The User Experience of a Crowdsourced Photo Tagging Game: Evaluation with Usability Tests

Lotta Söderholm

Helsinki August 30, 2013
Master’s thesis
UNIVERSITY OF HELSINKI
Department of Computer Science
There are big amounts of photos taken and uploaded to the internet every year, but few of them are labelled in any way that makes searching for them possible. At the same time people spent hours playing casual computer games like Solitaire or Mahjong.

Phototag, a memory game, was designed to take advantage of this by crowdsourcing the tagging of photos. A prototype of the game has been developed and the purpose of this study was to evaluate the entertainment value of the game and the value of the tags produced during game play. The tests were done with two groups of participants. One group played the game with random photos and another group with family members’ photos. The tests were conducted by modifying classical usability evaluation methods to fit the evaluation of the user experience of the game. Data was collected using think-aloud protocol, video material, questionnaires and logged data from the game. The tags added during game play were evaluated by the participants and the researcher.

The tests showed that the game had some playability problems: efficiency, learnability and memorability problems were found. A lot of unnecessary clicking was required during the game, which made the players easily frustrated. The players also found the writing of tags tiresome and many tried playing the game without adding any tags.

Some of the participants expressed that they liked the game and would play it again, whereas others said they would not. The participants that normally like playing more casual games were also more satisfied with the game.

Most of the tags produced describes the photos well and were the same tags the original photograph had added to the photos. More than half of the tags could be used as such, but there were also poor tags added such as codewords, acronyms, typos and words that only meant something to the player. This shows that the tags would need to go through quality checks before they can be used.

ACM Computing Classification System (CCS):

Human-centered computing -> Usability testing; Laboratory experiments; User studies;
Software and its engineering -> Software usability;
## Contents

1 Introduction 1

2 Description of the Phototag game 3  
   2.1 Motivation behind the game 4  
   2.2 The prototype 6

3 Usability and User Experience 11  
   3.1 Definition of User Experience 11  
   3.2 User Experience in games 14  
      3.2.1 Playability 15  
      3.2.2 Psychology of User Experience in games 16  
      3.2.3 Core elements of the gaming experience 18

4 Methods for evaluating User Experience 20  
   4.1 Inspection methods 20  
   4.2 Usability testing 21  
      4.2.1 How to manage a usability test 23  
      4.2.2 Think-aloud 25  
      4.2.3 Questionnaires 27  
      4.2.4 Metrics and how to collect data 28

5 Test protocol and execution 30  
   5.1 Objective for the tests 30  
   5.2 Metrics and data collected 31  
   5.3 Test participants 35  
   5.4 Settings for the game 36  
   5.5 Procedure for the tests 36  
   5.6 How the data was analysed 38

6 Results 40  
   6.1 Participants 40  
   6.2 Entertainment value 41  
      6.2.1 Playability 41  
      6.2.2 How the players felt about the game 42
6.2.3 Difficulty level ........................................... 44
6.2.4 Frustration ............................................... 45
6.2.5 Writing tags ............................................. 46
6.2.6 Replay value ............................................. 46
6.2.7 Significance of photos .................................. 47
6.3 Value of tags ............................................... 48
  6.3.1 Adding of tags ........................................ 48
  6.3.2 Tags correlation to the photos .................... 49
  6.3.3 The value of the tags to the participants ....... 53

7 Discussion and conclusions .................................. 55

References ................................................................ 57

Appendices

1 Questionnaire

2 Finnish translation of the questionnaire

3 Interview questions concerning the tags

4 Finnish translation of interview questions

5 Interview questions for photo provider

6 Finnish interview questions for photo provider

7 Test layout
1 Introduction

When developing software it is very important to take into account the usability of the system, to make the system pleasing to use. The focus in usability research has long been in functionality and efficiency, but today the research is leaning more and more towards pleasure and aesthetics [Brown 2010]. When developing a computer game these features are even more important to take into account. When playing a game, it is not about performing a task efficiently or getting the job done, but about having fun and wanting to play. Ideally, there is a personal bond created between the player and the game, which makes the player want to play more. This makes the user experience when playing a game much more important than with task-driven applications and it also adds an extra challenge to the creation and evaluation of games.

In the development process it is important to test the usability and user experience of the system as early as possible [Holzinger 2005]. This should be done to reduce the overhead of having to come back and make major changes to the system later in the process. If big faults are found at the end of development, fixing them can be very expensive.

The purpose of this study is to evaluate the entertainment value of a game that is under development. The game is called Phototag and it is a version of the classical memory game, where you try to find a pair of identical cards from the table. In this digital version it is also possible to add notes to the cards to aid the search for them. The game can be played with different pictures on the cards. The pictures can be the player’s own photos, friends’ or families’ photos, or strangers’ photos. The entertainment value of the game might depend heavily on the photos used in the game. The main goal of the study is to evaluate if the game idea is interesting enough to be developed further and if the notes created during game play have value. How the nature of the photos used in the game influences the gaming experience will also be evaluated.

The game is still early in the development cycle and a first prototype of the game was developed in spring 2012. The prototype implements the basic idea for the game with the most basic features. This prototype is going to be used when evaluating the user experience for the game.

For the evaluation, literature on user experience and usability testing was considered and used to inform the planning and design of the study. This includes specific
research on user experience in games, since the priority of user experience in games can differ from task-based applications. By combining usability evaluation methods and research on user experience in games, a testing plan for this particular game was created.

