SAARLAND UNIVERSITY



Faculty of Mathematics and Computer Science Department of Computer Science Bachelor Thesis

"Enable or Disable Gamification?" Investigating the Effect of Choice in a Gamified Setting

^{submitted by} Lea Verena Schmeer Saarbrücken October 2019

Advisors

Dr. Pascal Lessel German Research Center for Artificial Intelligence, Saarland Informatics Campus, Saarbrücken, Germany

Maximilian Altmeyer German Research Center for Artificial Intelligence, Saarland Informatics Campus, Saarbrücken, Germany

Reviewers

Prof. Dr. Antonio Krüger, German Research Center for Artificial Intelligence, Saarland Informatics Campus, Saarbrücken, Germany

Dr. Michael Schmitz, Hochschule der Bildenden Künste Saar, Saarbrücken, Germany

Submitted July 12, 2019 Saarland University Faculty MI – Mathematics and Computer Science Department of Computer Science Campus - Building E1.1 66123 Saarbrücken Germany

Statement in Lieu of an Oath:

I hereby confirm that I have written this thesis on my own and that I have not used any other media or materials than the ones referred to in this thesis.

Saarbrücken, July 12, 2019

Declaration of Consent:

I agree to make both versions of my thesis (with a passing grade) accessible to the public by having them added to the library of the Computer Science Department.

Saarbrücken, July 12, 2019

Acknowledgements

On this occasion, I would like to thank Pascal Lessel and Maximilian Altmeyer for their immense support and engagement. Without their guidance, this venture would not have been possible. Furthermore, I would like to express my gratitude to Prof. Dr. Antonio Krüger for giving me the opportunity to write my bachelor thesis at the German Research Center for Artificial Intelligence, as well as to Dr. Michael Schmitz for reviewing this thesis.

My deepest appreciation goes to my family and my friends, for their moral support, their continuous encouragement and their help whenever it was needed.

Last but not least, a special thanks to everyone who participated in the study.

Preface

Please note that parts of the work presented in this thesis, including ideas, applications, studies as well as results, tables and conclusions have already been published as a conference paper:

Pascal Lessel, Maximilian Altmeyer, Lea Verena Schmeer and Antonio Krüger. 2019."Enable or Disable Gamification?"–Analyzing the Impact of Choice in a Gamified Image Tagging Task. In *Proceedings of the 37th Annual ACM Conference on Human Factors in Computing Systems (CHI '19).* ACM, Article 150, 12 pages. http://doi.acm.org/10.1145/3290605.3300380

Abstract

While the effects of tailoring gamification to users' preferences are found to be largely positive, customizing a gamification setup often takes time and effort on the user's part. Therefore, in this bachelor thesis, a very simple form of customization was explored: With only a single press of a button users chose whether or not they want to use game elements at all.

In a study, the effectiveness of this approach was investigated. Within the context of an online image tagging platform, based on previous work, one half of the participants chose to enable or disable gamification for their task after having seen both options in a tutorial. The other half did not have this choice and was assigned randomly to either the gamified or the non-gamified version of the task instead. To gain insights on how user performance and intrinsic motivation were impacted by having this choice, the overall tag quality and the amount of generated tags were compared between participants who deliberately chose to enable gamification and participants who were randomly assigned to this version, as well as between participants who chose to disable gamification and those randomly assigned to the non-gamified version.

The results indicate that, while having a choice did not have an effect on people who found the gamification setup to be appealing, it motivated those who did not want to use the game elements and, as a result, disabled gamification: they provided significantly more tags than participants who completed the same non-gamified task but without having a choice, demonstrating that the option to disable gamification poses an easy but useful customization approach.

Contents

1	Int	roduction	1
	1.1	Motivation	1
	1.2	Choice in a Gamified Setting	2
	1.3	Research Questions	3
		1.3.1 Significance of the Research Questions	3
		1.3.2 Overview of the Approach	4
	1.4	Outline of the Thesis	5
2	Rel	lated Work	6
	2.1	Autonomy and Choice	6
	2.2	Autonomy in Gamification	8
	2.3	Relation between Gamer Types and Game Elements	9
	2.4	Image Tagging	10
		2.4.1 Image Annotation Tasks as Game with a Purpose	10
		2.4.2 Using Image Tagging Tasks to Investigate Gamification	12
3	Stu	idy Platform	15
	3.1	Image Tagging Platform	15
	3.2	Additional Features	17
		3.2.1 Choice	17
		3.2.2 Tutorial	18
		3.2.3 Bonus rounds	20
		3.2.4 Questionnaires	20
	3.3	Implementation	21
4	Use	er Study	22
	4.1	Conditions	22
	4.2	Hypotheses	23
	4.3	Method	24
	4.4	Measures and Coding Process	28
	4.5	Participants	28
	4.6	Results	29
		4.6.1 Main Results	29

		4.6.2 Supporting Results	31		
	4.7	Discussion	36		
5	Cor	nclusion	38		
	5.1	Overview	38		
	5.2	Contributions	39		
	5.3	Limitations	39		
	5.4	Future Work	40		
A	Qu	estionnaires	42		
	A.1	Questionnaire on the Perception of the Game Elements	42		
	A.2	Intrinsic Motivation Inventory	43		
	A.3	Hexad	45		
	A.4	Conclusive Questionnaire	46		
B	B Set of Rules Formulated by the Independent Raters				
Bibliography					

Chapter 1 Introduction

Even though the effect of tailoring gamification to users' individual preferences was shown to be largely positive [17], a universal solution to this approach has not yet been found. While the concepts of personalization and customization serve as good prospects in this regard, they, too, have their drawbacks: Personalization, on one hand, is difficult to realize due to the sheer amount of factors that can influence users' preferences (e.g. [1, 3, 13, 15, 30]); customization, on the other hand, often increases user effort [24]. With this predicament in mind, we explore a simple and effortless form of customization, in which the users—with only a single press of a button—choose whether or not they want to use game elements at all, i.e. they choose to enable or disable a fixed gamification setup. In a user study, we examine how having this choice affected their performance and intrinsic motivation.

1.1 Motivation

Over the last decade, the concept of gamification, i.e. the employment of game design elements and game principles in a non-game context [8], has gained widespread popularity. This is not least due to the fact that it poses an easy and effective way to increase people's motivation and enjoyment of a task, sub-sequently making it a useful addition to, inter alia, fitness and health related systems [14], or even crowdsourcing systems [32].

While extensive research confirmed the positive effects of gamification in various contexts [11], the findings implicate that the game design elements are perceived differently from user to user: In particular, the perception of game design elements was found to depend largely on factors like users' gender [15], their age [3], culture [1], player type [30] and their personality traits [13]. Consequently,

game design elements that are appealing to some people might not be appealing to others, and in turn, the same applies to different gamification setups: If a user dislikes the employed elements within a specific system, the positive effects of gamification are reduced [25], showcasing that one-size-fits all approaches for gamified systems can be fairly problematic.

Consequently, in an attempt to overcome this problem and to increase the effectiveness of gamification, the tailoring of gamification setups is being investigated [16, 17, 25]. Contrary to fixed game designs, a tailored setup is supposed to cater to each user's individual needs, providing a more fitting, well-perceived user experience. One possible way to tailor a gamification setup is by personalizing it. That is, having the system adapt itself automatically based on people's likes and preferences [24]. This approach is difficult to put into practice, however, because optimally all aforementioned factors (gender, age, culture, etc.) would have to be taken into account. Existing personalization approaches often focus solely on player type or personality traits, resulting in inappropriate setups for some users, still [4]. To our knowledge, an all-encompassing universal solution to this approach has not yet been found.

Another possibility for tailoring gamification is to allow for customization within the system, i.e. to give users the possibility to adapt the game design elements and game setup to their liking at runtime. While sophisticated customization approaches yielded significantly positive results in recent studies [16, 17], they, too, are problematic. The configuration of game elements can be a very lengthy, complex process, taking time and effort on the users' part. Additionally, providing too many configuration options at once was shown to result in "choice overload" [12], where the users felt overwhelmed by a too wide range of choices and were, in turn, affected negatively.

As far as we are aware, simple, non-sophisticated customization approaches that aim to minimize user effort have not yet been investigated, raising the question of whether positive effects can be achieved when using only a small selection of customization options. To answer this question, we present a system which applies a very simplified customization approach, where the users are given only a bare minimum of configuration options. They can neither select nor adapt single game design elements, but can decide, for a fixed gamification setup, whether or not they want to use all game elements, i.e. they have the single choice of whether they want to enable or disable gamification. Investigating this approach is doubly interesting, as it not only allows us to explore the feasibility of low-sophistication customization but, even further than that, enables us to gain insights on how having a choice affects users' behavior in this gamified context.

1.2 Choice in a Gamified Setting

Choice is, by nature, an integral part of customizable gamification: Users are provided with different game design elements and configuration options that

enable them to make choices, based solely on their personal preferences, on how they want to adapt the system.

Previous studies already confirmed that providing people with a choice is beneficial in various regards: Not only do people feel less anxious when being offered a choice [29], but also, their performance increases considerably [35]. Additionally, for customizable gamification approaches, with users being allowed to freely choose different aspects of the game setup for themselves, we expect them to feel more autonomous as compared to one-size-fits-all approaches, where no choices are allowed. Since, as stated by the Self-Determination Theory (SDT) [27], the need for autonomy poses an important basis of intrinsic motivation and behavior (see Section 2.1), an increased feeling of autonomy should affect people's intrinsic motivation positively. Consequently, in the case of customizable gamification, being allowed to make autonomous choices regarding the game elements and the game setup should be perceived as intrinsically motivating.

Taking this, as well as the general benefits of choice into account, the question arises of whether having a choice in itself might be a substantial factor in the success of customizable gamification intervention. In order to gain a better understanding of this issue, the main focus of this thesis is to examine the impact of choice in the herein presented gamified system.

1.3 Research Questions

The goal of this thesis is to investigate the following research questions:

- **RQ1.** Does having the choice to enable or disable a fixed gamification setup increase users' intrinsic motivation?
- **RQ2.** Does having the choice to enable or disable a fixed gamification setup positively affect users' performance?

1.3.1 Significance of the Research Questions

Formulating our research questions in this way and specifically using a system for which users can only decide to enable or disable a fixed gamification setup enables us to investigate the effect of choice in an appropriate manner. This would not be the case with more sophisticated configuration options: The amount and complexity of the configuration options ultimately determine to which extent a system is customizable, i.e. the system's degree of customization. The more elements can be adapted in a detailed manner, the higher the system's degree of customization, and, the other way around, the less customization options are provided and the less complex they are, the lower the system's degree of customization. A system with a high degree of customization allows for a very precise configuration, as users can adjust every aspect of the gamification. However, with each decision the users are asked to make on their gamification setup, the configuration process gets more and more time-consuming and the risk of "choice overload" increases. Furthermore, with a high-degree-of-customization approach, drawing conclusions about the effect of choice specifically becomes virtually impossible, as one would have to differentiate between the impact of choice and the impact of using a game configuration that is fully adapted to the users' liking.

To resolve these limitations, the customization approach presented in this thesis represents the lowest degree of customization. As described above, it consists of only the single, elementary configuration option to enable or disable gamification, which can be completed in a single step, with the single click of a button. Not only do we not have to worry about "choice overload", but, since the users can only decide whether or not they want to use gamification at all—without being able to change the game setup itself in the slightest—the risk of introducing unwanted additional variables is minimized as well. This way, any positive effect we find for our system can unambiguously be attributed to the users benefitting from having a choice.

1.3.2 Overview of the Approach

In line with RQ1 and RQ2, we conduct a study on how being offered the choice to enable or disable gamification affects people's performance and motivation, basing the underlying system on an effective gamification setup from previous work [21] (explained in more detail in Section 2.4.2). In their research, the authors investigated the effects of individual game elements on an online image tagging platform, where participants completed either a gamified or a non-gamified version of the same image tagging task. They found that their employed game elements were perceived positively by participants, as those who completed the gamified version of the task produced more tags than participants who did the non-gamified version of the task. For this reason, as well as for the fact that both of these task versions already existed in a well-constructed manner, we adopt parts of their platform and image tagging task for our approach.

