com/

Aristotle. (n.d.). *Nicomachean Ethics*. http://classics.mit.edu/Aristotle/nicomachaen.2.ii.html

Atkinson, R. C., & Shiffrin, R. M. (1968). Human Memory: A Proposed System and Its Control Processes. In R. J. Sternberg, S. T. Fiske, & D. J. Foss (Eds.), *Scientists Making a Difference* (pp. 115–118). Cambridge University Press. https://doi.org/10.1017/cbo9781316422250.025

Baddeley, A., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, *8*, 47–89. https://doi.org/10.1016/j.cub.2009.12.014

Baddeley, A., & Wilson, B. A. (1994). When implicit learning fails: Amnesia and the problem of error elimination. *Neuropsychologia*. https://doi.org/10.1016/0028-3932(94)90068-X

Bailey, K. D. (1994). *Typologies and taxonomies: An introduction to classification techniques* (Vol. 102). Sage.

Ball, K., Berch, D. B., Helmers, K. F., Jobe, J. B., Leveck, M. D., Marsiske, M., Morris, J. N., Rebok, G. W., Smith, D. M., Tennstedt, S. L., Unverzagt, F. W., & Willis, S. L. (2002). Effects of cognitive training interventions with older adults: A randomized controlled trial. *Journal of the American Medical Association*, *288(18)*, 2271–2281. https://doi.org/10.1001/jama.288.18.2271

Banakou, D., Kishore, S., & Slater, M. (2018). Virtually Being Einstein Results in an Improvement in Cognitive Task Performance and a Decrease in Age Bias. *Frontiers in Psychology*, *9*, 917. https://doi.org/10.3389/fpsyg.2018.00917

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. *Intl. Journal of Human–Computer Interaction*, *24*(6), 574–594. https://doi.org/10.1080/10447310802205776

Barata, G., Gama, S., Jorge, J., & Gonçalves, D. (2013). Improving participation and learning with gamification. *Gamification '13: Proceedings of the First International Conference on Gameful Design, Research, and Applications*, 10–17. https://doi.org/10.1145/2583008.2583010

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, *128(4)*(4), 612–637. https://doi.org/10.1037/0033-2909.128.4.612

Barsalou, L. W. (1999). Language comprehension: Archival memory or preparation for situated action? *Discourse Processes*, *28(1)*, 61–80. https://doi.org/10.1080/01638539909545069

Bartle, R. (1996). Richard A. Bartle: Players Who Suit MUDs. *MUSE Web Site, April*. http://mud.co.uk/richard/hcds.htm

Beierwaltes, W. H. (1985). Playing with Medicine: A Historical Perspective. *Henry Ford Hospital Medical Journal*, *33*(1), 27–33.

Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P., & Berta, R. (2013). Assessment in and of Serious Games: An Overview. In *Advances in Human-Computer Interaction*. https://doi.org/10.1155/2013/136864

Berengueres, J., Alsuwairi, F., Zaki, N., & Ng, T. (2013). Gamification of a recycle bin with emoticons. *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 83–84. https://doi.org/10.1109/HRI.2013.6483512

Berlyne, D. E. (1960). *Conflict, arousal, and curiosity.* McGraw-Hill Book Company. https://doi.org/10.1037/11164-000

Berthold, K., & Renkl, A. (2010). How to foster active processing of explanations in instructional communication. *Educational Psychology Review*, *22*(1), 25–40.

Besharat, A., Ladik, D. M., & Carrillat, F. A. (2014). Are maximizers blind to the future? When today's best does not make for a better tomorrow. *Marketing Letters*, *25*(1), 77–91.

Betsch, C. (2004). *Präferenz für Intuition und Deliberation - Messung und Konsequenzen von individuellen Unterschieden in affekt- und kognitionsbasiertem Entscheiden*. 50.

Betsch, C., & Kunz, J. J. (2008). Individual strategy preferences and decisional fit. *Journal of Behavioral Decision Making*, *21*(5), 532–555.

Bifulco, I., Francese, R., Lettieri, M., Liscio, L., Passero, I., & Tortora, G. (2011). The TIE Project: Agile Development of a Virtual World Serious Game on Waste Disposal. *DMS*, 204–209.

Birk, M. V., Toker, D., Mandryk, R. L., & Conati, C. (2015). Modeling Motivation in a Social Network Game Using Player-Centric Traits and Personality Traits. *Proceedings of User Modeling, Adaptation and Personalization*, 18–30. https://doi.org/10.1007/978-3-319-20267-9_2

Björk, S. (2012). *Gameplay Design Patterns*. http://virt10.itu.chalmers.se/index.php/Main_Page

Björk, S., & Holopainen, J. (2005). Games and Design Patterns. In K. Salen & E. Zimmerman (Eds.), *The Game Design Reader: A Rules of Play Anthology* (pp. 410–437). MIT Press. https://books.google.de/books?hl=de&lr=&id=aZv6AQAAQBAJ&oi=fnd&pg=PA410&ots=erC2JvALRs&sig=QpGGEtvbRwqkAUp5k7Kz-xrfLDk&redir_esc=y#v=onepage&q&f=false

Blohm, I., & Leimeister, J. M. (2013). Gamification: Design of IT-based enhancing services for motivational support and behavioral change. *Business and Information Systems Engineering*, *5*(4), 275–278. https://doi.org/10.1007/s12599-013-0273-5

Bock, O., Baetge, I., & Nicklisch, A. (2014). *hroot*. https://hroot.wiso.uni-hamburg.de/

Bogost, I. (2011a). *Persuasive games: Exploitationware.* Web Log Comment. http://www.gamasutra.com/view/feature/6366/persuasive_games_exploitationware.php

Bogost, I. (2011b). *Reality is alright.* Web Log Comment. http://bogost.com/writing/blog/reality_is_broken

Bojd, B., Song, X., Tan, Y., & Yan, X. (2022). Gamified challenges in online weight-loss communities. *Information Systems Research*.

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *Science*, *323*(5916), 892–895.

Bradshaw, G. L., & Anderson, J. R. (1982). Elaborative encoding as an explanation of levels of processing. *Journal of Verbal Learning and Verbal Behavior*, *21*(2), 165–174. https://doi.org/https://doi.org/10.1016/S0022-5371(82)90531-X

Brian Moriarty. (1997). *Listen: The Potential of Shared Hallucinations.*

Brigham, T. J. (2015). An introduction to gamification: adding game elements for engagement. *Medical Reference Services Quarterly*, *34*(4), 471–480. https://doi.org/10.1080/02763869.2015.1082385

Brooke, J. (1996). SUS-A “Quick and Dirty” Usability Scale. *Usability Evaluation in Industry*, *189*(194), 4–7. https://doi.org/10.1201/9781498710411-35

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*(1), 32–42. https://doi.org/10.3102/0013189X018001032

Brucker, J. (2010). Playing with a bad deck: The caveats of card sorting as a web site redesign tool. *Journal of Hospital Librarianship*, *10*(1), 41–53.

Bucciol, A., Montinari, N., & Piovesan, M. (2015). Do not trash the incentive! Monetary incentives and waste sorting. *Scandinavian Journal of Economics*, *117*(4), 1204–1229. https://doi.org/10.1111/sjoe.12122

Buchanan, G., & McKay, D. (2011). In the bookshop: Examining popular search strategies. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, 269–278. https://doi.org/10.1145/1998076.1998127

Buckley, P., & Doyle, E. (2017). Individualising gamification: An investigation of the impact of learning styles and personality traits on the efficacy of gamification using a prediction market. *Computers & Education*, *106*, 43–55.

Buclet, N., & Godard, O. (2000). *Municipal Waste Management in Europe : a Comparative Study in Building Regimes*. Springer Netherlands. https://books.google.de/books?id=hu8sBAAAQBAJ&dq=municipal+waste+management+germany&lr=&hl=de&source=gbs_navlinks_s

Bühler, C., Greenberg, P. T., & Ripin, R. T. (1930). *The first year of life*.

Burmester, M. (2006). *Digital Game Based Learning: Proceedings of the 4th International Symposium for Information Design, 2nd of June 2005 at Stuttgart Media University*. KIT Scientific Publishing.

Burton, M. A. (1994). The Irrigation Management Game: A role playing exercise for training in irrigation management. *Irrigation and Drainage Systems*, *7*(4), 305–318. https://doi.org/10.1007/bf00881557

Busch, M., Mattheiss, E., Orji, R., Fröhlich, P., Lankes, M., & Tscheligi, M. (2016). Player Type Models – Towards Empirical Validation. *CHI Extended Abstracts*, 1835–1841. https://doi.org/10.1145/2851581.2892399

Bygate, M. (1996). Effects of task repetition: Appraising the developing language of learners. In D. Willis, Jane and Willis (Ed.), *Challenge and change in language teaching* (pp. 136–146).

Cahill, K., & Chalut, R. (2009). Optimal results: What libraries need to know about Google and search engine optimization. *The Reference Librarian*, *50*(3), 234–247.

Caillois, R. (1961). *Man, Play, and Games*. University of Illinois Press. https://books.google.de/books?hl=de&lr=&id=bDjOPsjzfC4C&oi=fnd&pg=PR9&dq=caillois&ots=onccDI6Hmh&sig=bXzevwcxKt3r_dBYpesQxj3doAk#v=onepage&q=caillois&f=false

Cambridge University Press. (2022). *Definition for "reward."* https://dictionary.cambridge.org/dictionary/english-german/reward

Candy, L., Edmonds, E. A., & Patrick, D. J. (1996). Interactive knowledge support to conceptual design. In *AI System Support for Conceptual Design* (pp. 260–278). Springer.

Canossa, A., Drachen, A., & Sørensen, J. R. M. (2011). Arrrgghh!!! blending quantitative and qualitative methods to detect player frustration. *Proceedings of the 6th International Conference on Foundations of Digital Games*, 61–68.

Carignan, J., & Lawler Kennedy, S. (2013). Case study: Identifying gamification opportunities in sales applications. *International Conference of Design, User Experience, and Usability*, 501–507.

Carroll, J. M. (1991). *Designing interaction: Psychology at the human-computer interface*. CUP Archive.

Challco, G. C., Moreira, D., Mizoguchi, R., & Isotani, S. (2014). *Towards an Ontology for Gamifying Collaborative Learning Scenarios* (pp. 404–409). Springer, Cham. https://doi.org/10.1007/978-3-319-07221-0_50

Chandler, P., & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction*. https://doi.org/10.1207/s1532690xci0804_2

Chaparro, B. S., Hinkle, V. D., & Riley, S. K. (2008). The usability of computerized card sorting: A comparison of three applications by researchers and end users. *Journal of Usability Studies*, *4*(1), 31–48.

Chen, T., Peng, L., Yang, J., & Cong, G. (2021). Analysis of user needs on downloading behavior of English vocabulary APPs based on data mining for online comments. *Mathematics*, *9*(12), 1341.

Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science*, *25*(4), 471–533. https://doi.org/10.1207/s15516709cog2504_1

Chillarege, K. A., Nordstrom, C. R., & Williams, K. B. (2003). Learning from our mistakes: Error management training for mature learners. *Journal of Business and Psychology*, *17*(3), 369–385. https://doi.org/10.1023/A:1022864324988

Chou, Y. (2016). *Actionable Gamification: Beyond Points, Badges, and Leaderboards*. Packt Publishing Ltd.

Cialdini, R. B., Trost, M. R., & Newsom, J. T. (1995). Preference for consistency: The development of a valid measure and the discovery of surprising behavioral implications. *Journal of Personality and Social Psychology*, *69*(2), 318.

Clare, L., & Jones, R. S. P. (2008). Errorless learning in the rehabilitation of memory impairment: A critical review. *Neuropsychology Review*, *18*(1), 1–23. https://doi.org/10.1007/s11065-008-9051-4

Clare, L., Wilson, B. A., Breen, K., & Hodges, J. R. (1999). Errorless learning of face-name associations in early Alzheimer's disease. *Neurocase*, *5*(1), 37–46. https://doi.org/10.1093/neucas/5.1.37

Clarebout, G., & Elen, J. (2008). The complexity of tool use in computer-based learning environments. *Instructional Science*, *37*(5), 475–486. https://doi.org/10.1007/s11251-008-9068-3

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital Games, Design, and Learning: A Systematic Review and Meta-Analysis. *Review of Educational Research*, *86*(1), 79–122. https://doi.org/10.3102/0034654315582065

Clark, J. M., & Paivio, A. (1987). A dual coding perspective on encoding processes. In *Imagery and related mnemonic processes* (pp. 5–33). Springer.

Clark, K., & Sefton, M. (2001). Repetition and signalling: experimental evidence from games with efficient equilibria. *Economics Letters*, *70*(3), 357–362. https://doi.org/https://doi.org/10.1016/S0165-1765(00)00381-5

Clement, J. (2020). *Global market share of search engines 2010-2020*. https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/

Clibbon, K., Candy, L., & Edmonds, E. A. (1995). A logic-based framework for representing architectural design knowledge. *Proceedings. Computer-Aided Architectural Design Futures. International Conference, The Global Studio*, 42–53.

Clocher, B., & Hoffmann, G. (2014). *Die Müll AG*. https://muell-ag.de/?home

Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education*, *18*(6), 683–695. https://doi.org/10.1080/13562517.2013.827653

Cohen-Almagor, R. (2013). Internet history. In *Moral, ethical, and social dilemmas in the age of technology: Theories and practice* (pp. 19–39). IGI Global.

Cohen, P., West, S. G., & Aiken, L. S. (2014). *Applied multiple regression/correlation analysis for the behavioral sciences*. Psychology Press.

Cooley, M. (2000). Human-centered design. *Information Design*, 59–81.

Crawford, C. (2003). Chris Crawford on Game Design. In *Chris Crawford on Game Design*. https://doi.org/10.1093/carcin/bgs054

Crowder, R. G., & Melton, A. W. (1965). The Ranschburg phenomenon: Failures of immediate recall correlated with repetition of elements within a stimulus. *Psychonomic Science*, *2*(1), 295–296. https://doi.org/10.3758/BF03343462

Cruickshank, D. R., & Telfer, R. (1980). Classroom games and simulations. *Theory into Practice*, *19*(1), 75–80.

Csikszentmihalyi, M. (2000). *Beyond boredom and anxiety*. Jossey-Bass Publishers. https://www.medimops.de/mihaly-csikszentmihalyi-beyond-boredom-and-anxiety-experiencing-flow-in-work-and-play-gebundene-ausgabe-M00787951404.html?variant=UsedVeryGood&creative=Shopping&sitelink=&gclid=Cj0KCQjw0IDtBRC6ARIsAIA5gWtI_nCqle0ICKNdiclr67w_yvCPXe

Csikszentmihalyi, M., Abuhamdeh, S., & Nakamura, J. (2005). Flow. In A. J. E. & C. S. Dweck (Ed.), *Handbook of competence and motivation.* (pp. 598–608). Guilford Publications. https://www.lib.byu.edu/cgi-bin/remoteauth.pl?url=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2005-08058-032&site=ehost-live&scope=site

Csikszentmihalyi, M., Abuhamdeh, S., & Nakamura, J. (2014). Flow. In *Flow and the Foundations of Positive Psychology* (pp. 227–238). Springer Netherlands. https://doi.org/10.1007/978-94-017-9088-8_15

Curry, D., McCarragher, T., & Dellmann-Jenkins, M. (2005). Training, transfer, and turnover: Exploring the relationship among transfer of learning factors and staff retention in child welfare. *Children and Youth Services Review*, *27*(8), 931–948. https://doi.org/10.1016/j.childyouth.2004.12.008

Dahling, J. J., & Thompson, M. N. (2013). Detrimental relations of maximization with academic and career attitudes. *Journal of Career Assessment*, *21*(2), 278–294.

De Ruiter, J. (2002). Aspects of dealing with digital information: “Mature” novices on the internet. *Library Trends*, *51*(2), 199-209+261.

Deci, E L, Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. In *Psychological bulletin* (Vol. 125, Issue 6, pp. 627–668; discussion 692-700). https://doi.org/10.1037/0033-2909.125.6.627

Deci, Edward L., Koestner, R., & Ryan, R. M. (2001). Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of Educational Research*, *71*(1), 1–27. https://doi.org/10.3102/00346543071001001

Deci, Edward L., & Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. Springer US. https://doi.org/10.1007/978-1-4899-2271-7

Deci, Edward L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, *18*(1), 105.

Deci, Edward L. (1972). The effects of contingent and noncontingent rewards and controls on intrinsic motivation. *Organizational Behavior and Human Performance*, *8*(2), 217–229.

Dennis, A. R., & Carte, T. A. (1998). Using geographical information systems for decision making: Extending cognitive fit theory to map-based presentations. *Information Systems Research*, *9*(2), 194–203.

Deterding, S. (2016). Make-Believe in Gameful and Playful Design. In P. Turner & J. T. Harviainen (Eds.), *Digital Make-Belief* (pp. 101–124). Springer International Publishing Switzerland.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining "gamification." *MindTrek '11: Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments*, 9–15. https://doi.org/10.1145/2181037.2181040

Deutsche Digitale Bibliothek. (2019). *Deutsche Digitale Bibliothek*. https://www.deutsche-digitale-bibliothek.de/?lang=en

Devanur, N. R., & Hayes, T. P. (2009). The adwords problem: online keyword matching with budgeted bidders under random permutations. *Proceedings of the 10th ACM Conference on Electronic Commerce*, 71–78.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). {BERT:} Pre-training of Deep Bidirectional Transformers for Language Understanding. *CoRR*, *abs/1810.0*. http://arxiv.org/abs/1810.04805

Dichev, C., Dicheva, D., Angelova, G., & Agre, G. (2015). From Gamification to Gameful Design and Gameful Experience in Learning. *Cybernetics and Information Technologies*, *14*(4), 80–100. https://doi.org/10.1515/cait-2014-0007

Dickey, M. D. (2005). Engaging by design: How engagement strategies in popular computer and video games can inform instructional design. *Educational Technology Research and Development*. https://doi.org/10.1007/BF02504866

Dignan, A. (2011). Game Frame. In *Using Games as a Strategy for Success*. http://books.google.com/books?id=nkUr1RI21jcC

Ding, X., Liu, T., Duan, J., & Nie, J.-Y. (2015). Mining user consumption intention from social media using domain adaptive convolutional neural network. *Proceedings of the AAAI Conference on Artificial Intelligence*, *29*(1).

Dini, K. (2012). On video games, culture, and therapy. *Psychoanalytic Inquiry*, *32*(5), 496–505.

DiNucci, D. (1999). Fragmented Future. *Print Magazine*, 3. http://darcyd.com/fragmented_future.pdf

Dix, A., Finlay, J., Abowd, G. D., & Beale, R. (2004). Human-computer interaction: Pearson prentice hall. *Inc, England*.

Dixon, D. (2011). Player Types and Gamification. *CHI 2011 Workshop Gamification Using Game Design Elements in NonGame Contexts*, 12–15. https://doi.org/ACM 978-4503-0268-5/11/05

Donaghey, C. L., McMillan, T. M., & O'Neill, B. (2010). Errorless learning is superior to trial and error when learning a practical skill in rehabilitation: A randomized controlled trial. *Clinical Rehabilitation*, *24*(3), 195–201. https://doi.org/10.1177/0269215509353270

dos Reis, A. V., & Press, L. T. (2019). Sesame Credit: Social Compliance Gamification in China. *Journal of Games, Game Art, and Gamification*, *4*(1), 7–13.

Doyle, A. C. (1992). *The adventures of sherlock holmes*. Wordsworth Editions.

Drachen, A., Canossa, A., & Yannakakis, G. N. (2009). *Player Modeling using Self-Organization in Tomb Raider: Underworld*. https://doi.org/10.4067/S0716-58112003001400003

Dreher, A., Kuntze, S., & Winkel, K. (2014). Empirical Study of a Competence Structure Model Regarding Conversions of Representations--The Case of Fractions. In D. Nicol, C., Liljedahl, P., Oesterle, S., & Allan (Ed.), *Proceedings of the Joint Meeting of PME 38 and PME-NA 36* (pp. 425–432). ERIC.

Driskell, J. E., Willis, R. P., & Copper, C. (1992). Effect of Overlearning on Retention. *Journal of Applied Psychology*, *77*(5), 615–622. https://doi.org/10.1037/0021-9010.77.5.615

Druckman, J. N., & Kam, C. D. (2011). Students as experimental participants. *Cambridge Handbook of Experimental Political Science*, *1*, 41–57.

Duffield, J. A. (1991). Designing computer software for problem-solving instruction. *Educational Technology Research and Development*, *39*(1), 50–62. https://doi.org/10.1007/BF02298106

Dunn, J., & Clare, L. (2007). Learning face{\textendash}name associations in early-stage dementia: Comparing the effects of errorless learning and effortful processing. *Neuropsychological Rehabilitation*, *17*(6), 735–754. https://doi.org/10.1080/09602010701218317

Dunsmoor, J. E., Bandettini, P. A., & Knight, D. C. (2007). Impact of continuous versus intermittent CS-UCS pairing on human brain activation during Pavlovian fear conditioning. *Behavioral Neuroscience*, *121*(4), 635–642. https://doi.org/10.1037/0735-7044.121.4.635

Eco, U. (1983). *The Name of the Rose*.

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, *51*(4), 380.

Egenfeldt-Nielsen, S., Smith, J. H., & Pajares Tosca, S. (2008). *Understanding video games: The essential introduction*. Routledge.

El-Hilly, A. A., Iqbal, S. S., Ahmed, M., Sherwani, Y., Muntasir, M., Siddiqui, S., Al-Fagih, Z., Usmani, O., & Eisingerich, A. B. (2016). Game On? Smoking Cessation Through the Gamification of mHealth: A Longitudinal Qualitative Study. *JMIR Serious Games*, *4(2)*(2), e18. https://doi.org/10.2196/games.5678

El-Khuffash, A. (2013). Gamification. *Ryerson University, Torondo, Canada*.