The game and the idea behind the game are explained in Section 2. Earlier research about user experience and evaluation of games are presented in Section 3. In Section 4 usability testing and methods to evaluate user experience in games are discussed. In Section 5 the protocol and execution are explained in detail. The results of the tests are shown in Section 6 and in Section 7 the results from the study are discussed.
2 Description of the Phototag game

In 2009-2011 a three-year study on the gaming habits of Finnish people was done in co-operation between the universities of Tampere, Jyväskylä and Turku [Karvinen and Mäyrä 2011]. The study had all together 3335 participants from ages 10-75, with over 1000 answers for each year. The study showed that 73% of the participants played digital games sometimes and that the most popular game was the casual game Solitaire. Players, with an average age of 37, spent on average 3 hours a week playing computer games. In all age groups the playing of mobile phone games and facebook games was increasing and in the oldest age group, the amount that played any games increased from 75% to 91% in the three years the study was done. This shows that playing games is becoming more and more common all the time and a lot of time is used playing. The idea behind the photo tagging memory game Phototag is to take advantage of this time spent playing and to have the players producing something at the same time as they are playing and having fun.

The game that was evaluated in this study is a modified version of the game Concentration, also known as Memory or Pairs. In this classical memorygame, you have a pack of paired cards that are shuffled and placed face down on a surface. The player gets to flip two cards, one at a time, trying to find a pair. If the cards are not a pair, they will be placed back onto the surface face down on their original place. By remembering which cards are where, pairs gets easier to find. The flow of a single round is illustrated in Figure 1. The game is over when all the pairs in the game have been found.
In the next sections, the motivation behind Phototag and reasons for its development will be explained and the layout of the prototype game will be presented with the help of screenshots.

### 2.1 Motivation behind the game

Today taking photos have been made very simple for people. Professional cameras are now affordable by the public and almost every phone has a decent camera [Ames and Naaman 2007]. Photos are taken daily and often uploaded directly to some cloud service. A cloud service is a service that is always accessible through the net. A humorous example of the amount of cameras today can be seen in Figure 2, where is shown two photos, one from 2005 and another from 2013. The one from 2005 is from when Pope John Paul II’s body was carried away for public viewing and the one from 2013 is right after the election of the new pope. The bright whites spots in the second photo are lights from cameras and cameraphones. Even though the photos might be a bit misleading, it still shows the frequency of cameras today.
Figure 2: A picture, showing the difference between photos taken after the death of the pope in 2005 and the election of a new pope in 2013 [NBCNEWS photoblog 2013].

With these big amounts of photos taken by people, searching for a photo can get tough. Unfortunately, at the moment there does not exist any good automated way of identifying pictures, other than by adding labels to them manually. Few do this and users tend to rather organise their photos with a folder based system [Rodden and Wood 2003]. Ahn and Dabbish have developed a game, where users label images while playing [Ahn and Dabbish 2004]. The game is played by two people over the internet and the players are not able to communicate with each other. Both players are showed the same image and are then supposed to agree on one term to describe the image. This is done by writing suggestion and when both use the same term, the game shows them the next image. The players strive to agree on as many images as possible in 2.5 minutes. Ahn and Dabbish had the game running on the internet and during four months it produced 1.3 millions labels for 293,000 images [Ahn and Dabbish 2004].

The game used in this study also has a similar purpose as the game developed by
Ahn and Dabbish. This game, Phototag, has two purposes. The first one is to be entertaining in itself, the other one is to use crowdsourcing to generate tags for photos.

Crowdsourcing is a concept that have become popular in recent years. It means using a large group of people, most commonly an online community, to carry out a task by dividing the task into subtasks that one person can easily do [Doan et al. 2011].

In the scope of this thesis, tags are identified as metadata that can be associated with a picture to describe it. They are labels that describe parts of the picture, items in the picture or what is happening in the picture. When playing the game, the players have a possibility to add notes that contains tags to cards to aid their playing. These tags are automatically saved and linked to the photos.

The game described here is a prototype and in the actual game the photos would be retrieved from a cloud based service. The service would provide people with an efficient way to store, backup and manage photos using the cloud. A client storing photos in the service could give their permission to the game to use their photos. Then others would play the game with their photos and generate tags for them, without knowing where the photos came from. The tags added by players would be automatically linked to the original photos. By users offering their photos to be used in games, they get free tags for the photos in return. This means that the owner of the photos would not need to manually add tags to aid searching for photos.

The game can also have value in itself. The player playing the game does not need to know that the tags entered will be used for anything, they might play the game for the pure gaming value. The game can have value for two groups of people, the ones providing the photos and the ones playing the game. One of the purposes of this study was to determine if the game has playing value in itself.

2.2 The prototype

The game used in this study is a prototype of a game still in development. This prototype will now be presented and it was used as this in the tests.

Phototag is almost like the classical memorygame, but made into a computer game and with some added features. It can be played with larger amounts of cards than in the classical memorygame and you can add a note to each card to make them easier to find. The note consists of words, which are called tags, that represent the
photo. The game is a single player game and the goal is to find all pairs as quickly as possible.

The game starts with a board of cards face down, as can be seen in Figure 3. The board is made up of pairs of cards with the same photo on them. The main idea of the game is to find all the pairs on the board, by picking up a pair of cards and hoping that they match.