We then conduct our study in the following way: Like in [21], we differentiate between a non-gamified and a gamified task version and spit participants in three conditions. In a non-gamified condition, participants are given neither gamification nor choice and have to complete the plain, non-gamified version of the image tagging task. In a gamified condition, participants are not given a choice either, but complete the gamified version of the task. Lastly, and most importantly, we add a third condition, i.e. a choice condition, in which participants are allowed to choose for themselves whether they want to complete the image tagging task with or without using gamification. With this setup, we are able to analyze the effect of having a choice by comparing participants who have a choice and use it to enable gamification to participants who complete the same gamified task without having a choice, as well as participants who have a choice and use it to disable gamification to those who are assigned to the non-gamified task without having a choice. The amount of generated tags, as well as the overall tag quality serve as indicators of participants intrinsic motivation and performance.

1.4 Outline of the Thesis

In the following chapter, we take a look at existing research on gamification and customization approaches, and review related work on choice in various contexts. Subsequently, we give a detailed description of the implemented image tagging platform in which we examine the individual versions of the tagging task closely and compare them to the initial implementation from [21]. After formulating our hypotheses, we explain the method and conduction of our user study, which we follow up by presenting the results and discussing their significance. In a concluding chapter, we argue about possible limitations of the study, and why, based on our findings, offering users the choice to enable or disable gamification is a feasible and useful form of tailoring that should be more commonly employed.

Chapter 2 Related Work

In this chapter, we give an overview on related research which influenced the subject of this thesis and guided the conception of our study. To this end, we take a closer look at existing works on autonomy and choice in general, followed by works discussing the effects of autonomy in a gamified setting. Afterwards, in order to better understand how individual and contextual differences can impact the perception of gamification, we present research on how game elements affect the individual gamer types. Lastly, we also consider different approaches for creating an image annotation platform—as this is an integral part of our user study—and review existing implementations.

2.1 Autonomy and Choice

The importance of autonomy as a fundamental basis of intrinsic motivation has been highlighted in previous research. One prominent example of this is the Self-Determination Theory (SDT) by Deci et al. [27], which identified the need for autonomy as an innate and instinctive necessity, and further, as one of the three basic psychological needs (together with the need for competence and the need for relatedness) that, only if satisfied, foster well-being and growth, allowing for people to feel more intrinsically motivated.

Similarly, Zuckerman et al. researched the relation between self-determination and intrinsically motivated behavior [35] and conducted the following experiment: Participants were given a time frame of 30 minutes to solve three puzzle tasks. Before starting the experiment, however, all participants were assigned to one of two conditions, a task-choice condition or a task-no-choice condition, and were each paired with a participant in the respective other condition. Participants in the task-choice condition chose the puzzle they wanted to start with, and specified how much of their time they wanted to allocate to it. After completion (or after the allocated time was up), the same selection process was repeated for the second and third puzzle. Afterwards, the participants in the task-no-choice condition completed the same task; however, instead of getting to choose puzzle order and time allotment themselves, they were assigned the configurations of their preceding task-choice partner. After completing the task, all participants were left alone in the room for eight minutes. During this time, they could either read magazines, continue solving puzzles, or do something else entirely. The amount of time out of the eight minutes that they spent on the free choice task of solving puzzles was used as measure of their intrinsic motivation. It was found that participants in the task-choice condition spent significantly more time on the puzzles than those in the task-no-choice condition, showing that they were indeed more intrinsically motivated. In order to make sure that the greater intrinsic motivation effect was not mediated by performance differences, i.e. task-choice participants selecting puzzles that were uniquely easier for them, the average time needed to solve a puzzle and the average number of unsolved puzzles were analyzed, and no differences between the two conditions could be detected for these measures, so that the significant main effect for choice remained.

These results indicate that people who are given the freedom of choice—whether it be the choice of what to do or the choice of how to do it—are more intrinsically motivated for their activity than people who do the exact same activity without having a choice. As a consequence, we aim to find similar effects for our implemented image tagging platform: We hypothesize that users, who are given the choice to either enable or disable the gamification intervention for the image tagging task, feel inherently more intrinsically motivated, as their need for autonomy is satisfied. By extension, due to their increased intrinsic motivation, we expect them to perform significantly better than those, who do the same image tagging task but without having a choice. This assumption is further backed by findings from Amabile et al., who, in a study similar to [35], were able to derive positive effect of choice on intrinsic motivation in a learning environment. They had preschool children work on collages, but only half of them were allowed to choose the materials themselves, while the other half could only use materials as specified by the experimenters. The collages were rated independently by artist judges and compared between conditions, showing that collages done by children in the choice condition were far more creative. Furthermore, two weeks after the initial experiment, children in the choice condition showed a greater interest in doing collages again than those in the no-choice condition.

In conclusion, autonomy and choice have been shown to be beneficial to user motivation in different contexts and for various tasks. While this recognition already provides reasonable grounds to expect positive results for the herein presented approach, an in-depth review of research on the effects of autonomy within gamified settings is still necessary. Therefore, in the following section, we take a closer look at related works which combine autonomy and gamification by integrating choice into gamified systems.

2.2 Autonomy in Gamification

Nicholson [23] presented choice as a valuable concept of meaningful gamification, i.e. gamification that, in a specific context, engages and motivates the users in a way that allows for long-term changes of behavior: He argued that, instead of trying to increase users' motivation through rewards, gamified systems should aim to be more intrinsically motivating, because intrinsic motivation is an important factor in encouraging certain behaviors. As a consequence, he proposed to—in line with the SDT—put more control in the hands of the users, allowing for a greater feeling of autonomy. To achieve this shift of control from the system to the users, he discussed different ways of integrating choices into a gamification setup; for example, letting the users choose for themselves which activities they want to complete or which goals they want to set.

Taking it even further than that, Lessel et al. [16] introduced the concept of bottomup gamification. Bottom-up gamification, contrary to fixed gamification setups, allows for users to gamify a system to their liking by providing them with the option to choose fitting game elements for themselves at runtime. With this approach, users are free to make autonomous decisions in regards to the gamification intervention. To determine whether people would generally be interested in using bottom-up gamification in their everyday life, and by extension, if the increase of autonomy would be perceived positively, Lessel et al. conducted an online survey where people indicated to be overall open to the approach. The survey was followed by a twelve-day-long user study with a prototype of a bottom-up task management application. In the app, users could, for instance, name and categorize tasks, prioritize them, add due dates; most importantly though, they had the option to add game elements whenever a new task was created. For the most part, these could be combined freely and without restriction, allowing users to create game setups based on their own personal preferences. Before, during and after the study, users completed questionnaires to assess their gaming affinity, perception of the app, and experience with the app respectively and additionally, all interactions with the app were logged.

The result was that participants created far more gamified tasks than nongamified tasks. Moreover, all participants stated that they perceived being able to choose game elements for themselves positively. A majority of them even agreed that they would want to keep using the application in their everyday life. In conclusion, the users greatly appreciated being able to adapt the application to their liking. Even though configuring game elements and making adjustments required a lot more effort than simply using a top-down game configuration, participants readily and actively made use of the system. This observation is a first indication that users felt more motivated due to the higher level of autonomy in the bottom-up gamification approach and consecutively, that the positive findings for choice do extend to gamified settings, further reinforcing our hypothesis. While users did show great interest for bottom-up gamification in [16], the study lacked an actual comparison with 'regular' gamification. Consequently, in a second study, Lessel et al. compared the effectiveness of their approach with the effectiveness of top-down gamification [17]. For this purpose, they created an online crowd-sourcing platform, where participants had to complete microtasks involving receipts. The participants were assigned to either a baseline condition without any game elements, a top-down condition with fixed game elements, a bottom-up condition, or one of two mixed conditions which combined both approaches. They were then asked to solve as many microtasks as they wanted. Using the amount of solved tasks as a measure of motivation, the authors were able to provide evidence that participants solved more tasks when they had customized the game configurations to their liking, thus reconfirming the positive effect of bottom-up gamification. However, the authors were unsure of true reason behind these results: While at first, they assumed that participants who actively made use of the customization options were motivated by the particular game elements they selected (i.e. that they simply chose the most suitable and effective game setup for themselves), they later raised the question of whether participants being allowed to make choices for themselves might have had an impact on their findings to some degree, as well.

Their questioning motivated the research presented in this thesis: Since Lessel et al. were, due to the ambiguity of their findings, unable to draw clear conclusions for their approach, we decided to examine the effect of choice in a gamified setting more isolatedly, as this seemed to be the logical next step. As a result, their work initially shaped the design of our user study and influenced the conception of our image tagging platform to a great extent.

2.3 Relation between Gamer Types and Game Elements

In both experiments by Lessel et al., not all participants made use of the offered choices [16, 17]: Even though many different customization options were available to them, they kept their initial game configuration for the entire duration of the experiments. Lessel et al. argued that participants who did not make any changes to the game setup were either not interested in the gamification intervention, or did not notice the game elements in the first place [17]. However, another reason why some participants were more interested in making choices than others might be that their attitudes towards the different customization options varied. More specifically, while the SDT and related works praise autonomy to be highly beneficial, its efficacy within gamified settings might hinge on additional, user-dependent factors.

This assumption would fit in with previous research on gamification, where the users' perception of different game elements and principles was, as mentioned above, shown to largely depend on their age [3], gender [15], cultural background [1], personality [13] and player type: Orji et al. [26], for example, conducted a

study on the relation between player types and individual persuasion strategies and provided clear indications that people's perception of game elements was moderated by their BrainHex player type [22], confirming the need for adaptability of gamified systems. Similarly, Tondello et al. [30] emphasized the importance of considering users' needs in the design of gamification approaches and, as a result, focused their research on accurately mapping user preferences to Hexad player types. The Hexad framework [19] is comprised of six different user types, each of them modeled after people's extrinsic and intrinsic motivations. More specifically, the framework distinguishes between Philanthropists, i.e. people who are motivated by purpose, Socialisers, who are motivated by relatedness, Free Spirits, who like freedom and autonomy, Achievers, who seek competence, Players, who are motivated by extrinsic rewards, and finally Disruptors, who gain motivation from triggering changes. In order to accurately assign people to their Hexad user type, Tondello et al. developed the User Types Hexad Scale, which consisted of 24 items in total, all of them relating to a specific user type and measuring people's perception thereof. In a user study, they used their scale to determine the correlation between user types and the perception of various game design components and were able to find that all Hexad user types preferred different game elements.

The recognition that the users' perception of game elements is affected by their player type impacted our approach twofold. For one, we decided to investigate whether different player types were more likely to enable or disable the gamification intervention in our user study, which is why we opted to include the User Types Hexad Scale. Secondly, we took special care to design the gamified task appropriately: Since we required the applied gamification intervention to have a positive effect on participants' behavior, we considered game elements which had no negative connotations for any of the player types.

2.4 Image Tagging

In this section, we will take a closer look at related works which impacted the conception and design of our online image tagging platform. To this end, we present previously applied image annotation tasks and discuss their suitability in regards to our research topic and user study.

2.4.1 Image Annotation Tasks as Game with a Purpose

Image tagging tasks have often served as a framework for applying game design elements and principles (e.g. [21, 31, 33]). Ahn and Dabbish [32], for example, introduced several gamified image annotation tasks in an attempt to obtain meaningful, accurate labels for images on the Internet. Since the tagging of images is generally difficult to automate, they implemented various systems that rely on actual people to do the task instead. Consequently, in order to make the image tagging more attractive, Ahn and Dabbish applied various gamification elements and game principles, resulting in enjoyable, competitive and fast-paced 'games with a purpose' (GWAP).

One of these GWAPs was the ESP game, a two-player online image annotation game [31]. The game setup can be described as follows: Two random players were assigned as partners. Both were shown the same image, and had to guess what their partner was typing for the image, i.e. they had to agree on it. However, they were not allowed to communicate with each other, so that they only had the shared image to go off on. Once they both entered the same string, they moved on to the next image, which was again processed in the same way. Ultimately, the players' goal was to agree on as many images as possible. An analysis of the user statistics showed that the ESP game was very popular in general and even further, that some players had played it for more than 50 hours. These statistics clearly demonstrate that the image tagging was perceived as enjoyable by users, showcasing its suitability in the context of gamification.

A second GWAP for image annotation is Peekaboom [33]. Instead of collecting descriptive labels, Peekaboom aimed to locate objects within an image. Similar to the ESP game, Peekaboom was implemented as a two-player collaborative game. One player played as 'Peek', and the other one played as 'Boom'. While Boom was shown an image and the name of some depicted object from the start, Peek could only see those parts of the image that Boom chose to reveal (see Figure 2.1). The players' goal was for Peek to guess the name of Boom's object, so naturally, Boom would reveal only those areas that were associated with the object. Once Peek guessed the word correctly, the players switched roles and got a new image and object name. With this approach as well, Ahn and Dabbish were able to show the feasibility of gamified image annotation tasks.