Eldredge, N., & Cracraft, J. (1980). Phylogenetic patterns and the evolutionary process. *Method and Theory in Comparative Biology*, *8*, 1–349.

Elliot, S., & Webster, J. (2017). Editorial: Special issue on empirical research on information systems addressing the challenges of environmental sustainability: an imperative for urgent action. *Information Systems Journal*, *27*(4), 367–378. https://doi.org/10.1111/isj.12150

Elverdam, C., & Aarseth, E. (2007). Game classification and game design: Construction through critical analysis. *Games and Culture*, *2*(1), 3–22.

Epstein, S. (1993). *Implications of cognitive-experiential self-theory for personality and developmental psychology.*

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, *49*(8), 709.

Epstein, S. (2003). Cognitive-experiential self-theory of personality. *Comprehensive Handbook of Psychology*, *5*, 159–184.

Epstein, S. (2014). *Cognitive-experiential theory: An integrative theory of personality*. Oxford University Press.

Epstein, S., & Brodsky, A. (1993). *You're smarter than you think: How to develop your practical intelligence for success in living*. Simon & Schuster.

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive–experiential and analytical–rational thinking styles. *Journal of Personality and Social Psychology*, *71*(2), 390.

Epworth Bells Crowle and Isle of Axholme Messenger. (1874). *Healthful Play*. https://blog.britishnewspaperarchive.co.uk/2021/05/19/an-exploration-of-the-history-and-importance-of-play/

Estellés-Arolas, E., & González-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, *38*(2), 189–200.

European Parliament. (2018). *The circular economy package: new EU targets for recycling*. News. http://www.europarl.europa.eu/news/en/headlines/priorities/circular-economy/20170120STO59356/the-circular-economy-package-new-eu-targets-for-recycling

Europeana Foundation. (2019). *Europeana Collections*. https://www.europeana.eu/en/collections

Evans, J. J., Wilson, B. A., Schuri, U., Andrade, J., Baddeley, A., Bruna, O., Canavan, T., Sala, S. Della, Green, R., Laaksonen, R., Lorenzi, L., & Taussik, I. (2000). A comparison of "errorless" and "trial-and-error" learning methods for teaching individuals with acquired memory deficits. *Neuropsychological Rehabilitation*, *10*(1), 67–101. https://doi.org/10.1080/096020100389309

Evans, J. R., & Mathur, A. (2018). The value of online surveys: A look back and a look ahead. *Internet Research*.

Explosion AI. (2019). *spaCy - Industrial-Strength Natural Language Processing*. https://spacy.io/

Fabricatore, C. (2007). *Gameplay and game mechanics: a key to quality in videogames*.

Fenn, J., & LeHong, H. (2011). *Hype Cycle for Emerging Technologies, 2011*. https://www.gartner.com/en/documents/1754719#:~:text=Technologies at the Peak of,and natural language question answering.

Ferri, A. J. (2007). *Willing suspension of disbelief: Poetic faith in film*. Lexington Books.

Ferro, L. S., Walz, S. P., & Greuter, S. (2013). Towards personalised, gamified systems. *Proceedings of The 9th Australasian Conference on Interactive Entertainment Matters of Life and Death - {IE}{\textquotesingle}13*. https://doi.org/10.1145/2513002.2513024

Fields, B., Keith, S., & Blandford, A. (2005). Designing for expert information finding strategies. In *People and computers XVIII—Design for life* (pp. 89–102). Springer.

Fileni, F. (1988). Educational and cognitive aspects of videogames. *Simulation-Gaming in Education and Training*, 252–260.

Filho, W. L., Brandli, L., Moora, H., Kruopiene, J., & Stenmarck, Å. (2016). Benchmarking approaches and methods in the field of urban waste management. *Journal of Cleaner Production*, *112*, 4377–4386. https://doi.org/10.1016/j.jclepro.2015.09.065

Fillingham, J. K., Hodgson, C., Sage, K., & Lambon Ralph, M. A. (2003). The application of errorless learning to aphasic disorders: A review of theory and practice. *Neuropsychological Rehabilitation*, *13*(3), 337–363. https://doi.org/10.1080/09602010343000020

Fincher, S., & Tenenberg, J. (2005). Making sense of card sorting data. *Expert Systems*, *22*(3), 89–93.

Finixio Ltd. (2021). *Business 2 Community*. https://www.business2community.com/mobile-apps/app-ratings-and-reviews-2021-benchmarks-02396808#:~:text=The average iOS app has,average 2%2C586 app store reviews.

Fotaris, P., Mastoras, T., Leinfellner, R., & Rosunally, Y. (2016). Climbing up the leaderboard: An empirical study of applying gamification techniques to a computer programming class. *Electronic Journal of E-Learning*, *14(2)*(2), 94–110.

Freeman, L. (2004). The development of social network analysis. *A Study in the Sociology of Science*, *1*(687), 159–167.

Frese, M., Brodbeck, F., Heinbokel, T., Mooser, C., Schleiffenbaum, E., & Thiemann, P. (1991). Errors in training computer skills: On the positive function of errors. *Human-Computer Interaction*, *6*(1), 77–93. https://doi.org/10.1207/s15327051hci0601_3

Friedrich, H. F., & Mandl, H. (2006). Lernstrategien: Zur Strukturierung des Forschungsfeldes. *Handbuch Lernstrategien*, 1–23.

Frost, R. O., Marten, P., Lahart, C., & Rosenblate, R. (1990). The dimensions of perfectionism. *Cognitive Therapy and Research*, *14*(5), 449–468. https://doi.org/10.1007/BF01172967

Fullerton, T. (2008). Game Design Workshop: A Playcentric Approach to Creating Innovative Games. In *Technology*. https://doi.org/10.1007/s13398-014-0173-7.2

Gagn?, R. M. (1984). Learning outcomes and their effects: Useful categories of human performance. *American Psychologist*, *39*(4), 377–385. https://doi.org/10.1037/0003-066x.39.4.377

Galactic Cafe. (2013). *The Stanley Parable*. https://www.stanleyparable.com/

Gall, S. N.-L. (1981). Help-seeking: An understudied problem-solving skill in children. *Developmental Review*, *1*(3), 224–246. https://doi.org/10.1016/0273-2297(81)90019-8

game – Verband der deutschen Games-Branche e. V. (2019). *Jahresreport der deutschen Games-Branche 2019*. https://www.game.de/wp-content/uploads/2018/08/game-Jahresreport-2019_web.pdf

game – Verband der deutschen Games-Branche e. V. (2020). *Jahresreport der deutschen Games-Branche 2020*. https://www.game.de/wp-content/uploads/2020/08/game-Jahresreport-2020.pdf

GameFreak. (1996). *Pokémon*.

Gamified UK. (2018). *Gamified.uk*. https://www.gamified.uk/

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming*, *33*(4), 441–467. https://doi.org/10.1177/1046878102238607

Gaston, J., & Cooper, S. (2017). To Three or not to Three: Improving Human Computation Game Onboarding with a Three-Star System. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, 5034–5039. https://doi.org/10.1145/3025453.3025997

Gawronski, B., & Creighton, L. A. (2013). *Dual process theories.*

Geary, D. C. (2008). An Evolutionarily Informed Education Science. *Educational Psychologist*, *43*(4), 179–195. https://doi.org/10.1080/00461520802392133

Geary, D. C., & Berch, D. B. (2016). Evolution and children's cognitive and academic development. In *Evolutionary perspectives on child development and education* (pp. 217–249). Springer.

Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment*, *1*(1), 20. https://doi.org/10.1145/950566.950595

Geen, R. G. (1995). *Human motivation: A social psychological approach.* Thomson Brooks/Cole Publishing Co.

Gemeinsame Stelle dualer Systeme Deutschlands GmbH. (2021). *mülltrennung-wirkt.de Eine Initiative der dualen Systeme*. https://www.muelltrennung-wirkt.de/

Gibson, J. J. (1977). The Theory of Affordances. In R. Shaw & J. Bransford (Eds.), *R. E. Shaw & J. Bransford (eds.), Perceiving, Acting, and Knowing.*

Gigerenzer, G., & Selten, R. (2002). *Bounded rationality: The adaptive toolbox*. MIT press.

Gitlab.com/inkscape/inkscape. (2003). *Inkscape*. inkscape.org

Glenberg, A. M., & Robertson, D. A. (1999). Indexical understanding of instructions. *Discourse Processes*, *28*(1), 1–26. https://doi.org/10.1080/01638539909545067

Goldberg, L. R. (1990). An alternative" description of personality": the big-five factor structure. *Journal of Personality and Social Psychology*, *59*(6), 1216.

González-Briones, A., Valdeolmillos, D., Casado-Vara, R., Chamoso, P., Coria, J. A. G., Herrera-Viedma, E., & Corchado, J. M. (2018). GarbMAS: Simulation of the Application of Gamification Techniques to Increase the Amount of Recycled Waste Through a Multi-agent System. In F.-C. A. De La Prieta F., Omatu S. (Ed.), *Distributed Computing and Artificial Intelligence, 15th International Conference* (pp. 332–343). Springer International Publishing. https://doi.org/10.1007/978-3-319-94649-8_40

Google. (2004). *Google Scholar*. https://scholar.google.com/

Google. (2007). *GoogleGuide making searching even easier*. http://www.googleguide.com/directory.html

Google Ads. (n.d.). *About keyword matching options*. https://support.google.com/google-ads/answer/7478529?hl=en

Google Trends. (2022). *Google Trends - Gamification*. https://trends.google.com/trends/explore?date=all&geo=DE&q=gamification

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*. https://doi.org/10.1016/S0092-6566(03)00046-1

Graves, L., & Cherubini, F. (2016). *The rise of fact-checking sites in Europe*.

Green, R. (1996). Typologies and taxonomies: An introduction to classification techniques. In *Journal of the American Society for Information Science* (Vol. 47, Issue 4). https://doi.org/10.1002/(SICI)1097-4571(199604)47:4<328::AID-ASI10>3.0.CO;2-Y

Greenfield, P. M., & Savage-Rumbaugh, E. S. (1993). Comparing communicative competence in child and chimp: the pragmatics of repetition. *Journal of Child Language*, *20*(1), 1–26. https://doi.org/DOI: 10.1017/S0305000900009090

Greenwald, A. G. (1982). Ego task analysis: An integration of research on ego-involvement and self-awareness. *Cognitive Social Psychology*, 109–147.

Greitzer, F. L., Kuchar, O. A., & Huston, K. (2007). Cognitive science implications for enhancing training effectiveness in a serious gaming context. *Journal on Educational Resources in Computing (JERIC)*, *7*(3), 2-es.

Griffiths, M. D., & Hunt, N. (1995). Computer game playing in adolescence: Prevalence and demographic indicators. *Journal of Community & Applied Social Psychology*, *5*(3), 189–193. https://doi.org/10.1002/casp.2450050307

Groos, K. (1901). *The play of man (EL Baldwin, Trans.). New York, NY, US: D Appleton & Company*.