![Figure 3: The board of cards.](image)

The player starts the game by picking up a random card by clicking on it. In Figure 4 the player has clicked a card and the system shows the card to the player. When the player is looking at the photo, he/she can add a note to the photo. The system saves the note together with the card and this will make it easier to find the card later on. After closing (putting down) the card, the player can pick up another card and hope that it matches the first one. If it does not match, the player can also add a note to this second card. Another round in the game starts when the second card is closed. One round in the game means two different cards looked at.
Figure 4: The player has picked up a card and is looking at the front of the card. Here the player is able to add a note to the photo.

If the player picks up a photo that looks familiar, but can not remember where the other pair is, the player can use the search feature. This is done by entering terms to search for into the search field. The system then searches with these terms for tags in the notes that have been added to the photos. If the search term is found the system highlights those cards with yellow. This way the player can find the other pair. In Figure 5 can be seen how the cards, that are tagged with the search term, are highlighted. Now the player can pick up the highlighted card and if he/she remembered correctly how the photo was tagged, a pair might be matched.
Figure 5: The player searches for tags that he/she has added to the photos. The system highlights in yellow the cards tagged with those terms.

The game also has another added feature. By clicking the magnifying glass in the top right corner, the player can see which cards have been seen at least once. This way the player can make sure to pick up an unseen card in the next round. This feature is illustrated in Figure 6. This is to prevent the player from being stuck with picking the same cards over and over again.
Figure 6: The player clicks the magnifying glass-icon and the system highlights in red the cards that have already been seen.

When all the pairs in the game are found the game is over. The player can then choose to play another round of the game.

There are four difficulty levels for the game:

- Easy - a board with ten pairs using random photos.
- Medium - a board with 27 pairs using random photos.
- Hard - a board with 56 pairs using random photos.
- Insane - a board with 100 pairs using very similar photos.

The purpose of this study is to evaluate if the game idea is interesting enough to be developed further and the results of the tests will also give an indication to how the game should be developed. Some added features, that could add value to the game, have been discussed during development of the game. A couple of examples are multiplayer support, integration with social media and timebattle mode. These features are not in the scope of this thesis, but the results of the thesis will give an indication on what new features might be the most important.
3 Usability and User Experience

The international standard ISO 9241-11 defines usability like this [ISO 9241-11]:

Usability: the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.

Usability describes how well, easily and efficiently a product can be used by its users, which include how the users felt about the product. In 1993, Nielsen listed five characteristics that define good usability in systems: learnability, efficiency, memorability, low error rate and satisfaction [Nielsen 1993, pp. 26]. These characteristics are important in any system, but they might not be primary in the user’s interactions with a system [Holzinger 2005]. Part of the usability community consider the experience a user has when encountering and using a system more relevant. The mentioned usability characteristics influences the user’s experience, but user experience goes beyond those characteristics and focuses on the personal outcome a user gets from the interaction. User experience looks at the complete interaction with the system, including the user’s feelings and thoughts.

Usability and user experience concern all human-made objects, but in this thesis these will be considered only concerning software and computer systems.

3.1 Definition of User Experience

Norman et al. introduced the term user experience (UX) in the mid 1990s [Norman et al. 1995] and today the term is widely used in the technology industries [Law et al. 2009]. Even though the term is widely used, it still lacks a clear definition and is understood in different ways. UX is often used as a synonym for usability, user interface, interaction experience, emotion, general experience or as an umbrella term for all of these. UX is also often associated with broader experiences, such as brand experience and service experience. Brand experience include not only the experience from interacting with the product, but also the company, other products and services. Service experience include face-to-face interactions, which user experience does not. Law et al. recommends that the term user experience should be used only for experiences that happen when a user is interacting with a product, system, service or object, through a user interface [Law et al. 2009]. In this thesis
user experience will be used for describing the experience when a user is interacting with a computer system through a user interface.

UX is often seen from different perspectives. Hassenzahl and Tractinsky have examined three different perspectives on UX, which were gathered from the existing literature on UX [Hassenzahl and Tractinsky 2006]. They named the perspectives: non-instrumental, emotion and affect, and the experiential. The first perspective, the non-instrumental, concentrated on the features of the systems that are not instrumental or task-based: the beauty and aesthetics. This also include specific non-instrumental needs like surprise, diversion, intimacy and stimulation. The second perspective, emotion and affect, concentrated on the feelings and emotions the experience brings the user. The last perspective, the experiential, emphasises the unique context of the experience and the fact that the experience is temporary. The unique context is formed by the user, the system and the surroundings of the experience. These perspectives overlap and none of them captures the complete nature of UX. Therefore Hassenzahl and Tractinsky argues that none of these perspectives are alone the right definition for UX, but that they together form the UX, as can be seen in Figure 7 [Hassenzahl and Tractinsky 2006].

Figure 7: Three different perspectives on UX combined [Hassenzahl and Tractinsky 2006].

Hassenzahl and Tractinsky end their paper by stating that UX is a consequence of a user’s internal state, the characteristics of the system and the context [Hassenzahl
and Tractinsky 2006]. In a similar way Roto et al. declared that there are many factors that can affect the UX, but that they can all be classified under these three main categories [Roto et al. 2010]:

**The user.** When a user is interacting with a system, his/her mood, feelings and expectations affect the UX. For example a strategic game can seem boring to the user, if he/she is having a bad day and can not concentrate on the game.