Figure 2.1: Screenshots of Peek (on the left) and Boom (on the right) which are taken from [33].

In summary, even though the GWAPs were developed outside of empirical studies, they nevertheless provide a good overview on possible implementations for enjoyable gamified image tagging tasks and, more importantly, demonstrate the potential of using image annotation as a basis for investigating different gamification approaches.

2.4.2 Using Image Tagging Tasks to Investigate Gamification

To broaden our understanding on how to implement gameful image tagging tasks for only a single user (since GWAPs were designed for two players only) and further than that, on how to sensibly employ such tasks within empirical user studies, we considered the research by Feyisetan et al. [9], who examined the effects of gamification in a paid, crowd-sourced microtask setting. In this setting, they compared a non-gamified, state-of-the-art image-annotation task against a self-developed gamified one. Their self-developed task, called Wordsmith, incorporated points, levels, leaderboards and badges as game elements, as well as a progress bar, and alerts for when a goal was in reach. They furthermore provided an image skip feature and the option to terminate Wordsmith at any time to make sure the task completion was not perceived as mandatory, but rather as something the participants actively chose to do.

In one of their experiments, participants were assigned to either the non-gamified image annotation task or Wordsmith, and asked to tag a single image with two labels. After completion, participants in the Wordsmith condition were allowed to continue labeling images if they wanted to; however, they were not paid for any additional work. Nonetheless, the authors found that each participant in the Wordsmith condition exceeded the amount of specified work by far. Additionally, when evaluating the produced tags in terms of quality, they discovered that labels from the Wordsmith condition were overall more descriptive. Consequently, the results indicate a positive effect of gamification on users' performance and especially on their intrinsic motivation. However, for their approach as well, it can be argued that—since participants in the Wordsmith condition could decide for themselves how long they wanted to keep annotating images—autonomy and choice played a role in Feyisetan et al.'s experiment setup, and thus might have influenced their findings to some extent.

Apart from strengthening our assumption that having a choice might benefit participants within gamified settings, their research, or more specifically their image tagging task, gave us some idea of how to implement a suitable system ourselves. Since the results for Wordsmith were positive throughout the entire experiment, we believe that, even if autonomy and choice played a role for their findings, the game elements and principles were overall perceived well by participants, which was—as mentioned above— one of the main concerns for our user study. Consequently, for the conception of our implementation, we mainly took their applied game elements, i.e. points, levels, leaderboards and badges, into consideration.

Lastly, we looked at a user study by Mekler et al. [21], in which the effects of individual game elements, namely points, levels and leaderboards, on users' intrinsic motivation and performance were explored. For this purpose, they reused an online platform that was implemented for a previous study [20], on which abstract paintings could be labeled with regard to their emotional content: The platform consisted of 15 images of abstract paintings, which were shown, one after the other, to the participants. They only stayed visible for five seconds, after which they would disappear and could not be viewed again. Instead, a textfield was revealed, where users could provide tags, in form of keywords, on the overall mood of each respective image. With this non-gamified baseline task in place, Mekler et al. additionally created three gamified versions of the same task, using different gamification approaches which combined non-customizable game elements. In one task version, users were awarded points for each submitted tag, with their score being displayed throughout the entire tagging process. The second version also awarded points per tag, but further used a leaderboard which allowed users to compare their score to other (fictitious) scores (as depicted in Figure 2.2). Lastly, the third version consisted of a combination of points and levels, i.e. the users collected points for each entered tag and moved up the levels as their score hit certain thresholds.

Having these four versions of the image tagging task available on the platform, Mekler et al. conducted their study by randomly splitting participants between them, resulting in four distinct conditions, namely a plain condition, a points condition, a points plus leaderboard condition, as well as a points plus levels condition. In each condition, participants were asked to complete their respective image tagging task and to provide as many tags as possible.





Figure 2.2: Screenshot of the gamified image tagging task that was used in the points plus leaderboard condition (taken from [21]).

The authors then used the amount of generated tags as a measure of their performance, and additionally assessed the tag quality by letting two independent evaluators rate all tags in terms of how well they fit the images. After they finished the task, participants also completed a questionnaire, determining their intrinsic motivation, the satisfaction of their need for competence and autonomy, as well as their general causality orientation. The causality orientation measured participants general motivational orientation, meaning to which extent they were oriented towards simulating their intrinsic motivation, to which extent they were oriented towards being controlled by external rewards and to which extent they believed that achieving desired outcomes is beyond their control [7].

The results showed a significant effect of game elements on the amount of generated tags: Participants in the gamified conditions produced more tags than participants in the non-gamified condition and, further than that, participants in the leaderboard and level conditions outperformed those in the points condition. What is more, participants' performance did not suffer under the higher amount of generated tags, as no change in quality could be observed across conditions. Surprisingly, the results for the questionnaires showed that participants' intrinsic motivation was not affected by any of the game elements, thus reaffirming our hypothesis that, in order to effectively increase the intrinsic motivation within a gamified setting, means of autonomy and choice need to be considered.

Nevertheless, all three of their gamified approaches fulfilled the main requirement for our image tagging task, since they overall affected participants' behavior positively. Furthermore, their platform provided an additional, well thought out, non-gamified version of the same image tagging task, as well as sensible measures for participants' motivation and performance, so that it already suited our approach very well. For this reason, we decided to mostly base the implementation for our platform on theirs and, in the following chapter, present the resulting online image tagging platform in a more detailed manner.

Chapter 3 Study Platform

In this chapter, we discuss the concept of the platform that was implemented to conduct our user study on. We decided on using a system and context for which we already knew gamification to have a positive impact on users' behavior and, as a consequence, chose to adopt Mekler et al.'s online image tagging platform [20, 21]. The resulting implementation is described in the following sections.

3.1 Image Tagging Platform

We replicated the following aspects: We kept the context of affective image classification and used the same set of images, originally taken from [18] (see Figure 3.1). They mostly consisted of abstract art, depicting different colors and textures; more specifically, they did not depict objects that would evoke certain emotions in the viewer, i.e. all images were open to a free interpretation [18]. We also complied with Mekler et al. concerning their ordering for the tagging task, showing the 15 images at random.

In terms of the image tagging task, we kept Mekler et al.'s plain, non-gamified version as a baseline for our platform. Additionally, since our intention was to investigate the effect of having a choice to enable or disable a fixed gamification setup and therefore, offering more than one gamified task version was unnecessary, we decided to implement only the gamification intervention of points and a leaderboard. We opted for points and a leaderboard specifically because the positive motivational effects found in [21] were more significant for this combination than when only points were used; another reason was that these elements are widely popular in general [11]. It is important to note that on Mekler et al.'s platform, the social aspect of the leaderboard was faked. It's entries, i.e.



Figure 3.1: Examples of images used on the platform (taken from [18]).

usernames and other players' scores, were completely fictive and therefore equal for all platform users, giving everyone the same chance of rising in ranks (the users being left unaware of this fact). We implemented the leaderboard in the same way but, as opposed to Mekler et al.'s 1000, 3000, 6000 and 1000 points, displayed scores of 1300, 3100, 6300 and 9500 points for the fake players. We felt that these values were more authentic, while still being close to the original scores—especially in the sense that the gaps between the individual scores were more or less consistent and getting to the first position took a considerate amount of work. Since, like in [20, 21], participants were awarded 100 points per generated tag, 96 tags in total were needed to get to the top.

Another feature that we adopted from Mekler et al.'s platform was the tutorial preceding the main image tagging task. This tutorial was mandatory for all users and was identical to the main task in format and context, i.e. abstract images were shown for five seconds, then disappeared, and users were asked to provide tags on the depicted mood. By completing this tutorial, the users could familiarize themselves with the tagging process, and in particular, adapt to the images suddenly disappearing after five seconds. Compared to the tutorial provided on the original platform, we made some adjustments to our implementation; these will be discussed in detail later in this chapter.

Lastly, we also copied the overall design and user interface of Mekler et al.'s platform (Figure 3.2). In particular, the images were shown on the left side of the screen; right below, a countdown indicated how much time was left to view the respective image, counting all the way down from five seconds. After five seconds, the image disappeared with a flipping animation, revealing the textfield in its stead, where platform users could then enter their tags. Note that for the tagging itself, no time limit was set, so that the users could take as much time as they needed to come up with tags. Only after they decided to proceed, the next image was shown. If users completed the gamified version, the game elements were displayed on the right side of the screen, with the personal score right above the leaderboard. Besides these interface design, we also reused most of the existing instructions and explanations by Mekler et al., which is why we decided to make the platform available only in German.



The image will disappear in 5 seconds

Figure 3.2: Screenshot of the gamified image tagging task on our platform. For the sake of this thesis, the platform was translated into English.

3.2 Additional Features

We tried to keep our platform close to the original in order to replicate Mekler et al.'s findings in respect to tag quantity and quality. Moreover, we concluded that using the same image tagging tasks and metrics would allow for an in-depth comparison between results, enabling us to put our approach into perspective. Nevertheless, we had to make adjustments to the platform to account for our user study later on and, in particular, added additional features and functionalities.

3.2.1 Choice

Since Mekler et al. used their platform to investigate the effect of individual, fixed game elements exclusively, it was not necessary for them to provide users with a choice regarding the gamification intervention. Therefore, we had to conceptualize this feature from scratch, paying close attention to two aspects in particular. Firstly, while we wanted to provide platform users with a choice, we did not want their choice to affect the gamification setup itself. Allowing users to adapt the individual game elements would have made it difficult for us to investigate the effect of choice specifically, since the resulting customized setup would likely affect users' motivation and performance positively on its own [16, 17]. Secondly, in order to avoid "choice overload" [12] having a negative impact, we aimed to keep user effort to a minimum. In other words, our goal was to offer a choice that could be made fast and easily, but would not change the individual game elements or the general setup of our gamification intervention.

Ultimately, we opted for the choice to enable or disable gamification as a whole. We let the platform users choose on their own which version of the tagging task they wanted to complete; either the plain, non-gamified version or the gamified version with points and a leaderboard.

In the introduction, you have seen two different versions of the tagging task. Now you can decide which version of the study you would like to take.							
Click Keep second version to maintain the last	version you have seen or click Use first version to take	e the version you saw					
first.							
As soon as you decide, the study starts.							
asse enter as many keywords as you can think of to describe a overall mood in the respective image.	Please enter as many keywords as you can think of to describe the overall mood in the respective image.	Score: 0					
u can separate keywords by pressing ENTER or TAB.	You can separate keywords by pressing ENTER or TAB.						
nce you are finished, click Next image.	Once you are finished, click Next Image.						
Add tag	Add tag	9620					
		600					
		1330 0					
		Ninii Vani neo20 snak You					
ample keywords for this image: happy, calm, dreamy	Example keywords for this image: happy, caim, deamy						
Next image	Next image						

Figure 3.3: Screenshot of the page where users could make their choice.

This was implemented in the following way: We included an extra page (depicted in Figure 3.3) which could be accessed between the start of the main task and the end of the tutorial, containing a short explanation of the choice that had to be made as well as screenshots of both task versions to help users make their decision. This process could be completed within a single step, simply by clicking on the selected task version. It is important to note that this page could only be accessed one single time, meaning that once users made their choice, they had to stick with it for the remainder of the study.

3.2.2 Tutorial

We also made substantial changes to the tutorial: On the original platform, Mekler et al. used the same non-gamified tutorial for all users, regardless of which task version they were assigned to. Consequently, even those who completed a gamified version of the main task only saw a tutorial without game elements. Since on our platform, we presented the option to enable or disable gamification right after the tutorial, we felt that it was necessary for us to provide an additional gamified version, so that users could get a first impression of the game elements before having to make their decision. This gamified tutorial used the same game elements as the gamified main task, namely points and a leaderboard. However, points that were achieved during the tutorial did not count towards the final score and were reseted right before the start of the main task. Secondly, we added a guided tour to the tutorial. The guided tour consisted of pop-up windows providing a step-by-step explanation of the tagging process and, if applicable, the game elements (see Figure 3.4). The first window popped up right at the beginning of the tutorial, informing the participants that the displayed image would disappear after five seconds. The second window appeared as soon as the textfield was revealed, giving a throughout explanation on what kind of tags we were looking for and how to proceed. If gamification was enabled for the tutorial, an additional window would pop up once the first tag was entered, showcasing both the score and the leaderboard. Users were explained that for each tag, 100 points would be awarded and notified that their tutorial points would not count towards the actual task.

the overall mood in the respective image. You can separate keywords by pressing ENTER or TAB. Once you are finished, click Next image. Add tag	The image is now gone. In its stead, a textifield appeared, where you can enter as many keywords as you want. The more keywords you come up with, the more you contribute to the study. In the course of this tutorial, we provide you with three example keywords per image, which you can use at will. You can find them right below the textifield. Click <i>Next image</i> to confirm your input and get to the next image.
Example keywords for this image: happy, calm, dreamy	
Next image	Close

Figure 3.4: Screenshot of a pop-up window that was part of the guided tour.