Größler, A., Maier, F. H., & Milling, P. M. (2000). Enhancing learning capabilities by providing transparency in business simulators. *Simulation & Gaming*, *31*(2), 257–278. https://doi.org/10.1177/104687810003100209

Guardiola, E. (2016). The gameplay loop: a player activity model for game design and analysis. *Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology*, 1–7.

Gustafsson, A., Katzeff, C., & Bang, M. (2009). Evaluation of a pervasive game for domestic energy engagement among teenagers. *Computers in Entertainment*, *7*(4), 1–19. https://doi.org/10.1145/1658866.1658873

Hackman, J. R., & Oldham, G. R. (1975). Development of the Job Diagnostic Survey. *Journal of Applied Psychology*, *60*(2), 159–170. https://doi.org/10.1037/h0076546

Halfbrick Studios pty. ltd. (2010). *Fruit Ninja*. https://www.halfbrick.com/games/fruit-ninja

Hamachek, D. E. (1978). Psychodynamics of normal and neurotic perfectionism. *Psychology: A Journal of Human Behavior*, *15*(1), 27–33.

Hamari, J. (2017). Do badges increase user activity? A field experiment on the effects of gamification. *Computers in Human Behavior*, *71*, 469–478. https://doi.org/10.1016/j.chb.2015.03.036

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? - A literature review of empirical studies on gamification. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 3025–3034.

Hamari, J., & Lehdonvirta, V. (2010). Game design as marketing: How game mechanics create demand for virtual goods. *International Journal of Business Science and Applied Management*, *5*(1), 14–29. https://doi.org/10.1108/00251741211203542

Hanrahan, M. (1998). The effect of learning environment factors on students' motivation and learning. *International Journal of Science Education*, *20*(6), 737–753.

Hanseth, O., & Lyytinen, K. (2004). *Theorizing about the design of Information Infrastructures: design kernel theories and principles*.

Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education*, *80*, 152–161.

Harackiewicz, J. M., Barron, K. E., & Elliot, A. J. (1998). Rethinking achievement goals: When are they adaptive for college students and why? *Educational Psychologist*, *33*(1), 1–21.

Harackiewicz, J. M., Manderlink, G., & Sansone, C. (1984). Rewarding pinball wizardry: Effects of evaluation and cue value on intrinsic interest. *Journal of Personality and Social Psychology*, *47*(2), 287.

Harteveld, C. (2011). Triadic Game Design. *Media*. https://doi.org/10.1007/978-1-84996-157-8

Hassenzahl, M. (2010). Experience Design: Technology for All the Right Reasons. *Synthesis Lectures on Human-Centered Informatics*. https://doi.org/10.2200/s00261ed1v01y201003hci008

Hawlitschek, F. (2020). Interview with Benjamin Scheffler on "The Future of Waste Management." *Business & Information Systems Engineering*, *63*, 207–211. https://doi.org/10.1007/s12599-020-00671-y

Herzberg, F. (1964). The motivation-hygiene concept and problems of manpower. *Personnel Administration*, *27*(1), 3–7.

Herzberg, F. (1974). Motivation-hygiene profiles: Pinpointing what ails the organization. *Organizational Dynamics*, *3*(2), 18–29. https://doi.org/10.1016/0090-2616(74)90007-2

Hewitt, P. L., & Flett, G. L. (1991). Perfectionism in the self and social contexts: conceptualization, assessment, and association with psychopathology. *Journal of Personality and Social Psychology*, *60*(3), 456.

Highhouse, S. E., Diab, D. L., & Gillespie, M. A. (2008). Are maximizers really unhappy? The measurement of maximizing tendency. *Judgment and Decision Making Journal*, *3*(5), 364.

Hintzman, D. L. (1976). Repetition and Memory. *Psychology of Learning and Motivation*, *10*, 47–91. https://doi.org/10.1016/S0079-7421(08)60464-8

Hinze, A., McKay, D., Vanderschantz, N., Timpany, C., & Cunningham, S. J. (2012). Book selection behavior in the physical library: Implications for ebook collections. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, 305–314. https://doi.org/10.1145/2232817.2232874

Hirsch Jr., E. D. (2000). You can Always Look it up, or can You? *American Educator*, *24(1)*(1), 4–9.

Hoffmann, G., & Martin, R. (2018). *Tako (formerly Kubun)*.

Hoffmann, G., Martin, R., & Weinhardt, C. (2019). Perfectionism in Games-Analyzing Playing Behaviors in an Educational Game. *2019 11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*, 1–4. https://doi.org/doi: 10.1109/VS-Games.2019.8864542.

Hoffmann, G., & Matysiak, L. (2019, September). Exploring Game Design for the Financial Education of Millenials. *2019 11th International Conference on Virtual Worlds and Games for Serious Applications ({VS}-Games)*. https://doi.org/10.1109/vs-games.2019.8864517

Hoffmann, G., & Pfeiffer, J. (2021). Gameful Learning for a More Sustainable World. *Business & Information Systems Engineering*, 1–24. https://doi.org/https://doi.org/10.1007/s12599-021-00731-x

Hölscher, C., & Strube, G. (2000). Web search behavior of internet experts and newbies. *Computer Networks*, *33*(1), 337–346. https://doi.org/10.1016/S1389-1286(00)00031-1

Honnibal, M. (2015). *spaCy*. https://spacy.io/models/en#en_core_web_trf

Hopelab. (2006). *Re-Mission*. https://www.re-mission2.org/

Huizinga, J. (1956). *Homo Ludens Vom Ursprung der Kultur Im Spiel*. https://philpapers.org/rec/HUIHLV

Hull, C. L. (1930). Simple trial and error learning: A study in psychological theory. *Psychological Review*, *37*(3), 241–256. https://doi.org/10.1037/h0073614

Hung, A. C. Y. (2017). A critique and defense of gamification. *Journal of Interactive Online Learning*, *15*(1).

Hunicke, R., Leblanc, M., & Zubek, R. (2004). *MDA: A Formal Approach to Game Design and Game Research*. http://www.aaai.org/Papers/Workshops/2004/WS-04-04/WS04-04-001.pdf

Hunkin, N. M., Squires, E. J., Parkin, A. J., & Tidy, J. A. (1998). Are the benefits of errorless learning dependent on implicit memory? *Neuropsychologia*, *36*(1), 25–36. https://doi.org/10.1016/S0028-3932(97)00106-1

Hunt, J. M. (1965). Intrinsic motivation and its role in development. In *Nebraska symposium on motivation, {Vol} 12* (pp. 189–282).

Huotari, K. (2012). Defining Gamification - A Service Marketing Perspective. *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, 17–22.

Idrobo, M. L., Saenz, M. F., Márceles, K., Chanchí, G. E., Vidal, M. I., & Burbano, C. L. (2018). Recycling: A Serious Game Focused on the Classification of Waste. In Z.-F. R. Mata-Rivera M. (Ed.), *International Congress of Telematics and Computing* (pp. 234–245). Springer, Cham. https://doi.org/10.1007/978-3-030-03763-5_20

Informa. (1997). *Game Developer (formerly known as Gamasutra).* https://www.gamedeveloper.com/

Institute for Systems Biology. (2002). *Cytoscape*. https://cytoscape.org/

International Organization for Standardization. (1998). *ISO 9241*.

Iosup, A., & Epema, D. (2014). An experience report on using gamification in technical higher education. *Proceedings of the 45th ACM Technical Symposium on Computer Science Education - SIGCSE '14*. https://doi.org/10.1145/2538862.2538899

Ivancic, K., & Hesketh, B. (2000). Learning from errors in a driving simulation: Effects on driving skill and self-confidence. *Ergonomics*, *43*(12), 1966–1984. https://doi.org/10.1080/00140130050201427

Jacob, E. K. (2004). Classification and Categorization : A Difference that makes a Difference. *Library Trends*, *52*(3), 515–540. https://doi.org/10.1007/s11406-014-9559-1

Jang, S., Kitchen, P. J., & Kim, J. (2018). The effects of gamified customer benefits and characteristics on behavioral engagement and purchase: Evidence from mobile exercise application uses. *Journal of Business Research*, *92*, 250–259. https://doi.org/10.1016/j.jbusres.2018.07.056

Jansz, J., & Tanis, M. (2007). Appeal of playing online First Person Shooter Games. *Cyberpsychology & Behavior : The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society*, *10*(1), 133–136. https://doi.org/10.1089/cpb.2006.9981

Järvinen, A. (2007). Introducing Applied Ludology: Hands-on Methods for Game Studies. *DiGRA Conference*, 1–11.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). Big five inventory. *Journal of Personality and Social Psychology*.

Johnson, A. (2004). We Learn from our Mistakes—Don't We? *Ergonomics in Design*, *12*(2), 24–27. https://doi.org/10.1177/106480460401200206

Jonassen, D. H., & Rohrer-Murphy, L. (1999). Activity theory as a framework for designing constructivist learning environments. *Educational Technology Research and Development*, *47*(1), 61–79. https://doi.org/10.1007/BF02299477

Jones, B. F. O. (1994). *Designing Learning and Technology for Educational Reform.* Publications Department, North Central Regional Educational laboratory, 1900 Spring Road, Suite 300, Oak Brook, IL 60521-1480 ($9.95, order #RPIC-DLT-94). https://eric.ed.gov/?id=ED378940

Jones, K. S. (1979). Experiments in relevance weighting of search terms. *Information Processing & Management*, *15*(3), 133–144.

Jones, R. S. P., & Eayrs, C. B. (1992). the Use of Errorless Learning Procedures in Teaching People With a Learning Disability: a Critical Review. *Mental Handicap Research*, *5*(2), 204–212. https://doi.org/10.1111/j.1468-3148.1992.tb00045.x

Jürgen Strohm. (2001). *Industrial Waste*. https://boardgamegeek.com/boardgame/2476/industrial-waste

Kahn, A. S., Shen, C., Lu, L., Ratan, R. A., Coary, S., Hou, J., Meng, J., Osborn, J., & Williams, D. (2015). The Trojan Player Typology: A cross-genre, cross-cultural, behaviorally validated scale of video game play motivations. *Computers in Human Behavior*, *49*, 354–361. https://doi.org/10.1016/j.chb.2015.03.018

Kappen, D., Mirza-Babaei, P., & Nacke, L. (2018). Gamification of Older Adults' Physical Activity: An Eight-Week Study. *Proceedings of the 51st Hawaii International Conference on System Sciences*, 1207–1216. https://doi.org/10.24251/hicss.2018.149

Karle, J. W., Watter, S., & Shedden, J. M. (2010). Task switching in video game players: Benefits of selective attention but not resistance to proactive interference. *Acta Psychologica*, *134*(1), 70–78.