**The system.** Naturally the system affects the UX and not only the systems properties, but also the brand and the user’s added properties. The user’s added properties are modifications done by the user. For example the experience will be different when using a gallery application showing the user’s own pictures, versus the same gallery with strangers pictures.

**The context.** The context is the surroundings of the experience, the physical environment around the user and the information context. Information context is the information the user has about the system. The context of an experience can change and affect the user even though the system stays the same. For example the experience of playing a game can be different when playing the game at the office alone versus at home with friends.

UX can be studied from different point of views: as a phenomenon, a field of study and as a practice [Roto et al. 2010]. This distinction can be explained using an analog: health as a phenomenon, medicine as a field of study and a doctor’s work as a practice. For UX the difference is explained like this:

**UX as a phenomenon** means studying the theory behind the UX, trying to describe what it is and is not, and what it consists of. This include identifying the different types of UX.

**UX as a field of study** means studying the phenomenon: how experiences are formed and what people are experiencing. The field of study helps in finding methods to investigate and develop UX.

**UX as a practice** means using the studies that have been done about UX. For example evaluating UX in a product, demonstrating a certain UX in a product or delivering designs that enables a certain UX.
In this thesis UX will be considered as a subjective, dynamic event, that is influenced by the context, the system and the user. It concerns the feelings and emotions of the user and is affected by the aesthetics, usability and functional aspects of the system. UX will also mainly be considered as a phenomenon and a practice. The phenomenon will be discussed in the theory parts: how UX in games is formed and what subparts it consists of. UX will be considered as a practice when evaluating Phototag.

3.2 User Experience in games

UX does not consist of just one clear event that can be captured and measured, but of an infinite amount of smaller experiences [Thayer and Dugan 2009]. This fact that it is not a clearly measurable event and that the definition of UX is still a bit unclear, makes it difficult to evaluate UX. To make the evaluation of UX possible, it needs to be broken down into measurable parts that can be analysed using known usability assessment techniques [Thayer and Dugan 2009]. When evaluating UX, the type of system that the user is interacting with needs to be taken into account. It needs to be clear what to measure, how and in which context.

When studying UX in games, the particular nature of games needs to be taken into account. As in other digital systems, the functional and usability aspects are important, but in a game the non-instrumental qualities are even more important [Fierley and Engl 2010]. As already presented, the UX is always bound to the user and this is also true for UX in games. This means that the UX in games is also heavily bound to the psychological nature of humans [Takatalo et al. 2010]. In Figure 8 is visualised the relationship between the player and the game and how the UX is a combination of them. On one side we have the game, the way it works, looks and is played. On the other side is the user’s internal state, for example cognition, motivation and attention. Together these two form the UX. The quality, intensity, meaning, value and quality of the experience are all bound to the user and the game.
The term playability is often used to describe the functional and classical usability aspects of a game [Fierley and Engl 2010] and will be explained more clearly in Section 3.2.1. The psychological nature of games and a framework for evaluating UX in games are discussed in Section 3.2.2. In Section 3.2.3 a theory on the core elements of the gaming experience is discussed.

### 3.2.1 Playability

Nacke et al. argues that playability is the evaluative process directed towards games, whereas player experience is directed towards the player [Nacke et al. 2009]. Good playability of a game should be a prerequisite when evaluating UX in games and a good game design should not contain any problems that hinders the player from playing the game. Even though a game idea might be great, bad playability will make the game hard to play and the players will probably not find the experience pleasing.

For the player to have an undisturbed gaming experience, the functional aspects, such as navigation and controls of the game needs to be as natural as possible [Fierley and Engl 2010]. It is important that the player can immerse himself/herself into the game and does not need to consciously concentrate on the navigation and controls. This is important for the flow of the game.
Many of these functional aspects are often intertwined with the classical usability aspects: learnability, efficiency, memorability, low error rate and satisfaction, which are all important for the gaming experience. The faster and easier the player learns the controls and possible actions in the game, the faster he/she can concentrate solely on the game. Low efficiency in a game can also disturb the gaming experience. For example if a trivial action in the game takes a long time to complete and contains unnecessary steps, it can affect the flow of the game. Low error rate is important in any application, but in a game, even a small bug or crash can destroy the experience.

Good functional and usability aspects might be enough to make a successful program, but for a successful game, the non-instrumental qualities are also important to optimize [Fierley and Engl 2010]. These are for example beauty, aesthetics, originality and the user’s possibility to express oneself. These are important, because when playing a game, the focus is not on successfully completing a task, but on the subjective and personal user experience.

3.2.2 Psychology of User Experience in games

A goal for the UX in games is often to create strong feelings in the player and this bounds the UX to the psychological nature of humans [Takatalo et al. 2010]. When playing a game, the player evaluates how he/she is doing in the game, if he/she is good enough to reach the goal. This performance evaluation is done all the time, either consciously or subconsciously and good feelings of success often emerge when the player finally reaches his/her goal. A good game also draws the attention of the player and provides an escape from the real world. Even feelings that normally are considered bad, as frustration or anger, can be desired in a game. They can intensify the feeling of accomplishment when the goal is reached. All of these powerful emotions are a big part of the UX in games.