Additionally, we wanted to give further directions concerning the type of tags suitable for affective image classification. Many of the tags collected by Mekler et al. did not meet the set criteria [21], even though the instructions clearly stated that the authors were looking for keywords on the images' overall atmosphere. To make this even clearer, we decided to provide exemplary tags for the tutorial images. We supplied three tags for each of the tutorial images and displayed them right beneath the textfield where the platform users were safe to notice them. Furthermore, we mentioned the exemplary keywords in the guided tour, explaining that users could use them as a reference for their own respective tags.

As a last change to the tutorial, we shortened it by one image. On the platform by Mekler et al. the tutorial spanned three consecutive images. Instead of using all three, however, we decided cut the last tutorial image, effectively reducing the tutorial to two images. This way, users who would be given the choice to enable or disable gamification could complete exactly one gamified and one non-gamified image during the tutorial, familiarizing themselves with both task versions to the same extend.

3.2.3 Bonus rounds

Furthermore, we added the option to complete bonus rounds of the image tagging task. This was suggested in [21] as a way of increasing methodological robustness: In previous studies, free choice measures (like voluntarily tagging more images) yielded results that were comparable to participants' self-reports on their intrinsic motivation [5], so that the bonus rounds should serve as a reasonable indicaton thereof. Namely, the more intrinsically motivated the users, the more bonus rounds they are likely to complete.

Unlike the main task, these bonus rounds consisted of five images each. They were taken from the same set of abstract paintings [18] as the rest of the images used for the platform and thus fit well in terms of style and level of abstraction. The bonus rounds were accessible for the users as soon as the 15 main images were tagged. If users had completed the non-gamified version of the image tagging task, the bonus images were non-gamified as well, and similarly, if users had completed the gamified version, the game elements were kept for the bonus round. Points collected during the bonus round were added on top of the score from the main task; this way, the bonus rounds served as a good opportunity for users who wanted to get to the first position of the leaderboard.

Moreover, users were not limited to a single bonus round. Instead, once a bonus round was completed, they could, in theory, always start a new one; the platform keeping track of and displaying the number of finished rounds. In practice, we implemented a total of 20 bonus rounds, since it was highly unlikely to need more than that (20 bonus rounds amount to 100 extra images). Additionally, we were restricted by the amount of unused images available in the set from [18]. Thus, after the twentieth go, the option to start a new bonus round was simply omitted from the menu.

3.2.4 Questionnaires

We also chose to integrate questionnaires that were relevant for our user study directly into our platform. This way, it was possible for us to administer the questionnaires at fixed points throughout the image tagging process, as compared to Mekler et al. who, in their study, had participants fill out questionnaires only after the tagging task was fully completed. In particular, we implemented the option to access questionnaires between the the two tutorial images, as well as after the tutorial, i.e. right before starting the main tagging task. Additionally, we also provided questionnaires at the end of the tagging process, either directly after the 15 main images, or, in the case that users selected to complete additional bonus rounds, directly after the bonus images.

3.3 Implementation

The platform was implemented in form of a single page application, using *AngularJS*¹version 1.6.10. Furthermore, in order to match the design of Mekler et al.'s platform, our user interface was build with *Bootstrap*².

To store the data that was obtained during the study, we used a $mySQL^3$ database which we connected to via PHP^4 . Upon visiting the platform, participants were registered on this database; additionally, their ID was stored in a cookie and saved on their browser. As the users explored the platform, we continuously stored their current state, i.e. their exact location on the platform, by saving information on their progress to the database. This information entailed, inter alia, which version of the task they were assigned to, if the tutorial had been completed yet, which image they were currently working on, how many and which images they had worked on before, how many tags they had written in total, and which of the questionnaires had been filled out already. Each time the users proceeded to the next page, an updated version of their state was inserted into the database and marked as active, that is marked as the newest, most relevant entry. For those users who were given the choice to enable or disable gamification, their choice was noted as well, and their assigned task version was updated accordingly for all successive states.

With these states in place, the platform could easily provide users with a fitting route for proceeding through the platform. Furthermore, when a user re-visited our page and already had a cookie with an ID set—the cookies expired only after one year—their last registered state could easily be checked; this way, they could pick up exactly were they had left off before leaving the platform. This implementation ensured that the users would not partake in the image tagging task multiple times. It also stopped the users from reloading the page in an attempt to start over, as they were always led back to their last logged location.

Moreover, once an image was displayed during the task, it was flagged as seen by the user on the database. Therefore, even when reloading the page, the image would not be displayed a second time. Instead, the textfield would appear instantaneously, leaving no way for users to view the image for more than five seconds in total.

¹https://angularjs.org (last accessed July 3, 2019)

²https://getbootstrap.com (last accessed July 3, 2019)

³https://mysql.com (last accessed July 3, 2019)

⁴https://php.net (last accessed July 3, 2019)

Chapter 4 User Study

With our online platform in place, we conducted a user study in order to analyze the impact of choice in a gamified setting. More specifically, we utilized the implemented versions of the image tagging task to investigate whether the choice to enable or disable gamification has a positive effect on users' performance.

4.1 Conditions

For the user study, we distinguished between three conditions which were based on the online image tagging platform. Namely, to ensure that we reasoned about the effect of choice specifically, we differentiated between a No Gamification condition, a Gamification condition and a Choice condition:

- No Gamification condition. Participants in the No Gamification condition were given a non-gamified tutorial, in which both tutorial images were displayed without any game elements. Once the tutorial was done, they were assigned to the non-gamified task version without being given a choice on the matter. If they decided to complete any bonus rounds, the bonus images were non-gamified, as well.
- Gamification condition. The Gamification condition was the opposite to the No Gamification condition, in the sense that participants were given two gamified tutorial images and afterwards, had to complete the gamified version of the main tagging task without being offered a choice. In this condition, all bonus images were displayed in the gamified version.
- **Choice condition.** Participants in the choice condition completed a mixed tutorial. This meant that one tutorial image was shown in the gamified ver-

sion, while the other was displayed in the non-gamified, plain version, the ordering being counterbalanced (one half of participants saw the gamified image first, the other half the non-gamified image). After having seen both task versions during the tutorial, participants were allowed to make a choice on whether they wanted to enable or disable the gamification intervention for the main task. At this point, the Choice condition was split into two subconditions, namely Choice_{Gamification} and Choice_{No Gamification}. Participants who decided to enable gamification were assigned to the Choice_{Gamification} condition and similarly, participants who disabled gamification were assigned to the Choice_{No Gamification} condition. Once they were assigned, they remained in their respective condition for the remainder of the tagging process, i.e. for the entirety of the main task, as well as for any bonus rounds they chose to complete. Out of all participants, only those assigned to the Choice condition ever saw both versions of the tagging task; they were the only ones made aware of the fact that the option to enable or disable gamification existed on the platform.

With the participants being randomly assigned to either Gamification, No Gamification or Choice, we were able to investigate the effect of having a choice by comparing the overall performance, or more specifically, the quantity and quality of provided tags, across conditions.

4.2 Hypotheses

Building on the formerly presented related work, we derived the following hypotheses:

- **H1.** Participants who complete the gamified task generate more tags than those who complete the non-gamified task, without it resulting in a decrease of quality, regardless of whether they are given a choice.
- H2. Participants who have a choice and use it to enable gamification generate more tags than participants who do the same gamified task without having a choice, without it resulting in a decrease of quality.
- H3 Participants who have a choice and use it to disable gamification generate more tags than participants who do the same non-gamified task without having a choice, without it resulting in a decrease of quality.

H1 is based on the results obtained by Mekler et al. [21], who found that the gamification intervention of points and a leaderboard had a positive impact on users' motivation in the context of image tagging, resulting in an increased amount of generated tags. Regardless of this, the quality of tags did not suffer in any of their gamified conditions. As our platform and context were heavily inspired by theirs, we aimed to replicate their findings and, as a consequence,

expected more tags to be provided in the Gamification condition than in the No Gamification condition, without a loss of tag quality. Similarly, we expected participants in the Choice_{Gamification} to generate more tags than participants in the Choice_{No Gamification} condition, without it affecting the quality of generated tags negatively.

H2 and H3 follow from the previously presented related works, which not only show that having a choice is generally perceived well [16], but also that peoples' behavior can be positively affected when being offered a choice [2, 35]. As a consequence, we expected participants in the Choice_{Gamification} condition to provide more tags of equally good quality compared to participants in the Gamification condition, who, while completing the same task with the same gamification intervention, were not given a choice. Likewise, we expected participants in the Choice_{No Gamification} condition to generate more, equally qualitative tags compared to participants in the No Gamification condition, even though they complete the same non-gamified task.

4.3 Method

A link to the platform was shared via German social media, *Facebook*⁵ in particular. By clicking the link, people were passed on to the platform's webpage, where a small introductory text was displayed to provide a suitable framing for the image tagging task. Instead of disclosing information on the purpose of the user study or even the study in general, we claimed that the image tagging task solely served the purpose of advancing in the field of affective image classification. We explained that a participation would benefit us scientifically, as it would help us explore the moods and feelings that the images evoked within people. The same framing was used by Mekler et al. [21] and was perceived positively by their participants, so that we decided to adopt it for our study, as well.

After people confirmed that they wanted to participate and that they had read and understood our privacy policy, they were assigned to either the No Gamification condition, the Gamification condition or the Choice condition. We made sure to maintain a 1:1:2 distribution, meaning that compared to the other two conditions, double the amount of participants were assigned to the Choice condition. This was due to the fact that, as mentioned above, the Choice condition was to be split into two sub-conditions later on, namely Choice_{Gamification} for participants who enabled gamification and Choice_{No Gamification} for those who disabled gamification. Thus, by allotting more participants to the Choice condition, we hoped to achieve an overall even distribution of participants between No Gamification, Gamification, Choice_{No Gamification} and Choice_{Gamification}. Once par-

⁵https://facebook.com (last accessed July 3, 2019)

ticipants were assigned to a condition, they got started with their corresponding tutorial—those in the No Gamification condition with the non-gamified tutorial, those in the Gamification condition with the gamified tutorial and those in the Choice condition with the mixed tutorial. Depending on participants' condition, the tutorial was even further adjusted, the Choice condition being particularly affected by this: Since participants in the Choice condition completed a mixed tutorial, i.e. one gamified and one non-gamified tutorial image, we took special care to avoid any priming effects. To this end, the Choice condition was split in half; for one half of participants, the gamification intervention of points and a leaderboard was enabled for the first tutorial image and disabled for the second, while for the other half of participants, this order was reversed.

Furthermore, questionnaires were administered after each of the tutorial images: After the non-gamified tutorial image, participants in the Choice condition were asked to answer three questions, which related to the enjoyment subscale of the validated German short-scale of the Intrinsic Motivation Inventory (IMI) [34], on a 5-point Likert scale (ranging from 1='strongly disagree' to 5='strongly agree'). The questions for the IMI were formulated in such a way that not the gamification intervention itself, but rather the image tagging task in general was paramount (e.g.: "Image tagging was fun for me.", "I found the image tagging interesting").

After the gamified tutorial image, the same three items of the IMI were administered, followed by six additional items, which aimed to assess participants' enjoyment of the game elements specifically. These items were worded as follows: "Receiving points for image-tagging has motivated me to provide more tags", "I find points motivating in general", "I like the game element: 'Receiving points for actions"', "To be able to compare myself to others on a leaderboard has motivated me to provide more tags", "I find the comparison with others on a leaderboard motivating in general", "I like the game element: 'Comparison with others on a leaderboard"'. They, too, were rated on a 5-point scale, ranging from 1='strongly disagree' to 5='strongly agree'. Note that these questionnaires were translated only in the context of this thesis, as the online platform was German. The original, German version of each questionnaire can be found in appendix A.