Karlsruhe Institute of Technology. (2007). *KD²Lab*. https://www.kd2lab.kit.edu/

Karlsruhe Institute of Technology. (2009). *KIT Bibliothek*. https://primo.bibliothek.kit.edu/primo_library/libweb/action/search.do?vid=KIT&tab=kit_pc&srt=date

Karlsruhe Institute of Technology. (2012). *KIT-Katalog Plus*. https://www.bibliothek.kit.edu/suchen-finden.php

Karlsruhe University of Arts and Design. (2009). *GameLab Karlsruhe*. https://infoart.hfg-karlsruhe.de/category/gamelab/

Kato, P. M., Cole, S. W., Bradlyn, A. S., & Pollock, B. H. (2008). A Video Game Improves Behavioral Outcomes in Adolescents and Young Adults With Cancer: A Randomized Trial. *Pediatrics*, *122*(2), e305–e317. https://doi.org/10.1542/peds.2007-3134

Keith, N., & Frese, M. (2008). Effectiveness of Error Management Training: A Meta-Analysis. *Journal of Applied Psychology*, *93*(1), 59–69. https://doi.org/10.1037/0021-9010.93.1.59

Keller, J., Bohner, G., & Erb, H.-P. (2000). Intuitive und heuristische Urteilsbildung–verschiedene Prozesse? Präsentation einer deutschen Fassung des „Rational-Experiential Inventory "sowie neuer Selbstberichtskalen zur Heuristiknutzung. *Zeitschrift Für Sozialpsychologie*, *31*(2), 87–101.

Kelley, P., & Whatson, T. (2013). Making long-term memories in minutes: a spaced learning pattern from memory research in education. *Frontiers in Human Neuroscience*, *7*, 589. https://doi.org/10.3389/fnhum.2013.00589

Kelly, G. (1955). Personal construct psychology. *Nueva York: Norton*.

Kessels, R. P. C., & Haan, E. H. F. (2003). Implicit Learning in Memory Rehabilitation: A Meta-Analysis on Errorless Learning and Vanishing Cues Methods. *Journal of Clinical and Experimental Neuropsychology*, *25*(6), 805–814. https://doi.org/10.1076/jcen.25.6.805.16474

Kessels, R. P. C., van Loon, E., & Wester, A. J. (2007). Route learning in amnesia: A comparison of trial-and-error and errorless learning in patients with the Korsakoff syndrome. *Clinical Rehabilitation*, *21*(10), 905–911. https://doi.org/10.1177/0269215507077309

Kim, J. (2013). *Trophies, Plaques, and Rewards: An Application of Cognitive Evaluation Theory to Volunteer Motivation.*

Kim, Y. J., & Shute, V. J. (2015). The interplay of game elements with psychometric qualities, learning, and enjoyment in game-based assessment. *Computers and Education*, *87*, 340–356. https://doi.org/10.1016/j.compedu.2015.07.009

King. (2012). *Candy Crush Saga*. https://www.king.com/game/candycrush

King, D., Delfabbro, P., & Griffiths, M. (2009). Video Game Structural Characteristics: A New Psychological Taxonomy. *International Journal of Mental Health and Addiction*, *8*(1), 90–106. https://doi.org/10.1007/s11469-009-9206-4

Kleemann, F., Voß, G. G., & Rieder, K. (2008). Un (der) paid innovators: The commercial utilization of consumer work through crowdsourcing. *Science, Technology & Innovation Studies*, *4*(1), 5–26.

Klopfer, E. (2008). *Augmented learning: Research and design of mobile educational games*. MIT press.

Knight, C., & Munro, M. (1999). Comprehension with[in] virtual environment visualisations. *Proceedings*

Seventh International Workshop on Program Comprehension, 4–11. https://doi.org/10.1109/WPC.1999.777733

Kolers, P. A., & Brison, S. J. (1984). Commentary: On pictures, words, and their mental representation. *Journal of Verbal Learning and Verbal Behavior*, *23*(1), 105–113. https://doi.org/10.1016/S0022-5371(84)90539-5

Koster, R. (2005). Theory of Fun for Game Design. *A Theory of Fun for Game Design*, 256. https://doi.org/10.1007/s13398-014-0173-7.2

Koufaris, M. (2002). Applying the technology acceptance model and flow theory to online consumer behavior. *Information Systems Research*, *13*(2), 205–223.

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of Cognitive, Skill-Based, and Affective Theories of Learning Outcomes to New Methods of Training Evaluation. *Journal of Applied Psychology*, *78*(2), 311–328. https://doi.org/10.1037/0021-9010.78.2.311

Kreimeier, B. (2002). *The case for game design patterns*.

Kruglanski, A. W. (1990). *Motivations for judging and knowing: Implications for causal attribution.*

Kruse, J., & Lenger, A. (2014). Zur aktuellen Bedeutung von qualitativen Forschungsmethoden in der deutschen Volkswirtschaftslehre {\textendash} Eine programmatische Exploration. *Zeitschrift Für Qualitative Forschung*, *14*(1), 105–138. https://doi.org/10.3224/zqf.v14i1.15455

Kuechler, B., & Vaishnavi, V. (2008). Theory development in design science research: Anatomy of a research project. *Proceedings of the 3rd International Conference on Design Science Research in Information Systems and Technology, DESRIST 2008*. https://doi.org/10.1057/ejis.2008.40

Kumu. (2019). *Kumu relationship mapping software*. https://kumu.io/

Kurtzman, G. W., Day, S. C., Small, D. S., Lynch, M., Zhu, J., Wang, W., Rareshide, C. A. L., & Patel, M. S. (2018). Social Incentives and Gamification to Promote Weight Loss: The LOSE IT Randomized, Controlled Trial. *Journal of General Internal Medicine*, *33(10)*(10), 1669–1675. https://doi.org/10.1007/s11606-018-4552-1

Landers, R. N., & Landers, A. K. (2014). An Empirical Test of the Theory of Gamified Learning: The Effect of Leaderboards on Time-on-Task and Academic Performance. *Simulation & Gaming*, *45*(6), 769–785. https://doi.org/10.1177/1046878114563662

Laran, J., & Janiszewski, C. (2011). Work or Fun? How Task Construal and Completion Influence Regulatory Behavior. *Journal of Consumer Research*, *37*(6), 967–983. https://doi.org/10.1086/656576

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, *11*(1), 65–100.

Lazzaro, N. (2004). Why We Play Games: Four Keys to More Emotion Without Story. *Game Developer Conference (GDC)*, 1–8. https://doi.org/10.1111/j.1464-410X.2004.04896.x

LeHong, H., & Fenn, J. (2013). *Hype Cycle for Emerging Technologies, 2013*. https://www.gartner.com/en/documents/2571624#:~:text=A Gartner Hype Cycle provides,or hanging on too long.

Leinhardt, G., & Steele, M. D. (2005). Seeing the complexity of standing to the side: Instructional dialogues. *Cognition and Instruction*, *23*(1), 87–163. https://doi.org/10.1207/s1532690xci2301_4

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet*

Physics Doklady, *10*(8), 707–710.

Levin, J. R. (1988). Elaboration-based learning strategies: Powerful theory = powerful application. *Contemporary Educational Psychology*, *13*(3), 191–205. https://doi.org/10.1016/0361-476X(88)90020-3

Lieberoth, A. (2015). Shallow gamification: Testing psychological effects of framing an activity as a game. *Games and Culture*, *10*(3), 229–248. https://doi.org/10.1177/1555412014559978

Lim, H., Lee, S. G., & Nam, K. (2007). Validating E-learning factors affecting training effectiveness. *International Journal of Information Management*, *27*(1), 22–35. https://doi.org/10.1016/j.ijinfomgt.2006.08.002

LimeSurvey GmbH. (2003). *LimeSurvey*. https://www.limesurvey.org/

Lisitsyna, L. S., Pershin, A. A., & Kazakov, M. A. (2015). *Game Mechanics Used for Achieving Better Results of Massive Online Courses* (pp. 183–192). Springer, Cham. https://doi.org/10.1007/978-3-319-19875-0_17

Littleton, K., Ashman, H., Light, P., Artis, J., Roberts, T., & Oosterwegel, A. (1999). Gender, task contexts, and children's performance on a computer-based task. *European Journal of Psychology of Education*, *14*(1), 129–139. https://doi.org/10.1007/BF03173115

Liu, M., & Reed, W. M. (1994). The relationship between the learning strategies and learning styles in a hypermedia environment. *Computers in Human Behavior*, *10*(4), 419–434. https://doi.org/10.1016/0747-5632(94)90038-8

Lotfi, E., & Mohammed, B. (2014). Players performances analysis based on educational data mining case of study: Interactive waste sorting serious game. *International Journal of Computer Applications*, *108*(11), 13–18. https://doi.org/10.5120/18954-0217

Lounis, S., Pramatari, K., & Theotokis, A. (2014). *Gamification is all about fun: The role of incentive type and community collaboration*.

Loveland, K. K., & Olley, J. G. (1979). The effect of external reward on interest and quality of task performance in children of high and low intrinsic motivation. *Child Development*, 1207–1210.

Lowendahl, J.-M. (2014). *Hype Cycle for Education, 2014*. https://www.gartner.com/en/documents/2806424

Lucas, W., & Topi, H. (2004). Training for Web search: Will it get you in shape? *Journal of the American Society for Information Science and Technology*, *55*(13), 1183–1198.

Luo, Y., Zelenika, I., & Zhao, J. (2019). Providing immediate feedback improves recycling and composting accuracy. *Journal of Environmental Management*, *232*, 445–454. https://doi.org/10.1016/j.jenvman.2018.11.061

Lynch, T., & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research*, *4*(3), 221–250. https://doi.org/10.1177/136216880000400303

Maedche, A., & Staab, S. (2001). Ontology Learning for the Semantic Web. *IEEE Intelligent Systems*, *16*, 72–79. https://doi.org/10.1109/5254.920602

Mahmood, A. H., & Murad, I. M. A. (2018). Approaching the Language of the Second Language Learner: Interlanguage and the Models Before. *English Language Teaching*, *11*(10), 95–108. https://doi.org/10.5539/elt.v11n10p95

Maiden, N. (2008). *Requirements Engineering Lectures*.

Mäkitalo-Siegl, K., Kohnle, C., & Fischer, F. (2011). Computer-supported collaborative inquiry learning and classroom scripts: Effects on help-seeking processes and learning outcomes. *Learning and Instruction*, *21*(2), 257–266. https://doi.org/10.1016/j.learninstruc.2010.07.001

Marczewski, A. (2017). The ethics of gamification. *XRDS: Crossroads, The ACM Magazine for Students*, *24*(1), 56–59.

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, *50*(4), 370.

Mauri, M., Elli, T., Caviglia, G., Uboldi, G., & Azzi, M. (2017). RAWGraphs: A Visualisation Platform to Create Open Outputs. *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter*, 28:1--28:5. https://doi.org/10.1145/3125571.3125585

Max-Neef, M., Elizalde, A., & Hopenhayn, M. (1990). *Human scale development: an option for the future*. Development Alternatives Centre [Centro de Alternativas de Desarrollo](CEPAUR).

Mayer, R. E. (1980). Elaboration techniques that increase the meaningfulness of technical text: An experimental test of the learning strategy hypothesis. *Journal of Educational Psychology*, *72*(6), 770–784. https://doi.org/10.1037/0022-0663.72.6.770

Mayer, R. E. (2002). Multimedia learning. In *Psychology of Learning and Motivation* (Vol. 41, pp. 85–139). Academic Press. https://doi.org/https://doi.org/10.1016/S0079-7421(02)80005-6

Mayer, R. E. (2009). *Multimedia Learning* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511811678

Mayer, R. E. (2019). Computer games in education. *Annual Review of Psychology*, *70*, 531–549.