Takatalo et al. have developed a Presence-Involvement-Flow Framework (PIFF), which is meant to integrate the big number of UX subcomponents in games into one framework [Takatalo et al. 2010]. As can be seen in Figure 9, it is based on two broad concepts: adaption and flow. These concepts are relevant to both the functional components of the game, (e.g. the mechanics, navigation, story) and the psychological components of the UX (e.g. cognition, emotions, motivation).
In PIFF, presence and involvement are grouped together under adaption, which describes the way a player forms a relationship with the game [Takatalo et al. 2010]. Presence describes the player’s experience of being in a game world and is in PIFF divided into interaction, physical presence, attention, role engagement, co-presence and arousal. These subcomponents aid the player’s feeling of being part of the game world: their attention is in the game, they connect with their role figure and with other agents in the game. Involvement in adaption, describes the level of motivation the player feels for the game: how interesting and important the game is for the player. Takatalo et al. describes presence and involvement as indicators for the shift between the real world and the game world.

The second broad concept in PIFF, flow, is defined by Csikszentmihalyi as a positive and enjoyable experience stemming from interesting activity that is considered worth doing for its own sake [Csikszentmihalyi 1975]. In PIFF, flow is divided into cognitive evaluation and emotional outcomes [Takatalo et al. 2011]. When playing a game, the player evaluate their performance all the time, and when the challenges and the skill of the player are in balance, a good flow emerges. This provides emotional outcomes, such as enjoyment, control and happy feelings. The player’s evaluation
of their performance is represented by cognitive evaluation and the user’s feelings that emerged from playing, represent the emotional outcome. To achieve flow there are many factors involved, for example: clear goals, sense of control and instant feedback.

### 3.2.3 Core elements of the gaming experience

Calvillo-Gámez et al. have developed a theory of the Core Elements of the Gaming Experience (CEGE) [Calvillo-Gámez et al. 2010]. Calvillo-Gámez et al. argues that a positive experience when playing games emerges from the player’s perception of the video game and the interaction with it. They define two main components of the experience: *puppetry*, which is the player’s interaction with the game that makes the game their own, and *video game*, which is the game play, concept, rules and environment in the game.

Puppetry is described by Calvillo-Gámez et al. as the interaction between the player and the video game, and that it is affected by three conditions: control, ownership and facilitators [Calvillo-Gámez et al. 2010]. Control of the game is achieved by the possible actions the player can do in the game. Together with external factors, such as previous experiences, control of the game produces a feeling of ownership in the game. This is possible when the player recognises that the actions performed are his/hers. The player’s perception of the video game is formed by the environment and the game play which produces enjoyment. The environment, aka. the sound and graphics of a game, is a big part of the game that affects how the player perceives the game.

In Figure 10 can be seen the core elements of the gaming experience as defined by Calvillo-Gámez et al. The figure also shows the variables of the elements. The facilitators depends on the time the player is ready to play, previous experiences with similar games and the aesthetic values of the game. Here the aesthetic value of the game is the personal value for the player. Control is produced by the small actions the player can do in the game, the players understanding of the goal and by keeping the player occupied. Control also include the learning process for the controls of the game, remembering how to manipulate the game and getting used to the point-of-view in the game. The control and facilitators element then together produce the feeling of ownership with the strategies (big actions) the player uses, the player’s personal goals, the way a player can perform actions that are alien to their everyday life, and the rewards of the game.
Figure 10: The core elements of the gaming experience and its variables [Calvillo-Gámez et al. 2010].

Many of the variables in Calvillo-Gámez et al. theory can be compared to the other theories and concepts already presented. For example the variables memory and controllers, that is part of the element control, corresponds strongly to the usability concepts memorability and learnability. The element facilitators correspond to the subjectiveness of a UX and the ownership factor to the psychology behind the UX that Takatalo et al. [Takatalo et al. 2010] tried to explain.
4 Methods for evaluating User Experience

From early on in game development, game developers have tried to evaluate the UX in their games [Bernhaupt 2010]. The first basic evaluations were the developers trying to play the game themselves and trying to understand what made their game fun or boring.

During different phases of game production, there are different methods and approaches used to evaluate the UX. Even before the production of a game has begun, some evaluation of the game can be done [Bernhaupt 2010]. Without any working game, mostly paper prototyping, peer and expert evaluations are used. In the early stages of development where a working version of the game is available, user testing is the most beneficial, combined with focus groups, interviews, informal play testing and questionnaires [Brown 2010]. The best evaluation method to use for a specific game depends on the type of game and the interaction techniques used in the game [Bernhaupt 2010]. For example exertion games can be evaluated through the movements of the player and social games can be evaluated through the social experience the players get. The type of evaluation to use also depends on the goals for the test.

Classical usability evaluation methods can also be used to test UX [Koeffel et al. 2010]. Methods used to evaluate usability are often divided into three categories: analytic evaluation, evaluation by experts and evaluation by users [Leventhal and Barnes 2008, pp. 214-219]. Analytic evaluation provides ways to predict how the system or interface of a system will work. Evaluation by experts are often so called inspection methods using either heuristics or walk-throughs. In evaluation with users, a real potential user of the system is asked to use the system and data derived from that interaction is analysed. To evaluate playability and UX in games, mostly inspection methods and usability testing is used [Bernhaupt 2010] and these methods will be explained in the next sections.

4.1 Inspection methods

Inspection methods are a set of methods for identifying usability problems and improving the usability of an interface by checking it against established standards. Two of these methods are cognitive walk-through [Polson et al. 1992] and heuristic evaluation [Nielsen 1994].