After both tutorial images and the corresponding questionnaires were completed, participants were passed on to the next page, where they where allowed to choose which version of the image tagging—gamified or non-gamified—they wanted to keep for the main task. On the page, we provided a short informative text, in which we explained this choice to participants. We were careful to keep the explanation neutral to fit our framing; we did not use terms like 'game', 'gamification' or 'game elements', but rather talked about 'task versions' in general. Hence, our wording was along the lines of "Now you can decide which version of the task you would like to use for the study". Besides the informative text, screenshots of both task versions were displayed on the page in order to refresh participants' memories. The task version they saw last was displayed on the left side of the screen, with the button underneath reading "keep second version"; the screenshot depicting the other, first task version was shown on the

right, its button reading "use first version" (see Figure 4.1). After participants made their choice, they were assigned to the $Choice_{Gamification}$ condition or the $Choice_{No Gamification}$ condition respectively, and started working on their selected main task. Apart from this, no further changes were made to the image tagging task, so that they proceeded as normal from here on.



Figure 4.1: The screenshot on the left shows the page for participants who completed the the gamified tutorial first; the screenshot on the right shows the page for participants who, vice versa, completed the non-gamified tutorial first. While the screenshots' positions are swapped, the rest of the texts and buttons remain the same in both screens.

For participants in Gamification and No Gamification, the tutorial was also adjusted in terms of questionnaires: In the Gamification condition, we administered a questionnaire right at the end of the tutorial, i.e. after both tutorial images had been tagged. It consisted of the three items of the enjoyment scale of the validated short-scale IMI, as well as the six items to asses participants' perception of the game elements, which are listed above. After finishing the tutorial, participants were assigned to the gamified version of the main task without having a choice. This was similar for participants in the No Gamification condition, who, after tagging both images, were asked to fill out the three items relating to the enjoyment scale of the IMI. Afterwards, they were mandatorily assigned to the non-gamified version of the main task.

The main task then consisted of the same 15 images for all participants, which were shown in a randomized order. As mentioned in the previous chapter, these images corresponded with those used on Mekler et al.'s platform [20, 21]. After the main task, participants in all conditions also had the option to complete a bonus round, consisting of five images, or to proceed without tagging any further images (Figure 4.2). If they chose to do the bonus round, it was presented in the same version as the main task, i.e. non-gamified for participants in the No Gamification and in the Choice_{No Gamification} condition and gamified for participants in Gamification and Choice_{Gamification}. Once completed, they could always decide to start another bonus round or, as before, proceed without tagging any more images instead.

If participants decided to proceed on the platform, they were administered another set of questionnaires. Regardless of their condition, they were requested to fill out the Gamification User Types Hexad Scale⁶ [30] in order to get information on their player type. Afterwards, participants were asked to provide answers to the complete German short-scale IMI, the subscales assessing their enjoyment, pressure, autonomy and competence. For participants in the No Gamification condition, this concluded the study. Participants in the other conditions, however, were given additional questions in relation to 'having a choice'. More specifically, participants in the Gamification condition were asked to rate the statement "I would have liked it if I could have decided whether I wanted to use game elements or not" on a 5-point scale (1='strongly disagree', 5='strongly agree'), while participants in the Choice condition rated the statement "I liked that I was able to decide whether I want to use game elements or not" on the same scale. Lastly, participants in both conditions were asked to rate the statement "I would have liked to have more choices". Here, they were also provided an additional text field, where they could explain in detail which choices they would have liked to have. Once both statements were rated, their participation was completed.

Done!

Thank you for participating in the image tagging task. We would like to ask you some **concluding questions**. Click *Survey*, to get to the questions.

If you want to help us even more, you can tag **five additional images**. These are new images you have not seen yet. The bonus round has the same form as the previous tagging and can be repeated as many times as you want before you continue on to the survey. Click *Bonus round*.

Please note: you have to complete the survey before before we can assess your contribution to the study.



Figure 4.2: Screenshot of the page on which participants could decide to either complete a bonus round or to continue on to the concluding questionnaire. After each completed bonus round, they were brought back to this page and could decide anew.

⁶German version provided by Marc Busch who is one of the authors.

4.4 Measures and Coding Process

With the assigned condition as the independent variable, we mainly focused on the amount of generated tags as a dependent variable in order to argue about participants' performance. However, we only considered tags that were generated during the main task, meaning that tags from the tutorial and the bonus rounds did not count towards the study. This was, for one, due to the fact that we wanted to keep our results comparable to those of Mekler et al. [21]; another reason was that, while not all participants completed the same amount of bonus runs or even any bonus runs at all, the main task was mandatory for everyone. As a consequence, taking only tags from the main task into consideration made the most sense for accurately comparing participants' performance.

As a secondary variable, we also considered the quality of generated tags. To this end, we let all tags be coded by two separate coders based on a set of rules that were established beforehand and refined throughout the coding process (see appendix B). The coders rated all generated tags on a 3-point scale, the value 1 meaning that the tag neither suited the image nor related to a mood, value 2 meaning that the tag either fit the image but did not relate to a mood, or, the other way around, that the tag related to a mood but did not fit the image very well, and value 3 meaning that the tag related to both the image and a mood. We calculated the inter-rater agreement as $\kappa = .86$, which is considered almost perfect [28], so that these ratings were overall reliable.

4.5 Participants

90 participants completed the study. i.e. they at least tagged the 15 main images and filled out all questionnaires and forms. To ensure that only valid responses were taken into consideration, we excluded all participants for which the standard deviation in either the Hexad or the IMI questionnaire was less or equal to 0.5, in which case we concluded that they must have simply clicked through the questionnaire without truthfully answering the questions. Additionally, we removed extreme outliers regarding the amount of generated tags: We calculated the mean and standard deviation of the number of tags generated in the main task for each condition, and excluded all participants who deviated from this mean by more than 2.5 times the respective standard variation. This process was repeated until no more outliers could be found. In the end, we removed 13 participants—6 of them due to their responses to the questionnaires and 7 because they were outliers in terms of generated tags—and were left with a total of 77 valid responses. Out of those, 47 participants were female, 26 male, 2 did not specify a gender and 2 identified as other. Furthermore, 41 participants were aged 18–24 years, 19 were aged 25–31 years, 5 were aged 32–38 years, 1 was aged 39–45 years, 5 were aged 46–52, and finally, 6 participants were aged 52 years or older. Figure 4.3 shows their distribution across the individual conditions.



Figure 4.3: 16 of the 77 participants were assigned to the No Gamification condition, 19 were assigned to the Gamification condition and 42 to the Choice condition. Out of those, 19 participants chose to enable gamification, while 23 chose to disable it.

4.6 Results

In the next sections, we discuss the data that was obtained in the user study. In line with our hypotheses, we first and foremost focused on the quantity and quality of generated tags, presenting the corresponding findings as our main results. Subsequently, participants' responses to the questionnaires were analyzed as well; they are presented as supporting results.

4.6.1 Main Results

Taking the the amount and the quality of generated tags from the main task into consideration, the following results were found:

MR1: The game elements motivated participants to generate more tags

We first analyzed the number of generated tags independently of their quality (see Table 4.1). To compare the mean amount of tags across conditions, we calculated a one-way ANOVA, for which we found a significant difference (Welch's F(3, 39.702)=13.535, p<.001). We report the Welch's F here, as Levene's test for homogeneity of variance was almost significant for our dataset (p=.05) and Welch's F is more robust in this case [10]. For the same reason, we used the Games-Howell post-hoc procedure to further compare the individual conditions. We found that participants in the Gamification condition generated significantly more tags than participants in the No Gamification condition (M=62 vs M=37.7, p<.001) and similarly, that participants in Choice_{Gamification} generated significantly more tags than participants in Choice_{No Gamification} (M=70.9 vs M=51.2, p<.05). This matches the results reported in [21], where participants who used the gamification intervention of points and a leaderboard also generated more tags than participants in the non-gamified condition.

MR2: Having a choice motivated participants who disabled gamification to generate more tags

The Games-Howell post-hoc test further revealed a significant difference for the amount of generated tags in the $Choice_{No \ Gamification}$ and No Gamification conditions (M=51.2 vs M=37.7, p<.05; M being the mean amount of generated tags), meaning that participants who were offered a choice and used it to disable gamification provided significantly more tags than participants who did the same non-gamified task without being given a choice. Surprisingly, however, we did not find a significant effect between $Choice_{Gamification}$ and Gamification (M=70.9 vs M=62, p=.585), meaning that participants who chose to enable gamification did not provide significantly more tags than participants who completed the same gamified task but without being offered a choice.

		Number of generated tags					
Condition	n	Μ	SD	Mdn	Min	Max	Sum
No Gamification	16	37.7	11.1	36	18	58	603
Gamification	19	62	17.9	63	33	90	1178
Choice _{Gamification}	19	70.9	24.8	70	18	121	1348
Choice _{No Gamification}	23	51.2	18.5	51	13	84	1178

Table 4.1: Detailed overview of the amount of tags generated in the main task across conditions (n=number of participants in the condition, M=mean amount of generated tags, SD=standard deviation, Mdn=median, Min=minimum and Max=maximum).

MR3: Neither gamification nor choice motivated participants to complete more bonus rounds

At this point, we also compared the mean number of completed bonus rounds across conditions (see Table 4.2). We calculated a second one-way ANOVA with the mean amount of bonus rounds as the dependent variable and found a significant difference (Welch's F(3, 34.220)=3,244, p<.05). The homogeneity of variance was violated (Levene's test being significant with p<.01), which is why we again report the more robust Welch's F. We compared the Gamification condition to the No Gamification condition (M=0.58 vs M=0.25, p=.407) and the Choice_{Gamification} condition to the Choice_{No Gamification} condition (M=0.33 vs M=0.09, p=.308), using the Games-Howell post-hoc procedure. Neither of these tests revealed a significant difference, i.e. participants in the gamified conditions did not complete significantly more bonus rounds than participants in the non-gamified conditions. The Games-Howell post-hoc procedure further revealed no significant effect between the Choice_{Gamification} and Gamification conditions (M=0.33 vs M=0.58, p=1), as well as between Choice_{No Gamification} and No Gamification (M=0.09 vs M=0.25, p=.407) vs M=0.25, p=.407) and the choice_{No Gamification} conditions.

p=.581), so that having a choice did not motivate participants to do more bonus rounds either. We believe the reason for this might be that even when using gamification or being offered a choice, the task itself was not particularly enjoyable for participants, keeping them from wanting to tag any additional images.

		Completed bonus rounds						
Condition	n	Μ	SD	Mdn	Min	Max		
No Gamification	16	0.25	0.45	0	0	1		
Gamification	19	0.58	0.77	0	0	3		
Choice _{Gamification}	19	0.58	1.17	0	0	5		
Choice _{No Gamification}	23	0.09	0.29	0	0	1		

Table 4.2: The amount of completed bonus rounds in the individual conditions.

MR4: Neither gamification nor choice affected the quality of generated tags

We also took the quality of participants' performance into account by comparing the tag quality across conditions. The mean quality was generally rather high (No Gamification: M=2.6, Gamification: M=2.4, Choice_{Gamification}: M=2.5, Choice_{No Gamification}: M=2.5) and only marginally deviated across conditions. Accordingly, no significant difference between the conditions could be found (Welch's F(3, 39.53)=2.438, p=.079). This fits in with Mekler et al.'s [21] results, who, in their study, reported neither an increase nor a decrease of tag quality for the gamified conditions when compared to their non-gamified baseline condition, so that the tag quality seems to be unaffected by both gamification and choice.

4.6.2 Supporting Results

Taking participants' responses to the individual questionnaires into consideration, we report the following supporting results:

SR1: Not all participants in the Choice condition enjoyed the game elements to the same extent and chose accordingly

In order to determine possible factors that might have influenced participants' decision to disable or enable gamification, we analyzed the data collected in the Choice conditions more closely. Firstly, we checked for ordering bias, i.e. we investigated whether the ordering of the two tutorial images had an impact on participants' choice. Out of the 23 participants in the Choice condition who decided to disable gamification, eleven saw the non-gamified tutorial image first and twelve saw it second; similarly, nine out of the 19 participants in the Choice condition who chose to enable gamification saw the gamified tutorial image first,

ten saw it second. Overall, about 50% of participants selected whichever task version they saw first during the tutorial, while the other 50% selected whichever task version they saw second. Since this is a relatively even contribution, we rule out any possibility of ordering bias.

We then took a closer look at the results obtained in the tutorial questionnaires. In particular, we derived participants' answers to the six-item questionnaire measuring their perception of the individual game elements (Figure 4.4).



Figure 4.4: Mean ratings to the 6 statements measuring participants' perception of the game elements (all rated on a 5-point scale).