McClelland, D. C. (1967). The Achieving Society (1961). *Chapter*, *8*, 301–335.

McGuinness, D. L., Fikes, R., Hendler, J., & Stein, L. A. (2002). DAML+ OIL: an ontology language for the Semantic Web. *IEEE Intelligent Systems*, *17*(5), 72–80.

Menon, B. M., Unnikrishnan, R., Muir, A., & Bhavani, R. R. (2017). Serious game on recognizing categories of waste, to support a zero waste recycling program. *2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH)*, 1–8.

Merriam-Webster. (2020). *Definition of browse*. https://www.merriam-webster.com/dictionary/browse

Merriam-Webster Incorporated. (2022a). *Definition for "reward."* https://www.merriam-webster.com/dictionary/reward

Merriam-Webster Incorporated. (2022b). *Definition of "Classification."* https://www.merriam-webster.com/dictionary/classification

Merriam-Webster Incorporated. (2022c). *Definition of "Taxonomy."* https://www.merriam-webster.com/dictionary/taxonomy

Meske, C., Brockmann, T., Wilms, K., & Stieglitz, S. (2017). Social collaboration and gamification. In *Gamification* (pp. 93–109). Springer.

Might and Delight. (2013). *Shelter*. https://store.steampowered.com/app/244710/Shelter/

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *ArXiv Preprint ArXiv:1301.3781*.

Mikuskova, E. B., Hanák, R., & Cavojova, V. (2015). Appropriateness of two inventories measuring intuition (the PID and the REI) for Slovak population. *Studia Psychologica*, *57*(1), 63.

Miller, G. A., & Gildea, P. M. (1987). How children learn words. *Scientific American*, *257*(3), 94–99. https://doi.org/10.1038/scientificamerican0987-94

Misuraca, R., Faraci, P., Gangemi, A., Carmeci, F. A., & Miceli, S. (2015). The Decision Making Tendency Inventory: A new measure to assess maximizing, satisficing, and minimizing. *Personality and Individual Differences*, *85*, 111–116. https://doi.org/10.1016/j.paid.2015.04.043

Mitchell, B. L. (2012). *Game design essentials*. John Wiley & Sons.

Mohanty, H. (2015). *Big Data: An Introduction*. https://doi.org/10.1007/978-81-322-2494-5_1

Murray, Henry A. (2008). *Explorations in Personality*.

Murray, Henry Alexander. (1938). *Explorations in Personality: A Clinical and Experimental study of fifty men of college age*.

Myers, I. B., & Myers, P. B. (2010). *Gifts differing: Understanding personality type*. Nicholas Brealey.

Nacke, L. E. ., Bateman, C. ., & Mandryk, R. L. . (2011). BrainHex: Preliminary results from a neurobiological gamer typology survey. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6972 LNCS*, 288–293. https://doi.org/10.1007/978-3-642-24500-8_31

Nacke, L. E., Bateman, C., & Mandryk, R. L. (2014). BrainHex: A neurobiological gamer typology survey. *Entertainment Computing*, *5*(1), 55–62. https://doi.org/10.1016/j.entcom.2013.06.002

Nacke, L. E., & Deterding, C. S. (2017). The maturing of gamification research. *Computers in Human Behaviour*, 450–454.

Navarro-Prieto, R., Scaife, M., & Rogers, Y. (1999). Cognitive Strategies in Web Searching. *Human Factors & the Web*, *1999*, 1–12.

Neuberg, S. L., Judice, T. N., & West, S. G. (1997). What the Need for Closure Scale measures and what it does not: Toward differentiating among related epistemic motives. *Journal of Personality and Social Psychology*, *72*(6), 1396.

Newell, A., & Simon, H. A. (1972). *Human problem solving* (Vol. 104, Issue 9). Prentice-hall Englewood Cliffs, NJ.

Newman, R. S. (2000). Social Influences on the Development of Children{\textquotesingle}s Adaptive Help Seeking: The Role of Parents, Teachers, and Peers. *Developmental Review*, *20*(3), 350–404. https://doi.org/10.1006/drev.1999.0502

Nguyen, T. T., Harper, F. M., Terveen, L., & Konstan, J. A. (2018). User personality and user satisfaction with recommender systems. *Information Systems Frontiers*, *20*(6), 1173–1189.

Nickerson, R. C., Varshney, U., & Muntermann, J. (2013). A method for taxonomy development and its application in information systems. *European Journal of Information Systems*, *22*(3), 336–359. https://doi.org/10.1057/ejis.2012.26

Nickerson, R. S., & Adams, M. J. (1979). Long-term memory for a common object. *Cognitive Psychology*, *11*(3), 287–307. https://doi.org/10.1016/0010-0285(79)90013-6

Niles, J. D. W. (1977). Lamkin: the motivation of horror. *Journal of American Folklore*, 49–67.

Nintendo EPD. (2017). *The Legend of Zelda: Breath of the Wild*. https://www.zelda.com/breath-of-the-wild/

Nitsch, R., Bruder, R., & Kelava, A. (2016). Schülerhandlungen als Elemente fachdidaktisch motivierter

Kompetenzmodellierungen. *Journal Für Mathematik-Didaktik*, *37*(2), 289–317.

Norman, D. (2013). *The design of everyday things: Revised and expanded edition*. Basic books.

O'Neil, H. F., Wainess, R., & Baker, E. L. (2005). Classification of learning outcomes: Evidence from the computer games literature. *The Cirriculum Journal*, *16*(4), 455–474.

Ohlsson, S. (1996). Learning from performance errors. *Psychological Review*, *103*(2), 241. https://doi.org/10.1037/0033-295X.103.2.241

Olsson, C. M., Björk, S., & Dahlskog, S. (2014). The conceptual relationship model: understanding patterns and mechanics in game design. *Digital Games Research Association (DiGRA), Snowbird, Utah, USA (2014)*, 1–16.

Oviatt, S. (2006). Human-centered Design Meets Cognitive Load Theory: Designing Interfaces That Help People Think. *Proceedings of the 14th ACM International Conference on Multimedia*, 871–880. https://doi.org/10.1145/1180639.1180831

Oxford Dictionary. (2019). *Lexico Definition of Anime*. https://www.lexico.com/en/definition/anime

Page, M., Wilson, B. A., Shiel, A., Carter, G., & Norris, D. (2006). What is the locus of the errorless-learning advantage? *Neuropsychologia*, *44*(1), 90–100. https://doi.org/10.1016/j.neuropsychologia.2005.04.004

Panksepp, J. (2004). *Affective neuroscience: The foundations of human and animal emotions*. Oxford university press.

Parkin, A. J., & Streete, S. (1988). Implicit and explicit memory in young children and adults. *British Journal of Psychology*, *79*(3), 361–369. https://doi.org/10.1111/j.2044-8295.1988.tb02295.x

Parlett, D. (2016). What'sa ludeme? *Game & Puzzle Design*, *2*(2), 81–84.

Pek, J., Wong, O., & Wong, A. (2018). How to address non-normality: A taxonomy of approaches, reviewed, and illustrated. *Frontiers in Psychology*, *9*, 2104. https://doi.org/10.3389/fpsyg.2018.02104

Pellegrini, A. D. (2009). *The role of play in human development*. Oxford University Press, USA.

Perls, F., Hefferline, G., & Goodman, P. (1951). Gestalt therapy. *New York*, *64*(7), 19–313.

Perrig, W., & Kintsch, W. (1985). Propositional and situational representations of text. *Journal of Memory and Language*, *24*(5), 503–518.

Peukert, C., Pfeiffer, J., Meißner, M., Pfeiffer, T., & Weinhardt, C. (2019). Shopping in virtual reality stores: the influence of immersion on system adoption. *Journal of Management Information Systems*, *36*(3), 755–788.

Pfeiffer, J., Benbasat, I., & Rothlauf, F. (2014). *Minimally restrictive decision support systems*.

Piaget, J., & Cook, M. (1952). *The origins of intelligence in children* (Vol. 8, Issue 5). International Universities Press New York.

Pink, D. H. (2011). *Drive: The surprising truth about what motivates us*. Penguin.

Pinter, A. (2005). Task Repetition with 10-year-old Children. In *Teachers Exploring Tasks in English Language Teaching* (pp. 113–126). https://doi.org/10.1057/9780230522961_11

Pivec, M., Dziabenko, O., & Schinnerl, I. (2003). Aspects of game-based learning. *3rd International Conference on Knowledge Management, Graz, Austria*, *304*.

Plume, A., & van Weijen, D. (2014). *Publish or perish? The rise of the fractional author…*.

https://www.researchtrends.com/issue-38-september-2014/publish-or-perish-the-rise-of-the-fractional-author/

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. *Journal of Applied Psychology*, *88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Poels, K., Kort, Y. De, & IJsselsteijn, W. (2007). "It is always a lot of fun!": Exploring dimensions of digital game experience using focus group methodology. *Future Play 2007*, 83–89. https://doi.org/10.1145/1328202.1328218

Pramana, G., Parmanto, B., Lomas, J., Lindhiem, O., Kendall, P. C., & Silk, J. (2018). Using mobile health gamification to facilitate cognitive behavioral therapy skills practice in child anxiety treatment: open clinical trial. *JMIR Serious Games*, *6*(2), e8902.

Prather, D. C. (1971). Trial-and-error versus errorless learning: training, transfer, and stress. *The American Journal of Psychology*, *84*(3), 377–386. https://doi.org/10.2307/1420469

Purvis, A., Howell, R. T., & Iyer, R. (2011). Exploring the role of personality in the relationship between maximization and well-being. *Personality and Individual Differences*, *50*(3), 370–375.

Putnam, A. L. (2015). Mnemonics in education: Current research and applications. *Translational Issues in Psychological Science*, *1*(2), 130–139. https://doi.org/10.1037/tps0000023

Putz, L. M., Treiblmaier, H., & Pfoser, S. (2018). Field trips for sustainable transport education: Impact on knowledge, attitude and behavioral intention. *International Journal of Logistics Management*, *29*(4), 1424–1450. https://doi.org/10.1108/IJLM-05-2017-0138

Rammstedt, B., & 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. *Journal of Research in Personality*, *41*(1), 203–212.

Rapp, A., Hopfgartner, F., Hamari, J., Linehan, C., & Cena, F. (2019). Strengthening gamification studies: Current trends and future opportunities of gamification research. In *International Journal of Human-Computer Studies* (Vol. 127, pp. 1–6). Elsevier.

Reichwald, R., & Piller, F. (2006). Organisation der arbeitsteiligen Wertschöpfung: Entwicklungen und Trends auf dem Weg zur interaktiven Wertschöpfung. *Interaktive Wertschöpfung: Open Innovation, Individualisierung Und Neue Formen Der Arbeitsteilung*, 11–94.

Reiss, S. (2001). *Who am I - the basic 16 desires*. 250–262.

Reiss, S., & Sushinsky, L. W. (1975). Overjustification, competing responses, and the acquisition of intrinsic interest. *Journal of Personality and Social Psychology*, *31*(6), 1116.