In a cognitive walk-through a group of developers explore a system’s functionalities
with a task-oriented method [Polson et al. 1992]. The user’s behaviour is simulated step-by-step and the mental process that is required for each step in the task is analysed. After each step a set of inspection questions are answered. Advantages to cognitive walk-through is that it is independent from the users and helps the designers to see the system from a potential user’s perspective. It can help to define users’ goals and assumptions [Holzinger 2005] and to effectively identify problems that arise from interaction with the system. The independence from a user can also been seen as a disadvantage and the evaluation can easily become biased, which might lead to improper task selection and emphasis on low-level details.

In heuristic evaluation experts go through dialogues and interactive elements of the system and judge if they follow established usability rules [Nielsen 1994]. This can be done early in the development process, when there is only the design for the system to be tested, but also throughout the development process. The method can find typical minor or major problems effectively using recognized and accepted principles. The downsides to heuristic evaluation is separation from the end user and that the system is not really used. When a end user actually performs a task in the system, he/she might see the problems another way or have unknown needs that the evaluation can not identify [Holzinger 2005]. It is also unreliable in identifying domain-specific problems.

Heuristics were first presented by Malone in 1980 [Malone 1980], but was adopted more commonly when Nielsen introduced his ten heuristics in 1994 [Nielsen 1994]. In 2002 Federoff adopted Nielsens heuristics into the evaluation of games and introduced 40 heuristics to be used especially for games [Federoff 2002]. In 2004 Desurvire et al. defined a new set of heuristics to evaluate playability in games, which they called HEP (heuristic evaluation of playability) [Desurvire et al. 2004].

Even though using heuristics to evaluate games is not the most common way, there are emerging ready heuristics to be used especially for games [Koeffel et al. 2010]. Many usability problems are the same for a game as for any other digital product and in these areas heuristics are a low-cost way of identifying problems.

4.2 Usability testing

Testing with end users is the most fundamental usability method and is in some sense indispensable [Holzinger 2005]. It provides direct information about how people use a system and their exact problems with a specific interface. Because UX is so
strongly bound to the users and their experiences, a good evaluation method for UX in games is user testing.

Usability testing in games can be done in many different ways, but the key aspect is to have representative users actually playing the game [Brown 2010]. The tests can range from small tests with only a few users to large test groups of hundred users. Small tests are often done in labs, where users play the game with a development team measuring and observing the reactions of the player. Larger tests can also be done in labs, but are often done with the players playing the game at home with some kind of automatic data collection.

Data from usability tests can be collected in many ways: by observing the players [Nielsen 1993, pp. 207], with questionnaires [Leventhal and Barnes 2008, pp. 217-219] and automatic recording [Brown 2010]. The players are given the opportunity to interact naturally with the game and they can be asked to verbalise their thoughts while playing [Bernhaupt 2010]. The tester collects data by observing the interaction with the game and by recording thoughts that the players verbalises. The players can be interviewed or asked to fill in a questionnaire about their feelings and experiences about the game, after the playing is done. Data can also be collected automatically during game play, for example by measuring how long it took for the players to reach a certain level, where in the game they did mistakes or which weapon they preferred to use.

Nielsen addresses some possible problems with usability testing concerning the reliability and validity of the data collected [Nielsen 1993, pp. 166-170]. According to Nielsen the most common mistakes when planning a usability test are to choose the wrong test users, construct defective tasks or not to take into account the time constraints and social influences during a test. The test might be affected by the fact that the participants does not have enough time to get familiar with the equipment or they might feel the need to perform, which also influences the test. The reliability of the test might also be compromised by big differences between test participants. All people are unique and complex beings, which seldom behave or react in the same way, even when chosen from the same target group. Nielsen adds that usability tests are never a waste of time. Even if there are possible mistakes that can be made in a usability study, there is a lot of data that can be collected through usability testing. It always produces some results that then gives some indication to what decisions to make for the system.
4.2.1 How to manage a usability test

There are some common steps to be taken when conducting a usability test [Leventhal and Barnes 2008, pp. 206-219]. First a testing plan needs to be made and for this one needs to understand what it is that needs to be evaluated. When a plan for the test is made, it needs to be tested with a pilot test to remove problems from the plan. When the plan for the test is complete, the test can be conducted, then the data analysed and finally conclusions can be drawn from the data collected.

The plan for the test should include what questions the test is supposed to answer, the protocol for the test, number of participants, schedule, possible compensation to the test participants and resources that are available for the test [Kantner 1994]. A test plan can also include all test material that are needed for the test: instructions, questionnaires etc. When the first version of the plan is ready, a pilot test should be done before the actual tests to test the procedure [Nielsen 1993, pp. 174-175]. The pilot test can possibly show severe problems with the test plan. For example, the test tasks might be too hard or too easy or the tasks might take longer to do than planned. Even when the plan is done thoroughly, the best way to see how the test works in practice is to actually test run it. For pilot testing about two people are enough and they can be anybody that are convenient.

When the pilot tests have been done and the final plan for the test is ready, the recruiting of participants can take place. For an effective test, the test participants should reflect the characteristics of the targeted user and ideally be likely users of the system [Kantner 1994, Nielsen 1993, pp. 175]. This way the results of the study can reflect the experiences that can be expected from real users of the system. Depending on the system, one must also consider if the users should be novice or expert users and if training for the test participants are needed [Nielsen 1993, pp. 177-178]. The participants can be divided into subgroups to test different problems with different users. Dumas and Redish say that a typical usability test should include six to twelve participants in two to three subgroups [Dumas and Redish 1999, pp. 128]. According to them, three to five participants per subgroup is enough to be confident with the results that are collected. With more participants, the same problems will only be seen over and over again.