As described in the first section of this chapter, all six questionnaire statements were rated on a 5-point Likert scale. For each statement, we calculated an independent t-test to compare the respective mean score between the Choice_{Gamification} and Choice_{No Gamification} condition and found that the difference was significant every time (p<.05). This means that participants in Choice_{Gamification} rated the game elements as significantly more enjoyable than participants in Choice_{No Gamification}, hinting that when choosing to enable or disable gamification, they made a conscious decision based on their perceived enjoyment of the game elements.

This assessment is further supported by the results for the enjoyment subscale of the IMI from the tutorial: As shown in Table 4.3, these ratings were mostly consistent across conditions, and, overall, relatively neutral, indicating that participants in all conditions perceived the image tagging task as neither particularly enjoyable, nor particularly unenjoyable.

Since in the Choice condition, the subscale had to be filled out for both task versions, we checked whether the ratings changed between the gamified and the non-gamified tutorial image for participants in Choice_{Gamification}, as well as

Condition	Enjoyment subscale without gamification M SD Mdn			Enjoyment subscale with gamification		
Condition	141	30	wiun		30	wiun
No Gamification	9.8	2.9	10	-	_	_
Gamification	_	_	_	10.3	2.3	11
Choice _{No Gamification}	9.6	2.9	9	9.5	2.7	9
Choice _{Gamification}	10.4	3.2	11	9.2	3	10

Table 4.3: The table shows the ratings for the enjoyment subscale of the IMI after both the non-gamified and the gamified tutorial. The scores for all three 5-point scale statements are summed up for each participant [34], resulting in values ranging from 3–15 (9 being the neutral score).

for participants in Choice_{No Gamification}. We calculated paired t-tests to compare the reported ratings for both task versions condition-wise. For participants later in Choice_{Gamification}, we did not find a significant difference (M=9.6 vs M=9.5, t(18)=.195, p=.848), i.e. their enjoyment of the task did not change depending on whether gamification was enabled or disabled. Participants later in Choice_{No Gamification}, however, enjoyed the image tagging significantly more when gamification was disabled (M=10.4 vs M=9.2, t(22)=2.859, p<.05).

Taking the results of both tutorial questionnaires together, we conclude that especially those participants who did not enjoy using the game elements tended to disable the gamification when given the choice.

SR2: Participants' player type was not a predictor for their choice

Using the data obtained in the Choice condition, we also calculated a binary logistic regression to analyze whether participants' player type, which was derived from their answers to the HEXAD questionnaire, was a predictor for their choice—in other words, whether certain player types were more likely to enable or disable gamification than others. This regression analysis, however, did not reach statistical significance for any of the player types (p>.20), meaning that in our case, they did not seem to affect participants' decision.

SR3: Participants in the Choice condition appreciated the option to enable or disable gamification

We evaluated the answers to the concluding questionnaire in the Choice conditions, where participants gave their opinion on whether they liked being offered the choice to enable or disable gamification (see Table 4.5). Here, participants in both Choice_{Gamification}, as well as Choice_{No Gamification} gave high ratings (M=4.26), which clearly show that they appreciated having this option. Surprisingly, participants in the Gamification condition, who, in their concluding questionnaire, were asked whether they would have liked being offered a choice, indicated that they would not have ben especially interested (M=2.53). We believe this might be due to the fact that they never saw the non-gamified version of the tagging task during their run and were therefore unable to fully understand the value of having such a choice. This assumption would also fit the results for the second question of the concluding questionnaire, which was the same for both the Choice and the Gamification condition: Participants were asked whether they would have liked to have even more choices besides enabling or disabling the gamification. They all tended to disagree with this statement (Choice_{No Gamification}: M=2.32, Choice_{Gamification}: M=2.26, Gamification: M=2.47). The free texts submitted at the end of the concluding questionnaire did not provide any conclusive results on the matter and are not further discussed for readability reasons.

SR4: Participants in the gamified conditions experienced higher levels of competence and participants in the Choice conditions experienced higher levels of autonomy

Lastly, we analyzed the responses to the IMI, which was filled out by all participants after completing the main task. The resulting scores are depicted in Figure 4.5, where we can clearly see that those participants who completed the gamified version of the image tagging task gave higher ratings for the competence subscale, while participants in both Choice conditions provided higher ratings for autonomy.



Figure 4.5: Mean ratings for the individual subscales of the IMI. For all subscales, the scores of the three related 5-point scale statements are summed up for each participant [34], i.e. the mean values range from 3–15).

We calculated a one-way ANOVA for each of the subscales. The homogeneity of variance was not violated for any of them (Levene's test not being significant with p>.3) and, as expected, no significant difference was found for the enjoyment subscale (F(3,73)=.597, p=.619) or the pressure subscale (F(3,73)=.391, p=.76). For both competence (F(3,73)=6.505, p<.05) and autonomy (F(3,73)=4.365, p<.01), however, the ANOVAs were significant, so that we additionally used the Games-Howell post-hoc procedure for pairwise comparisons of the mean scores.

The scores for the competence subscale were significantly different between No Gamification and Gamification (M=6.7 vs M=8.8, p<.05), between No Gamification and Choice_{Gamification} (M=6.7 vs M=9.4, p<.05), as well as between Choice_{Gamification} and Choice_{No Gamification} (M=9.4 vs M=7, p<.05). Even though the difference between Choice_{No Gamification} and Gamification condition was not significant (M=7 vs M=8.8, p=.088), it seems as if the gamification intervention overall increased participants' feeling of competence. One reason for this might be the nature of the leaderboard, which was relatively easy to climb due to the other players and scores being fictive. It is also thinkable that participants gained a sense of achievement by being awarded points for each generated tag. Since the feeling of competence is considered one of the basic needs for self determined behavior and, as such, should have a positive impact on intrinsic motivation [6], we assume that participants in the gamified conditions were more intrinsically motivated than participants in the non-gamified conditions, which is one possible explanation for why they provided more tags during the task.

For the autonomy subscale, we found significant differences between the No Gamification condition and both Choice conditions respectively (for Choice_{Gamification}: M=8.25 vs M=11.8, p<.05; for Choice_{No Gamification}: M=8.25 vs M=11.1, p<.05), hinting that participants in these conditions recognized the option to enable or disable gamification as an actual choice and felt more autonomous because of it. Furthermore, just like the feeling of competence, autonomy poses an important basis for intrinsic motivation [27]. Thus, we reason that participants who were given a choice experienced a higher level of intrinsic motivation compared to participants in the No Gamification condition. However, we did not find any significant differences between the Gamification condition and the Choice_{No Gamification} condition (M=9.7 vs M=11.1, p=.496), nor between the Gamification condition and the Choice_{Gamification} condition (M=9.7 vs M=11.8, p=.196). One reason for this might be that participants in the Gamification condition felt more autonomous to begin with: We believe that participants in the Gamification condition might have perceived the task as more of a competition (due to the competitive game elements), for which they could decide themselves how much effort they wanted to put into it, resulting in an increased feeling of autonomy.

4.7 Discussion

First of all, we were able to replicate Mekler et al.'s [21] findings. The gamification intervention of points and a leaderboard positively affected participants' performance, in the sense that they provided more tags than participants who did the same image tagging task, but without the game elements (**MR1**). Furthermore, same as in Mekler et al.'s work, the significantly increased amount of tags in the gamified conditions did have a negative effect on the overall tag quality (**MR4**).

Moreover, we found the offered choice to be beneficial (**M2** and **M3**): When comparing the two non-gamified conditions, we found that the amount of generated tags increased significantly when participants were given choice—again, without the tags' quality being affected in any way. Thus, having the choice to enable or disable gamification had a positive effect on those who decided not to use it. In the gamified conditions, on the other hand, there was no significant difference for neither the amount of generated tags nor the tag quality, meaning that participants who enabled gamification were, in the very least, not negatively impacted by having a choice. What is more, the answers to the administered questionnaires clearly showed that participants in the Choice condition generally appreciated being allowed to enable or disable gamification and, overall, chose deliberately (**SR1, SR3**). Their decision was not affected by the presentation sequence of task versions in the tutorial, but rather reflected their actual perception of the game elements—participants who enjoyed them enabled them, while participants who did not enjoy them disabled them.

While we believe the concept of choice to be useful and profitable in this regard, there are also some aspects which we view critically: Firstly, as mentioned above, having a choice did not seem to affect participants who decided to enable gamification since they did not provide significantly more tags than participants who did the gamified task without being offered a choice and, while it can be argued that their performance was not negatively impacted either, we find the lack of positive effects concerning. One possible explanation can be found in the rating to the enjoyment subscale of the IMI, which were, to the most part, relatively low, meaning that the task was not overly enjoyable for participants. We believe that maybe those participants, who used the game elements and were motivated by them, could not be further motivated by additional incentives (like having a choice), as the task itself remained uninteresting nevertheless. In other words, we suspect choice to have a ceiling effect. In order to investigate this assumption more closely, a follow-up study would be necessary, which we will discuss in more detail in the next chapter.

A second concern is that participants in the Choice_{No Gamification} condition, who, as we know now, were positively affected by having a choice, still provided less tags overall than participants in the Gamification condition. Even though this difference was not statistically significant, the question arises as to whether the approach of not offering choices would ultimately pose the more feasible

solution. To our understanding however, this is not the case for various reasons: For one, the omission of choice would again result in a one-size-fits-all approach, which previous studies [25, 30] have already shown to be problematic (see the related work chapter). This is only reinforced by the fact that we were able to observe first-handedly how much the perception of the gamification setup differed across participants, as their responses to our questionnaire regarding the enjoyment of the game elements varied immensely. Even further than that, we know that participants, who disabled gamification, did so consciously and deliberately, because they did not appreciate using the individual elements (SR1). Forcing them to use them anyway does not seem advisable, at least from a user experience point of view. What is more, based on their enjoyment ratings, some participants in the Gamification condition as well, might have preferred not to use gamification, and would have disabled it if given the choice. Unfortunately, since we could only make assumptions based on their answers to our questionnaire, we could not identify these participants accurately. Unlike in the Choice condition, cleanly splitting them into two subconditions based on whether they would have or would not have disabled the game elements was therefore not an option. As a consequence, we were unable to analyze and compare their performance in more depth. However, the question arises as to how the participants in the Gamification condition, who disliked the gamification, would have performed if they had been given a choice: Potentially, if these participants had been assigned to the Choice condition and as a result would have been able to choose not to use the game elements, they might have performed better and generated even more tags than they did in the Gamification condition, so that by implication, we maybe would have achieved better results overall. Additionally, we know from the responses to the IMI that participants who were given a choice felt significantly more autonomous than participants in either the Gamification or the No Gamification condition, which is why we assume that they were most likely more intrinsically motivated (SR4). Taken together, the aforementioned aspects strongly suggest that providing participants with the option to disable the gamification intervention seems far more reasonable than forcing them to keep using a possibly non-optimal one-size-fits-all approach.

Overall, we arrive at the conclusion that the choice to enable or disable gamification benefitted participants in the image tagging task, having a positive effect on those who disliked the game elements, while not negatively affecting participants who enjoyed the game elements.

Chapter 5 Conclusion

In this chapter, we give a conclusive overview of the user study and discuss the significance of our results. Furthermore, we go over the limitations of our approach and provide directions for future work.

5.1 Overview

In this thesis, we explored a very fundamental form of customization for gamified systems, namely enabling or disabling gamification as a whole. Since in previous research, having a choice was generally shown to be beneficial [11], we hypothesized that offering users the choice to activate or to deactivate a fixed configuration of game elements should have a positive effect, as well. To investigate this further, we implemented an online platform based on existing literature, where gamification was shown to be successful, and provided two versions—one gamified and the other non-gamified—of the same affective image tagging task. Similar to Mekler et al. [20, 21], we displayed a leaderboard and awarded points for each generated tag as the gamified task version; for the non-gamified version, we only used the plain task without any additional elements. This platform served as the basis for our user study: Some participants were asked to complete the non-gamified task, some were asked to complete the gamified task, and others were allowed to choose for themselves after having seen both task versions in a tutorial, meaning that they decided whether they wanted to enable or disable gamification. We compared participants' performance across conditions, using the quantity and quality of generated tags as measures thereof.

5.2 Contributions

First of all, we replicated the core findings of Mekler et al.'s study [21]. In particular, we found that participants who were assigned to the gamified task generated significantly more tags than participants who were given the nongamified task and equally, participants who chose to use game elements provided significantly more tags than participants who chose not to use game elements. Thus, we were able to reaffirm the positive effect of points and a leaderboard on users' performance, which in itself is an important contribution.