Rhumbl. (2019). *Rhumbl*. https://rhumbl.com/

Rieber, L. P. (1996). Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. *Educational Technology Research and Development*, *44*(2), 43–58. https://doi.org/10.1007/BF02300540

Rieber, L. P. (2005). Multimedia Learning in Games, Simulations, and Microworlds. In R. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 549–567). Cambridge University Press. https://doi.org/10.1017/CBO9780511816819.034

Righi, C., James, J., Beasley, M., Day, D. L., Fox, J. E., Gieber, J., Howe, C., & Ruby, L. (2013). Card sort analysis best practices. *Journal of Usability Studies*, *8*(3), 69–89.

Ring, L. (2013). Towards an Assessment of Gamification: Comparing Game Design Elements with

Android Design Principles. *Umea's 16th Student Conference in Computing Science*, 49–60.

Robertson, M. (2010). *Can't Play, Won't Play'. Hide&Seek. October 6*. Web Log Comment. https://kotaku.com/cant-play-wont-play-5686393

Robertson, S. E., & Jones, K. S. (1976). Relevance weighting of search terms. *Journal of the American Society for Information Science*, *27*(3), 129–146.

Robertson, S., & Robertson, J. (2013). Mastering the Requirements Process Getting Requirements Right. *Work*. https://doi.org/0-201-36046-2

Robinson, D., & Bellotti, V. (2013). A Preliminary Taxonomy of Gamification Elements for Varying Anticipated Commitment. *Chi'13*, 1–6. http://gamification-research.org/wp-content/uploads/2013/03/Robinson_Bellotti.pdf%5Cnpapers3://publication/uuid/D7BED7E9-0534-4247-822E-CB9FABBCD21D

Rodrigues, Luis Filipe, Costa, C. J., & Oliveira, A. (2013). *The adoption of gamification in e-banking*. 47. https://doi.org/10.1145/2503859.2503867

Rodrigues, Luís Filipe, Costa, C. J., & Oliveira, A. (2016). Gamification: A framework for designing software in e-banking. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2016.04.035

Roelle, J., Berthold, K., & Renkl, A. (2013). Two instructional aids to optimise processing and learning from instructional explanations. *Instructional Science*, *42*(2), 207–228. https://doi.org/10.1007/s11251-013-9277-2

Rollings, A., & Morris, D. (2003). *Game architecture and design*. New Riders Publishing.

Roman, G.-C. (1985). A Taxonomy of Current Issues in Requirements Engineering. *Computer*, *18*(4), 14–23. https://doi.org/10.1109/MC.1985.1662861

Rose, J. A., O'Meara, J. M., Gerhardt, T. C., & Williams, M. (2016). Gamification: using elements of video games to improve engagement in an undergraduate physics class. *Physics Education*, *51*(5), 055007. https://doi.org/10.1088/0031-9120/51/5/055007

Ross, M. (1975). Salience of reward and intrinsic motivation. *Journal of Personality and Social Psychology*, *32*(2), 245.

Rouse, R. (2004). *Game Design: Theory and Practice, Second Edition: Theory and Practice, Second Edition*.

Rovio Entertainment. (2009). *Angry Birds*. https://www.angrybirds.com/

Rugg, G., & McGeorge, P. (1997). The sorting techniques: a tutorial paper on card sorts, picture sorts and item sorts. *Expert Systems*, *14*(2), 80–93.

Rugg, G., & McGeorge, P. (2005). The sorting techniques: A tutorial paper on card sorts, picture sorts and item sorts. *Expert Systems*, *22*(3), 94–107. https://doi.org/10.1111/j.1468-0394.2005.00300.x

Ryan, A. M., Pintrich, P. R., & Midgley, C. (2001). Avoiding seeking help in the classroom: Who and why? *Educational Psychology Review*, *13*(2), 93–114. https://doi.org/10.1023/A:1009013420053

Ryan, A. M., & Shin, H. (2011). Help-seeking tendencies during early adolescence: An examination of motivational correlates and consequences for achievement. *Learning and Instruction*, *21*(2), 247–256. https://doi.org/10.1016/j.learninstruc.2010.07.003

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation. *American Psychologist*, *55*, 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Sanmugam, M., Abdullah, Z., Mohamed, H., Mohd Zaid, N., Aris, B., & Van Der Meijden, H. (2016). The impacts of infusing game elements and gamification in learning. *2016 IEEE 8th International Conference on Engineering Education: Enhancing Engineering Education Through Academia-Industry Collaboration, ICEED 2016*, 131–136. https://doi.org/10.1109/ICEED.2016.7856058

Schamber, L., & Marchionini, G. (1996). Information Seeking in Electronic Environments. *Journal of Education for Library and Information Science*, *37*(1), 81. https://doi.org/10.2307/40324289

Schell, J. (2014). *The Art of Game Design: A Deck of Lenses*. https://doi.org/10.1201/b17723

Schenk, E., & Guittard, C. (2011). Towards a characterization of crowdsourcing practices. *Journal of Innovation Economics Management*, *1*, 93–107.

Schlechty, P. C. (1997). *Inventing better schools : an action plan for educational reform*. Jossey-Bass Publishers. https://eric.ed.gov/?id=ED403658

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, *3*(4), 207–218. https://doi.org/10.1111/j.1467-9280.1992.tb00029.x

Schultz, P. W., Oskamp, S., & Mainieri, T. (1995). Who recycles and when? A review of personal and situational factors. *Journal of Environmental Psychology*, *15*(2), 105–121. https://doi.org/10.1016/0272-4944(95)90019-5

Schwartz, B. (2016). On the meaning and measurement of maximization. *Judgment and Decision Making*, *11*(2), 126.

Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., & Lehman, D. R. (2002). Maximizing versus satisficing: happiness is a matter of choice. *Journal of Personality and Social Psychology*, *83*(5), 1178.

Scott, T. J., & O'Sullivan, M. K. (2005). Analyzing student search strategies: Making a case for integrating information literacy skills into the curriculum. *Teacher Librarian*, *33*(1), 21.

Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human-Computer Studies*, *74*, 14–31. https://doi.org/10.1016/j.ijhcs.2014.09.006

Sellers, M. (2017). *Advanced game design : a systems approach*. Boston : Addison-Wesley, 2017. https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1624977

Selten, R., & Stoecker, R. (1986). End behavior in sequences of finite Prisoner{\textquotesingle}s Dilemma supergames A learning theory approach. *Journal of Economic Behavior {\&}amp$\mathsemicolon$ Organization*, *7*(1), 47–70. https://doi.org/10.1016/0167-2681(86)90021-1

Sepehr, S., & Head, M. (2013). Competition as an element of gamification for learning: an exploratory longitudinal investigation. *Proceedings of the First International Conference on Gameful Design, Research, and Applications*, 2–9.

Serrat, O. (2017). The five whys technique. In *Knowledge solutions* (pp. 307–310). Springer.

Shaffer, D. W., Squire, K. R., Halverson, R., & Gee, J. P. (2005). Video games and the future of learning. *Phi Delta Kappan*, *87*(2), 105–111. https://doi.org/10.1177/003172170508700205

Sheldon, K. M. (2011). Integrating behavioral-motive and experiential-requirement perspectives on psychological needs: a two process model. *Psychological Review*, *118*(4), 552.

Sherman, C. B., Sherman, C., & Price, G. (2001). *The invisible Web: Uncovering information sources search engines can't see*. Information Today, Inc.

Sherry, J. L., Lucas, K., Greenberg, B. S., & Lachlan, K. (2006). Video Game Uses and Gratifications as Predictors of Use and Game Preference. *Playing Video Games: Motives, Responses, and Consequences*, *24*(2), 213–224. https://doi.org/10.4324/9780203873700

Shute, V. J., Ventura, M., Bauer, M., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. *Serious Games: Mechanisms and Effects*, *2*(l), 295–321. https://doi.org/10.4324/9780203891650

Sicart, M. (2008). Defining game mechanics. *The International Journal of Computer Game Research*, *8(2)*(2). http://gamestudies.org/0802/articles/sicart

Sicart, M. (2015). Playing the good life: Gamification and ethics. *The Gameful World: Approaches, Issues, Applications*, 225–244.

Silverstein, C., Henzinger, M., Marais, H., & Moricz, M. (1998). Analysis of a Very Large AltaVista Query Log. *Technical Report 1998-014, Systems Research Center, Compaq Computer Corporation*, 014. https://doi.org/http://doi.acm.org/10.1145/331403.331405

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1), 99–118.

Simons, A., Wohlgenannt, I., Weinmann, M., & Fleischer, S. (2020). Good gamers, good managers? A proof-of-concept study with Sid Meier's Civilization. *Review of Managerial Science*, *15*(4), 957–990. https://doi.org/10.1007/s11846-020-00378-0

Skinner, B. F. (1936). The Effect on the Amount of Conditioning of an Interval of Time before Reinforcement. *The Journal of General Psychology*, *14*(2), 279–295. https://doi.org/10.1080/00221309.1936.9713155

Smelcer, J. B., & Carmel, E. (1997). The effectiveness of different representations for managerial problem solving: comparing tables and maps. *Decision Sciences*, *28*(2), 391–420.

Solarski, C. (2015). The Aesthetics of Game Art and Game Design. In *Gamasutra*. http://www.gamasutra.com/view/feature/185676/the_aesthetics_of_game_art_and_.php?print=1

Soranzo, A., & Cooksey, D. (2015). Testing taxonomies: Beyond card sorting. *Bulletin of the Association for Information Science and Technology*, *41*(5), 34–39.

Squire, K. (2007). Games, learning, and society: Building a field. *Educational Technology*, *47*(5), 51–55.

Sreelakshmi, R., McLain, M. L., Rajeshwaran, A., Rao, B., Jayakrishnan, R., & Bijlani, K. (2015). Gamification to enhance Learning using Gagne's learning model. *2015 6th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 1–6. http://ieeexplore.ieee.org/document/7395197/

Stacey, P., & Nandhakumar, J. (2009). A temporal perspective of the computer game development process. *Information Systems Journal*, *19*(5), 479–497. https://doi.org/10.1111/j.1365-2575.2007.00273.x

Stanitsas, M., Kirytopoulos, K., & Vareilles, E. (2019). Facilitating sustainability transition through serious games: A systematic literature review. *Journal of Cleaner Production*, *208*, 924–936. https://doi.org/10.1016/j.jclepro.2018.10.157f

Steinkuehler, C., & Duncan, S. (2008). Scientific habits of mind in virtual worlds. *Journal of Science Education and Technology*, *17*(6), 530–543.