Even though the possibility of harm to the test participants during a testing of a computer system is small, the tester still has a responsibility towards the participant [Nielsen 1993, pp. 181-185]. The test can be very stressful for the participants and they can often feel a need to perform and please the tester. They can also easily feel
stupid when making mistakes, especially when learning to use the system. Therefore it is important to make the participants feel comfortable, calm and as much at ease as possible during the test. At the beginning of the test there are a couple of things that the participants should be told [Nielsen 1993, pp. 181-185]:

- It is the system that is going to be tested and not the participant.
- If the tester has designed the system, then it is best not to mention that. Otherwise it is good to mention that the tester has no personal stakes in the system, which makes it easier for the participant to be honest and not afraid of hurting the testers feelings.
- That the test is voluntary and that the participant can terminate it at any time.
- The participant should be reminded that they can ask questions at any time, but that the tester might not be permitted to answer them.
- Any additional instructions that will be needed for the test, for example instructions on how to think-aloud.

In usability tests for task-based applications, there are a number of tasks set for the participant to carry out [Leventhal and Barnes 2008, pp. 206-219]. The tasks can be formulated as scenarios for the participant. This is a example of a scenario taken from Dumas and Redish [Dumas and Redish 1999, pp. 172]:

You have just arrived at your desk after a short vacation. Check to see how many mail messages you have waiting for you. If there are any messages from Mr. Green, the Vice President of your company, read them.

During the test, the tester should try not to interfere and to restrain from answering questions concerning the tasks to be done [Nielsen 1993, pp. 190]. The test participant might struggle with the task, but help from the tester might make the participant feel more stupid for needing the help of the tester. Only if the participant seems to get desperate should the tester give a hint or two.

After the test, the participant can be asked questions about the system or asked to fill in a questionnaire [Nielsen 1993, pp. 191]. Questions the participant might have
about the system should be answered after the questionnaires or interviews to avoid bias. Nielsen also points out that it is important to end the session on a positive note.

Fierley and Engl suggest that this classical usability testing with semi-structured procedures is not suitable for game testing [Fierley and Engl 2010]. Undisturbed engagement with the game is necessary to allow immersion, flow and other important dimensions of game experience to unfold. In short and fast games the participants should be given the opportunity to start and play the game several times before an interview takes place. It can be seen that a task in game testing is to play the game, but the tasks should not be as specific as in classical usability testing.

4.2.2 Think-aloud

Think-aloud is one of the most valuable methods used in usability testing [Holzinger 2005]. It involves having a user continuously verbalise their thoughts while using the system and this enables the tester to understand how the user views the system. When the users interacts with the system, they are encouraged to say what they are thinking, why they are doing what they are doing, why certain decisions are made and most importantly what they are feeling.

The think-aloud protocol has its roots in cognitive psychology with Ericsson and Simon making the method popular in the 1980s [Ericsson and Simon 1984]. In the 1990s the usability community started using the protocols for usability testing [Nielsen 1993, pp. 195-200].

When employing think-aloud, the purpose is to capture the content of the working memory of the user [Ericsson and Simon, pp. 14-15]. This should be done with as little influence by the tester as possible. If the tester interfere too much, one can not be sure that the users verbalisations are the reflections of the working memory. For example: if the tester asks why the user did a certain action, the user might start wondering why they actually did the action and therefore either change their action or verbalise more than they thought about in the first place.

A challenge with using think-aloud as a method is to get the participants in the test to learn how to adopt it. Because think-aloud is unnatural, Ericsson and Simon propose that the test participants should be trained in thinking aloud before the actual test [Ericsson and Simon 1984, pp. 376]. This is to show them the difference between describing what they are doing ("I click this") and thinking aloud ("I now
need to click this, because I won't go there next"). Dumas and Redish recommend that the participants are told that their thoughts during the test are interesting, because their reactions to the system are valuable [Dumas and Redish 1999, pp. 280-281]. The participants can be told to act like they were alone in the room, speaking to themselves and that they will be reminded to continue speaking if they fall silent.

Before starting the test, the participants can be given a brief example on how to think-aloud [Dumas and Redish 1999]. The tester can demonstrate think-aloud while performing a task not related to the actual test. Preferably verbalising that he/she do not like something about the task he/she is doing. This way it is shown to the participants that the tester wants more than just explanations of what they are doing and that they are encouraged to express both positive and critical statements.

Even the best test participants will sometimes forget to think-aloud [Dumas and Redish 1999]. They can then be reminded a couple of times to keep on verbalising their actions. Some participants might still stay quiet. If it seems that they are struggling with thinking aloud, it might be better to ask them questions about what they are doing instead of making them think aloud. When asking a question, it is important to try and not bias the test. Ericsson and Simon advises to only use the words: "keep talking", as not to bias the test and interrupt the test participant to much [Ericsson and Simon 1984, pp. 376]. In practice this advice is seldom followed and Krahmer and Ummelen showed that more instructions from the tester can in some cases be beneficial [Krahmer and Ummelen 2004].