Secondly, we confirmed that having a choice is, overall, beneficial. We found that participants who disliked the game elements and therefore disabled them provided significantly more tags than participants who were assigned to the non-gamified task version without being offered a choice. Furthermore, participants who enjoyed the gamification intervention and therefore enabled it, though not positively affected, did not experience any drawbacks given our measurements and depend variables. Consequently, it seems that having a choice positively affects all those users who dislike the provided, fixed game elements, while not having a negative impact on those who enjoy the gamification anyway.

Yet another contribution of this thesis is the customization concept in itself: The customization is completed within only a single step, as it only takes the single click of a button to select the preferred task version. Especially in comparison to more sophisticated approaches with a high degree of customization (e.g. [16, 17]), this is significantly less work for users, completely eliminating the problem of "choice overload" [12].

All in all, our research indicates that offering users the choice to disable gamification poses a feasible, easy-to-realize customization option. In conclusion, we contribute an approach that should be commonly applied to gamified systems.

5.3 Limitations

First and foremost, our research was limited by the small sample size, which resulted in a relatively low amount of participants in each condition. This is mostly due to the fact that we recruited the participants via social networks and did not pay them money to complete the image tagging task. While we had the possibility to conduct the study on a paid platform, like e.g. Amazon Mechanical Turk, we chose not to do so, however, because we felt that the offered money would have acted as an additional, unwanted incentive that would have to be taken into account.

The second limitation follows directly from the setup of our study or, more specifically, the design of our conditions: Since participants in the Choice condition could choose freely whether they wanted to enable or disable gamification, we had no way of regulating the amount of participants in Choice_{Gamification} and Choice_{No Gamification} to ensure an even distribution. In the end, the group sizes were relatively balanced, but nevertheless, this should be kept in mind for potential re-runs of the study (see the future work section). In relation to the design of the individual conditions, another limitation emerges. Namely, participants in the Gamification and No Gamification condition only ever saw their own version of the tagging task and thus had no way of knowing what they missed out on. This becomes especially problematic when looking at the results for the concluding questionnaire in the Gamification condition, where participants indicated that they would not have appreciated the choice to enable or disable the game elements. As stated above, we believe that they responded this way because it was difficult for them to fully understand their options. While this assumption seems reasonable, we were unable to gain any conclusive insight into the matter.

Furthermore, even though we had a good reason for using only points and a leaderboard (i.e. replicating Mekler et al.'s platform), we feel like this setup was somewhat limiting due to its simplicity. It would have been interesting to use a wider range of game elements and have participants experience a more sophisticated gamification intervention. Similarly, the image tagging task poses another limitation: We know that participants did not find the tagging task enjoyable and presume that generating accurate tags was rather difficult for them, due to the images' high level of abstraction. As a consequence, we believe that it is possible for participants' performance and motivation to have suffered under the nature and design of the task.

5.4 Future Work

As one direction for future work, we believe that investigating the potential ceiling effect of choice in a gamified setting seems especially interesting. In the discussion section of the previous chapter, we noticed that participants who enabled gamification were not further motivated by having a choice, because they did not find the task enjoyable and were already positively affected by the gamification intervention itself. In this case, being offered a choice could neither improve their perception of the task, nor increase their performance, so that we hypothesized about a possible ceiling effect. To investigate this matter more closely, we primarily propose to consider different gamified tasks that are either particularly enjoyable, or particularly unenjoyable, and to then analyze whether providing a choice on top of the gamification intervention has any effect on participants who already enjoy the applied game elements. Furthermore, it would be interesting to see how being offered more than one choice in a gamified task would affect participants: If our assumption is correct, and having a choice does not improve the situation for participants who like the gamification in the first place, they should not be affected by any additional choices either. Alternative

approaches for future work follow directly from the limitations discussed in the previous section. For example, one could do a re-run of the study on a paid platform in order to recruit more participants who complete the image tagging task. This way, not only would it be possible to investigate whether the results are stable for larger sample sizes, but also, whether the participants being paid for their participation has any impact on the overall findings. Another option would be to apply a gamification setup which offers a wider range of game elements, as it would be interesting to see whether the higher level of sophistication has an impact on people's perception of the system and having a choice.

For future re-runs, we also suggest to add two more conditions to the study, namely one condition where participants see both task versions during the tutorial but are forced to complete the non-gamified main task, and another condition where participants also see both task versions during the tutorial but are forced to complete the gamified main task. With these conditions in place, it is possible to examine whether participants in the top-down conditions would be more interested in having choice if they were fully aware of their options and had seen the task both with and without game elements.

Lastly, we propose a further investigation on which aspects in particular, besides people's perception and enjoyment of the game elements, have a moderating effect on their decision to enable or disable gamification.

Appendix A Questionnaires

Below, we give an overview of the questionnaires used in the Study. For each of them, we provide the original, German version of the scale, question and items, followed by an English translation in parentheses.

A.1 Questionnaire on the Perception of the Game Elements

Each item was rated on a 5-point scale, labeled with 1=trifft nicht zu, 2=trifft eher nicht zu, 3=trifft teils-teils zu, 4=trifft eher zu, 5=trifft zu (1=strongly disagree, 2=disagree, 3=neither, 4=agree, 5=strongly agree). See Figure A.1 for a screenshot of how the questionnaire was presented on the platform.

Inwieweit treffen die folgenden Aussagen auf dich zu? (To what extent do you agree with the following statements?)

- 1. Punkte für das Bilder-Taggen zu erhalten hat mich motiviert mehr Tags zu erstellen. (Receiving points for image-tagging has motivated me to provide more tags.)
- 2. Ich finde Punkte allgemein motivierend. (I find points motivating in general.)
- 3. Ich mag das Spielelement: "Punkte für Aktionen zu erhalten". (I like the game element: "Receiving points for actions".)
- 4. Mich mit anderen auf einer Bestenliste vergleichen zu können hat mich motiviert mehr Tags zu stellen. (To be able to compare myself to others on a leaderboard has motivated me to provide more tags.)

- 5. Ich finde den Vergleich mit anderen auf einer Bestenliste allgemein motivierend. (I find the comparison with others on a leaderboard motivating in general.)
- 6. Ich mag das Spielelement: "Vergleich mit anderen auf einer Bestenliste". (I like the game element: "Comparison with others on a leaderboard".)

*Inwieweit treffen die folgenden Aussagen auf dich zu?					
	trifft nicht zu	trifft eher nicht zu	trifft teils-teils zu	trifft eher zu	trifft zu
Punkte für das Bilder-Taggen zu erhalten hat mich motiviert mehr Tags zu erstellen.	0	٢	٥	0	۲
Ich finde Punkte allgemein motivierend.	٢	0	٢	0	٥
Ich mag das Spielelement: "Punkte für Aktionen zu erhalten".	0	0	٥	0	0
Mich mit anderen auf einer Bestenliste vergleichen zu können hat mich motiviert mehr Tags zu stellen.	٢	۲		۲	۲
Ich finde den Vergleich mit anderen auf einer Bestenliste allgemein motivierend.	0	0	0	0	٢
Ich mag das Spielelement: "Vergleich mit anderen auf einer Bestenliste".	۲	٢	٢	۲	٢

Figure A.1: Screenshot of the questionnaire assessing participants perception of the game elements.

A.2 Intrinsic Motivation Inventory

The German, validated short-scale of the IMI was taken from [34]. Participants were asked to rate each item on a 5-point scale, originally labeled with 1=stimmt gar nicht, 2=stimmt wenig, 3=stimmt teils-teils, 4=stimmt ziemlich, 5=stimmt völlig (1=not at all true, 2=not true, 3=partly true, 4=true, 5=absolutely true).

Inwieweit treffen die folgenden Aussagen auf dich zu? (To what extent do you agree with the following statements?)

- 1. Das Bilder-Taggen hat mir Spaß gemacht. (Image tagging was fun for me.)
- 2. Ich fand das Bilder-Taggen sehr interessant. (I found the image tagging interesting.)
- 3. Das Bilder-Taggen war unterhaltsam. (Image tagging was enjoyable.)

- 4. Mit meiner Leistung beim Bilder-Taggen bin ich zufrieden. (I am satisfied with my image tagging performance.)
- 5. Beim Bilder-Taggen stellte ich mich geschickt an. (I demonstrated my skills in image tagging.)
- 6. Ich glaube, ich war beim Bilder-Taggen ziemlich gut. (I think I was quite good at image tagging.)
- 7. Ich konnte das Bilder-Taggen selbst steuern. (I was able to do the image tagging by myself.)
- 8. Beim Bilder-Taggen konnte ich wählen, wie ich es mache. (I was able to control the image tagging.)
- 9. Beim Bilder-Taggen konnte ich so vorgehen, wie ich es wollte. (I could proceed with the image tagging as I wanted.)
- 10. Beim Bilder-Taggen fühlte ich mich unter Druck. (I felt under pressure while image tagging.)
- 11. Beim Bilder-Taggen fühlte ich mich angespannt. (I felt stressed while image tagging.)
- 12. Ich hatte Bedenken, ob ich das Bilder-Taggen gut hinbekomme. (I had doubts whether I perform well in the image tagging.)

Item 1–3 of this questionnaire account for the enjoyment subscale, which was applied as an independent questionnaire during the tutorial (see Figure A.2). Item 4–6 make up the competence subscale, item 6–9 the autonomy subscale and item 9–12 the pressure subscale.

*Inwieweit treffen die folgenden Aussagen auf dich zu?					
	stimmt gar nicht	stimmt wenig	stimmt teils- teils	stimmt ziemlich	stimmt völlig
Das Bilder-Taggen hat mir Spaß gemacht.	۲	0	0	٢	0
Ich fand das Bilder-Taggen sehr interessant.	۰	۲	٥	۲	۲
Das Bilder-Taggen war unterhaltsam.	0	0	0	0	0

Figure A.2: Enjoyment subscale of the IMI as applied during the tutorial.

A.3 Hexad

For the Hexad user types questionnaire [30], participants were asked to rate each item on a 7-point scale, originally labeled with 1=trifft überhaupt nicht zu, 2=trifft nicht zu, 3=trifft eher nicht zu, 4=weder noch, 5=eher zutreffend, 6=trifft zu, 7=trifft voll und ganz zu (1=entirely disagree, 2=mostly disagree, 3=somewhat disagree, 4=neither agree nor disagree, 5=somewhat agree, 6=mostly agree, 7=entirely agree).

Inwieweit treffen die folgenden Aussagen auf dich zu? (To what extent do you agree with the following statements?)

- 1. Ich mag es, Hindernisse zu überwinden. (I like overcoming obstacles.)
- 2. Es ist mir wichtig, meine Aufgaben immer vollständig zu erledigen. (It is important to me to always carry out my tasks completely.)
- 3. Ich tue mich schwer, mich von einem Problem abzuwenden, bevor ich eine Lösung gefunden habe. (It is difficult for me to let go of a problem before I have found a solution.)
- 4. Ich mag es, schwierige Aufgaben zu meistern. (I like mastering difficult tasks.)
- 5. Ich provoziere gerne. (I like to provoke.)
- 6. Ich mag es, den Status Quo in Frage zu stellen. (I like to question the status quo.)
- 7. Ich sehe mich als Rebell. (I see myself as a rebel.)
- 8. Ich halte mich nicht gerne an Regeln. (I dislike following rules.)
- 9. Es ist mir wichtig, meinen eigenen Weg zu gehen. (It is important to me to follow my own path.)
- 10. Ich lasse mich oft von meiner Neugier leiten. (I often let my curiosity guide me.)
- 11. Ich probiere gerne Neues aus. (I like to try new things.)
- 12. Mir ist meine Unabhängigkeit wichtig. (Being independent is important to me.)
- 13. Es bereitet mir Freude, wenn ich anderen helfen kann. (It makes me happy if I am able to help others.)
- 14. Ich helfe anderen gerne dabei, sich in neuen Situationen zurechtzufinden. (I like helping others to orient themselves in new situations.)

- 15. Ich teile mein Wissen gerne mit anderen. (I like sharing my knowledge with others.)
- 16. Mir liegt das Wohl anderer am Herzen. (The well being of others is important to me.)
- 17. Ich mag Wettbewerbe, bei denen ich einen Preis gewinnen kann. (I like competitions where a prize can be won.)
- 18. Belohnungen sind eine tolle Möglichkeit, mich zu motivieren. (Rewards are a great way to motivate me.)
- 19. Es ist wichtig für mich, dass ich einen Nutzen von meinem Aufwand habe. (Return of investment is important to me.)
- 20. Wenn der Lohn stimmt, strenge ich mich gerne an. (If the reward is enough I will put in the effort.)
- 21. Mir ist Interaktion mit anderen wichtig. (Interacting with others is important to me.)
- 22. Ich bin gerne Teil eines Teams. (I like being part of a team.)
- 23. Es ist mir wichtig, mich als Teil einer Gemeinschaft zu fühlen. (It is important for me to feel like I am part of a community.)
- 24. Ich mag Gruppenaktivitäten. (I enjoy group activities.)