Stelmaszewska, H., & Blandford, A. (2004). From physical to digital: A case study of computer scientists' behaviour in physical libraries. *International Journal on Digital Libraries*, *4*(2), 82–92.

https://doi.org/10.1007/s00799-003-0072-6

Stewart, B. (2011). *Personality And Play Styles: A Unified Model*. https://www.gamasutra.com/view/feature/6474/personality_and_play_styles_a_.php

Sweller, J., Ayres, P., & Kalyuga, S. (2011). Cognitive Load Theory. In *Cognitive Load Theory*. https://doi.org/10.1007/978-1-4419-8126-4

Tailby, R., & Haslam, C. (2003). An investigation of errorless learning in memory-impaired patients: Improving the technique and clarifying theory. *Neuropsychologia*, *41*(9), 1230–1240. https://doi.org/10.1016/S0028-3932(03)00036-8

Tan, G. W., & Wei, K. K. (2006). An empirical study of Web browsing behaviour: Towards an effective Website design. *Electronic Commerce Research and Applications*, *5*(4), 261–271. https://doi.org/10.1016/j.elerap.2006.04.007

Tekinbas, K. S., & Zimmerman, E. (2003). *Rules of Play - Game Design Fundamentals*. MIT Press.

Thomas, C., & Sharp, V. (2013). Understanding the normalisation of recycling behaviour and its implications for other pro-environmental behaviours: A review of social norms and recycling. *Resources, Conservation and Recycling*, *79*, 11–20. https://doi.org/10.1016/j.resconrec.2013.04.010

Thompson, E. P., Roman, R. J., Moskowitz, G. B., Chaiken, S., & Bargh, J. A. (1994). Accuracy motivation attenuates covert priming: The systematic reprocessing of social information. *Journal of Personality and Social Psychology*, *66*(3), 474.

Thompson, M. M., Naccarato, M. E., & Parker, K. E. (1989). Assessing cognitive need: The development of the Personal Need for Structure and Personal Fear of Invalidity Scales. *Paper Presented at the Annual Meeting of the Canadian Psychological Association, Halifax, Nova Scotia, Canada*.

Tibshirani, R. J., & Efron, B. (1993). An Introduction to the Bootstrap. In *Monographs on Statistics and Applied Probability* (Vol. 57). Springer-Science+Business Media, B.V.

Toda, A. M., Valle, P. H. D., & Isotani, S. (2018). *The Dark Side of Gamification: An Overview of Negative Effects of Gamification in Education BT - Higher Education for All. From Challenges to Novel Technology-Enhanced Solutions* (A. I. Cristea, I. I. Bittencourt, & F. Lima (Eds.); pp. 143–156). Springer International Publishing.

Tondello, G. F., Wehbe, R. R., Diamond, L., Busch, M., Marczewski, A., & Nacke, L. E. (2016). The Gamification User Types Hexad Scale. *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*, 229–243. https://doi.org/10.1145/2967934.2968082

Tractinsky, N., Katz, A. S., & Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers*, *13*(2), 127–145. https://doi.org/10.1016/s0953-5438(00)00031-x

Travis, R. (2011). Practomimetic Learning in the Classics Classroom: a Game-based Learning Method from Ancient Epic and Philosophy. *New England Classical Journal*, *38*.

Tricot, A., & Sweller, J. (2014). Domain-specific knowledge and why teaching generic skills does not work. *Educational Psychology Review*, *26*(2), 265–283.

Tsoumakas, G., & Katakis, I. (2007). Multi-label classification: An overview. *International Journal of Data Warehousing and Mining (IJDWM)*, *3*(3), 1–13.

Tuunanen, J., & Hamari, J. (2012). Meta-synthesis of player typologies. *Proceedings of Nordic Digra 2012 Conference: Local and Global - Games in Culture and Society*.

Twitch. (2006). *Internet Game Database*. https://www.igdb.com/

Umanath, N. S., & Vessey, I. (1994). Multiattribute data presentation and human judgment: A cognitive fit perspective. *Decision Sciences*, *25*(5-6), 795–824.

United Nations. (2020). *SDG - Goal 12: Ensure sustainable consumption and production*. https://www.un.org/sustainabledevelopment/sustainable-consumption-production/

Unity Technologies. (2014). *Unity Analytics*. https://unity.com/products/unity-analytics

Universität Würzburg. (2019). *Virtuelle Bibliothek Würzburg*. http://vb.uni-wuerzburg.de/

Universitätsbibliothek Heidelberg. (2020). *Elektronische Zeitschriftenbibliothek*. http://rzbvm013.uni-regensburg.de/ezeit/search.phtml?bibid=AAAAA&colors=7&lang=de

Uthayan, K. R., & Anandha Mala, G. S. (2015). Hybrid ontology for semantic information retrieval model using keyword matching indexing system. *The Scientific World Journal*, *2015*.

Utz, S. (2000). *Social Information Processing in MUDs : The Development of Friendships in Virtual Worlds*. *Cmc*, 1–23.

Valve. (2007). *Portal (series)*. https://store.steampowered.com/app/400/Portal/

Van Aken, J. E. (2005). Management research as a design science: Articulating the research products of mode 2 knowledge production in management. *British Journal of Management*, *16*(1), 19–36.

Van Eck, R. N., Shute, V. J., & Rieber, L. P. (2017). Leveling Up: Game Design Research and Practice for Instructional Designers. In R. Reiser & J. Dempsey (Eds.), *Trends and Issues in Instructional Design and Technology* (pp. 227–285). Pearson.

Van Rees, R. (2003). Clarity in the usage of the terms ontology, taxonomy and classification. *CIB REPORT*, *284*(432), 1–8.

vBulletin Solutions Inc. (2019). *AnimeSuki*. https://forums.animesuki.com/

Veltkamp, M., Aarts, H., & Custers, R. (2008). On the emergence of deprivation-reducing behaviors: Subliminal priming of behavior representations turns deprivation into motivation. *Journal of Experimental Social Psychology*, *44*(3), 866–873.

Versionmuseum. (2020). *History of Google Search*. https://www.versionmuseum.com/history-of/google-search

Vessey, I., & Galletta, D. (1991). Cognitive fit: An empirical study of information acquisition. *Information Systems Research*, *2*(1), 63–84. https://doi.org/10.1287/isre.2.1.63

Vichitvanichphong, S., Talaei-Khoei, A., Kerr, D., Ghapanchi, A. H., & Scott-Parker, B. (2016). Good old gamers, good drivers: Results from a correlational experiment among older drivers. *Australasian Journal of Information Systems*, *20*, 1–21. https://doi.org/10.3127/ajis.v20i0.1110

Voigt, M., Niehaves, B., & Becker, J. (2012). Towards a unified design theory for creativity support systems. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7286 LNCS*(GroupSystems 2011), 152–173. https://doi.org/10.1007/978-3-642-29863-9_13

Völkl, A. (2000). *Anime-Community.de*. https://www.anime-community.de/

Vygotsky, L. S. (1967). Play and its role in the mental development of the child. *Soviet Psychology*, *5*(3), 6–18. https://doi.org/10.2753/RPO1061-040505036

Waybackmachine. (2020). *Google Directory*. http://directory.google.com

Webb, N. M., & Farivar, S. (1999). Developing productive group interaction in middle school mathematics.

In A. King & A. M. O'Donnell (Eds.), *Cognitive Perspectives on Peer Learning* (pp. 117–149). Lawrence Erlbaum Associates Publishers.

Webb, N. M., Ing, M., Kersting, N., & Nemer, K. M. (2013). Help seeking in cooperative learning groups. In *Help Seeking in Academic Settings* (pp. 45–88). Taylor and Francis. https://doi.org/10.4324/9780203726563

Webster{\textquotesingle}s New World College Dictionary (4th edition). (2005). *Reference Reviews*, *19*(8), 35–36. https://doi.org/10.1108/09504120510632705

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology*, *67*(6), 1049.

Webster, J., & Martocchio, J. J. (1993). Turning work into play: Implications for microcomputer software training. *Journal of Management*, *19*(1), 127–146. https://doi.org/10.1016/0149-2063(93)90049-S

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *Source: MIS Quarterly*, *26*(2). http://www.jstor.org/stable/4132319

Werbach, K., & Hunter, D. (2012). *For the win : how game thinking can revolutionize your business*. Wharton.

Wertheimer, M. (1938). *Gestalt theory.*

Wessel, C. (2010). *Semi-strukturierte Interviews im Software-Engineering: Indikationsstellung, Vorbereitung, Durchführung und Auswertung - Ein Fall-basiertes Tutorium.*

Westera, W., Nadolski, R. J., Hummel, H. G. K., & Wopereis, I. G. J. H. (2008). Serious games for higher education: A framework for reducing design complexity. *Journal of Computer Assisted Learning*, *24*(5), 420–432. https://doi.org/10.1111/j.1365-2729.2008.00279.x

Whalen, K. A., Berlin, C., Ekberg, J., Barletta, I., & Hammersberg, P. (2018). 'All they do is win': Lessons learned from use of a serious game for Circular Economy education. *Resources, Conservation and Recycling*, *135*, 335–345. https://doi.org/10.1016/j.resconrec.2017.06.021

Witt, M., Scheiner, C. W., & Robra-Bissantz, S. (2011). Gamification of online idea competitions: insights from an explorative case. *GI-Jahrestagung*, 392.

Wittwer, J., & Renkl, A. (2008). Why instructional explanations often do not work: A framework for understanding the effectiveness of instructional explanations. *Educational Psychologist*, *43*(1), 49–64. https://doi.org/10.1080/00461520701756420

Wolfe, M. B. W., Schreiner, M. E., Rehder, B., Laham, D., Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). Learning from text: Matching readers and texts by latent semantic analysis. *Discourse Processes*, *25*(2–3), 309–336. https://doi.org/10.1080/01638539809545030

Wood, J. R., & Wood, L. E. (2008). Card sorting: current practices and beyond. *Journal of Usability Studies*, *4*(1), 1–6.

World Digital Library. (2019). *World Digital Library.*

Yee, N. (2006). Motivations for play in online games. *Cyberpsychology & Behavior : The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society*, *9*(6), 772–775. https://doi.org/10.1089/cpb.2006.9.772

Zackariasson, P., Wåhlin, N., & Wilson, T. L. (2010). Virtual Identities and Market Segmentation in Marketing in and Through Massively Multiplayer Online Games (MMOGs). *Services Marketing Quarterly*, *31*(3), 275–295. https://doi.org/10.1080/15332969.2010.486689

Zentek, T., & Hoffmann, G. (2016). Analyse von gamification anreizmechanismen für crowd worker in participatory sensing. *Mensch Und Computer 2016-Tagungsband*.

ZeptoLab. (2010). *Cut the Rope*. www.zeptolab.com

Zhang, P., Tang, J., & Jeong, E. E. (2021). A Meta-review of Gamification Research. *International Conference on Information*, 361–373.

Zichermann, G., & Cunningham, C. (2011). *Gamification by design: Implementing game mechanics in web and mobile apps*. " O'Reilly Media, Inc."

Zillig, L. M. P., Hemenover, S. H., & Dienstbier, R. A. (2002). What do we assess when we assess a Big 5 trait? A content analysis of the affective, behavioral, and cognitive processes represented in Big 5 personality inventories. *Personality and Social Psychology Bulletin*, *28*(6), 847–858.

Zimek, A., Schubert, E., & Kriegel, H.-P. (2012). A survey on unsupervised outlier detection in high-dimensional numerical data. *Statistical Analysis and Data Mining*, *5*(5), 363–387. https://doi.org/10.1002/sam.11161