A.4 Conclusive Questionnaire

The questionnaire was rated on a 5-point scale, originally labeled with 1=trifft nicht zu, 2=trifft eher nicht zu, 3=trifft teils-teils zu, 4=trifft eher zu, 5=trifft zu (1=strongly disagree, 2=disagree, 3=neither, 4=agree, 5=strongly agree). For participants in the Choice_{Gamification} and Choice_{No Gamification} conditions, the question and items were worded as follows:

Inwieweit treffen die folgenden Aussagen auf dich zu? (To what extent do you agree with the following statements?)

- Ich finde es gut, dass ich entscheiden konnte, ob ich Spielelemente nutzen will oder nicht. (I liked that I was able to decide whether I want to use game elements or not.)
- 2. Ich hätte gerne noch weitere Entscheidungsmöglichkeiten gehabt. (I would have liked to have more choices.)

For participants in the Gamification condition, on the other hand, the wording of the first item was adapted:

Inwieweit treffen die folgenden Aussagen auf dich zu? (To what extent do you agree with the following statements?)

- 1. Ich hätte es gut gefunden, wenn ich entscheiden hätte können, ob ich Spielelemente nutzen will oder nicht. (I would have liked it if I could have decided whether I wanted to use game elements or not)
- 2. Ich hätte gerne noch weitere Entscheidungsmöglichkeiten gehabt. (I would have liked to have more choices.)

In both versions, the conclusive questionnaire was followed by a free text field, where participants were asked to elaborate on which choices they would have liked to have in particular.

Appendix B Set of Rules Formulated by the Independent Raters

As a reminder, the task was worded as follows: "Please enter as many keywords as you can think of to describe the overall mood in the respective image."

Since 'mood' was not further defined, all tags that in the broadest sense relate to a mood (be it in a more psychological way like e.g. 'sad', or in an artistic way like e.g. 'mysterious') are acceptable. Each tag is rated on a scale of 1–3:

A tag fits an image well (3), if:

- The tag describes a mood and fits the image.
- All word forms (nouns, adjectives, verbs, ...) are admissible. For example, if the tag 'love' is rated as 3, then 'loving' and 'loved' must be rated as 3.

A tag fits an image (2), if:

- The tag describes a mood, but does not fit the image.
- The tag does not describe a mood, but fits the image.
- The tag is a color that takes up most of the image
- The tag describes an element that can be seen in the image.
- Again, all word forms are rated equally. If the tag 'love' is rated as 2, then 'loving' and 'loved' must be rated as 2.

A tag does not fit an image (1), if:

- The tag does not describe a mood and does not fit the image in any way.
- The tag is a color that only appears marginally (or not at all) in the image.
- The tag is an article or consists only of random chars
- The tag is not a German word

Bibliography

- Malik Almaliki, Nan Jiang, Raian Ali, and Fabiano Dalpiaz. 2014. Gamified culture-aware feedback acquisition. In 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing. IEEE, 624–625. http://dx.doi.org/10.1109/UCC.2014.99
- [2] Teresa M. Amabile and Judith Gitomer. 1984. Children's artistic creativity: Effects of choice in task materials. *Personality and Social Psychology Bulletin* 10, 2 (1984), 209–215. http://doi.org/10.1177/0146167284102006
- [3] Max V. Birk, Maximilian A. Friehs, and Regan L. Mandryk. 2017. Age-Based Preferences and Player Experience. In *Proceedings of the 2017 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '17).* ACM, 157–170. http://doi.acm.org/10.1145/3116595.3116608
- [4] Marc Busch, Elke Mattheiss, Wolfgang Hochleitner, Christina Hochleitner, Michael Lankes, Peter Fröhlich, Rita Orji, and Manfred Tscheligi. 2016. Using Player Type Models for Personalized Game Design-An Empirical Investigation. *IxD&A* 28 (2016), 145–163.
- [5] Edward L. Deci, Richard Koestner, and Richard M. Ryan. 1999. A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation. *Psychological Bulletin* 125, 6 (1999), 627–668. http: //doi.org/10.1037/0033-2909.125.6.627
- [6] Edward L. Deci, Richard Koestner, and Richard M. Ryan. 2001. Extrinsic Rewards and Intrinsic Motivation in Education: Reconsidered Once Again. *Review of Educational Research* 71, 1 (2001), 1–27. https://doi.org/10. 3102/00346543071001001
- [7] Edward L. Deci and Richard M. Ryan. 1985. The general causality orientations scale: Self-determination in personality. *Journal of Research in Personality* 19, 2 (1985), 109–134. https://doi.org/10.1016/0092-6566(85) 90023-6
- [8] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From Game Design Elements to Gamefulness: Defining "Gamification". In Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments (MindTrek '11). ACM, 9–15. http://doi.acm. org/10.1145/2181037.2181040

- [9] Oluwaseyi Feyisetan, Elena Simperl, Max Van Kleek, and Nigel Shadbolt. 2015. Improving Paid Microtasks through Gamification and Adaptive Furtherance Incentives. In Proceedings of the 24th International Conference on World Wide Web (WWW '15). International World Wide Web Conferences Steering Committee, 333–343. http://doi.org/10.1145/2736277.2741639
- [10] Andy Field. 2009. Running one-way ANOVA on SPSS. In Discovering Statistics Using SPSS. SAGE Publications Ltd, 375–379.
- [11] Juho Hamari, Jonna Koivisto, Harri Sarsa, and others. 2014. Does Gamification Work?–A Literature Review of Empirical Studies on Gamification.. In Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS '14). IEEE, 3025–3034. http://dx.doi.org/10.1109/HICSS. 2014.377
- [12] Sheena S. Iyengar and Mark R. Lepper. 2000. When Choice is Demotivating: Can One Desire Too Much of a Good Thing? *Journal of Personality and Social Psychology* 79, 6 (2000), 995–1006. http://doi.org/10.1037/ /0022-3514.79.6.995
- [13] Yuan Jia, Bin Xu, Yamini Karanam, and Stephen Voida. 2016. Personalitytargeted Gamification: A Survey Study on Personality Traits and Motivational Affordances. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, 2001–2013. DOI:http: //dx.doi.org/10.1145/2858036.2858515
- [14] Daniel Johnson, Sebastian Deterding, Kerri Ann Kuhn, Aleksandra Staneva, Stoyan Stoyanov, and Leanne Hides. 2016. Gamification for health and wellbeing: A systematic review of the literature. *Internet Interventions* 6 (2016), 89–106. DOI:http://dx.doi.org/10.1016/j.invent.2016. 10.002
- [15] Jonna Koivisto and Juho Hamari. 2014. Demographic differences in perceived benefits from gamification. *Computers in Human Behavior* 35 (2014), 179–188. https://doi.org/10.1016/j.chb.2014.03.007
- [16] Pascal Lessel, Maximilian Altmeyer, Marc Müller, Christian Wolff, and Antonio Krüger. 2016. "Don't Whip Me With Your Games": Investigating "Bottom-Up" Gamification. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16). ACM, 2026–2037. http://doi.acm.org/10.1145/2858036.2858463
- [17] Pascal Lessel, Maximilian Altmeyer, Marc Müller, Christian Wolff, and Antonio Krüger. 2017. Measuring the Effect of "Bottom-Up" Gamification in a Microtask Setting. In Proceedings of the 21st International Academic Mindtrek Conference (AcademicMindtrek '17). ACM, 63–72. http://doi.acm.org/ 10.1145/3131085.3131086

- [18] Jana Machajdik and Allan Hanbury. 2010. Affective Image Classification Using Features Inspired by Psychology and Art Theory. In Proceedings of the 18th ACM International Conference on Multimedia (MM '10). ACM, 83–92. http://doi.acm.org/10.1145/1873951.1873965
- [19] Andrzej Marczewski. 2015. User Types. In Even Ninja Monkeys Like to Play: Gamification, Game Thinking & Motivational Design. CreateSpace Independent Publishing Platform, 165–177.
- [20] Elisa D Mekler, Florian Brühlmann, Klaus Opwis, and Alexandre N Tuch. 2013. Do Points, Levels and Leaderboards Harm Intrinsic Motivation?: An Empirical Analysis of Common Gamification Elements. In Proceedings of the First International Conference on Gameful Design, Research, and Applications (Gamification '13). ACM, 66–73. http://doi.acm.org/10.1145/ 2583008.2583017
- [21] Elisa D. Mekler, Florian Brühlmann, Alexandre N. Tuch, and Klaus Opwis. 2017. Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior* 71 (2017), 525–534. http://doi.org/10.1016/j.chb.2015.08.048
- [22] Lennart E. Nacke, Chris Bateman, and Regan L. Mandryk. 2014. BrainHex: A neurobiological gamer typology survey. *Entertainment computing* 5, 1 (2014), 55–62. https://doi.org/10.1016/j.entcom.2013.06.002
- [23] Scott Nicholson. 2015. A RECIPE for Meaningful Gamification. In *Gami-fication in Education and Business*. Springer International Publishing, 1–20. http://doi.org/10.1007/978-3-319-10208-5_1
- [24] Rita Orji, Kiemute Oyibo, and Gustavo F. Tondello. 2017. A Comparison of System-Controlled and User-Controlled Personalization Approaches. In Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization (UMAP '17). ACM, 413–418. DOI:http://dx.doi.org/ 10.1145/3099023.3099116
- [25] Rita Orji, Gustavo F. Tondello, and Lennart E. Nacke. 2018. Personalizing Persuasive Strategies in Gameful Systems to Gamification User Types. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). ACM, Article 435, 14 pages. http://doi.acm.org/10.1145/ 3173574.3174009
- [26] Rita Orji, Julita Vassileva, and Regan L. Mandryk. 2014. Modeling the efficacy of persuasive strategies for different gamer types in serious games for health. User Modeling and User-Adapted Interaction 24, 5 (2014), 453–498. https://doi.org/10.1007/s11257-014-9149-8
- [27] Richard M. Ryan and Edward L. Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation. *American Psychologist* 55, 1 (2000), 68–78. DOI:http://dx.doi.org/10.1037/0003-066X.55.1.68

- [28] Steve Stemler. 2001. An overview of content analysis. *Practical Assessment, Research & Evaluation* 7, 17 (2001), 137–146.
- [29] Ezra Stotland and Arthur L. Blumenthal. 1964. The reduction of anxiety as a result of the expectation of making a choice. *Canadian Journal of Psychology* 18, 2 (1964), 139–145. http://doi.org/10.1037/h0083492
- [30] Gustavo F. Tondello, Rina R. Wehbe, Lisa Diamond, Marc Busch, Andrzej Marczewski, and Lennart E. Nacke. 2016. The Gamification User Types Hexad Scale. In Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '16). ACM, 229–243. DOI:http://dx.doi. org/10.1145/2967934.2968082
- [31] Luis von Ahn and Laura Dabbish. 2004. Labeling Images with a Computer Game. In Proceedings of the 2004 SIGCHI Conference on Human Factors in Computing Systems (CHI '04). ACM, 319–326. http://doi.acm.org/10. 1145/985692.985733
- [32] Luis von Ahn and Laura Dabbish. 2008. Designing Games with a Purpose. Commun. ACM 51, 8 (2008), 58–67. http://doi.acm.org/10.1145/ 1378704.1378719
- [33] Luis von Ahn, Ruoran Liu, and Manuel Blum. 2006. Peekaboom: A Game for Locating Objects in Images. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06). ACM, 55–64. http://doi. acm.org/10.1145/1124772.1124782
- [34] Matthias Wilde, Katrin Bätz, Anastassiya Kovaleva, and Detlef Urhahne. 2009. Überprüfung einer Kurzskala intrinsischer Motivation (KIM). Zeitschrift für Didaktik der Naturwissenschaften 15 (2009), 31–35. https: //pub.uni-bielefeld.de/record/2404161
- [35] Miron Zuckerman, Joseph Porac, Drew Lathin, Raymond Smith, and Edward L. Deci. 1978. On the Importance of Self-Determination for Intrinsically-Motivated Behavior. *Personality and Social Psychology Bulletin* 4, 3 (1978), 443–446. https://doi.org/10.1177/014616727800400317