When analysing think-aloud data, not only the verbal content has value [Cooke 2010]. Non verbal sounds and silences are also important. Users often become silent when dealing with a hard cognitive task and sounds like "um", "ah" etc., can reveal the user's emotions at the moment. The body language of the user also indicate their feelings.

The term think-aloud most often mean the concurrent protocol, where users talk at the same time as they perform the task. Think-aloud can also be performed retrospective, which means that the task is gone over after it is done and the user can then explain why they performed the task as they did [Fierley and Engl 2010]. This is often done by watching the user do the task from a recording. Van den Haak et al. (2004) compared concurrent think-aloud and retrospective think-aloud and found that they both found the same amount of problems in the user interface [Van den Haak et al. 2004]. They also found that the users employing the concurrent
protocol did not perform worse at the test, even though they had to talk at the same time as performing the task.

Think-aloud might not be a feasible method for measuring UX in games, because it can interfere with the experience [Fierley and Engl 2010]. Aspects of the UX in games, as immersion, flow and attention, might not as easily emerge if the users have to verbalise their thoughts at the same time as playing. For this reason, retrospective think-aloud is more recommended for evaluation of UX in games. For small, more casual games, concurrent think-aloud is still a practical methods, if an immersive game experience is not that important.

In this study concurrent think-aloud will be used, because in this case the game is a simple and casual game. Even so, the game might acquire a lot of the player’s attention and at the testing event the player will not be reminded to think-aloud if it seems to take to much of the player’s attention.

4.2.3 Questionnaires

In the beginning or at the end of a test, the user can be asked to fill in a questionnaire. Before the test the questionnaire can contain questions about the preconceptions about the system and what systems they usually use. After the testing, the users can be given a questionnaire that contains question about how they felt about using the system: Was it boring, difficult etc.?

There are standard questionnaires used for usability of systems, but there are few suitable for game testing [Fierley and Engl 2010]. IJsselsteijn et al. have been developing The Game Experience Questionnaire, which is a promising example, but the questionnaire is still unpublished [IJsselsteijn et al. 2013]. The questionnaire is supposed to capture the gaming experience based on items like positive affect, competence, immersion, flow and challenge.

Calvillo-Gámez et al. have developed the Core Elements of the Gaming Experience Questionnaire (CEGEQ) to measure the core elements of the gaming experience [Calvillo-Gámez et al. 2010]. The questionnaire is based on their Core Elements of the Gaming Experience theory.

For this study, a self-developed questionnaire is used, with some questions taken from the CEGEQ and others derived from the PIFF framework. Because the game studied is a casual and simple game, neither framework are alone suitable and parts are taken from both. To the questionnaire is also added game specific questions
about the tags and user interface.

4.2.4 Metrics and how to collect data

When deciding what to evaluate, the tester needs to decide what metrics to use for the tests and what kind of data is wanted. There is a lot of possible data that can be collected and it is generally divided along two dimensions: qualitative versus quantitative and objective versus subjective data [Leventhal and Barnes 2008]. Data that does not consider the users’ opinion is objective, whereas subjective data is collected from the users’ opinions.

Objective data can for example be collected from the user’s performance when doing a task: e.g. the time it takes to perform a certain task or the errors a user makes [Leventhal and Barnes 2008]. Subjective data can for example be collected with a questionnaire and can deal with the feelings the user had about the system.

Quantitative data is data that can be numerically processed, whereas qualitative data is nonnumeric and needs to be interpreted by the tester [Leventhal and Barnes 2008]. Quantitative data is often collected automatically using for example questionnaires. Qualitative data is often collected by observing the user carrying out the test and by interviewing the user.

When planning a usability test, one must first decide what needs to be measured, what the metrics for the test should be. The metrics to use can be chosen in many different ways. One effective method is the Goal Question Metric approach (GQM) that has been developed by Basili et al. [Basili et al. 1994].

GQM is a top-down approach, using three steps [Basili et al. 1994]. First we have the conceptual level, where the goals for the evaluation is defined: what the purpose, issue and test objects are, including from which viewpoint the measurement should be done. Secondly we have the operational level, where from each goal a set of questions is derived. When the questions derived from one goal are all answered, the goal should be met. The third step is the quantitative level, where the metrics are defined. For each question metrics are defined, that together can answer the question. One metric can answer many questions and the metrics can also be of a qualitative nature as long as they answer the questions. An example of the GQM approach with the three levels and hierarchical structure can be seen in Figure 11.
Figure 11: An example showing the hierarchical structure of the GQM approach.
5 Test protocol and execution

Some methods to evaluate usability and user experience have been presented in the previous sections. On that basis it was decided that the best way to test the UX in Phototag is usability testing. This is because especially the UX is strongly bound to the user and the prototype of the game gives an opportunity to test the game on potential real users of the system. The tests were done employing think-aloud protocol and questionnaires.

Before the actual tests took place, two pilot tests were conducted. They showed that approximately one hour was enough time for the player to get an understanding of the game without making them too tired. It also showed that the participants needed to be more precisely instructed on how to think-aloud.

Next will be presented the protocol used for the tests. The goal for the tests is discussed in Section 5.1. The metrics and questionnaire questions for the tests are presented in Section 5.2. The participants of the tests were divided into two groups, details of the groups are presented in Section 5.3. In Section 5.4 the details of the photos and board used in the tests are presented. The exact procedure of the tests and setting are presented in Section 5.5. How the collected data will be analysed is shown in Section 5.6.

5.1 Objective for the tests

The purpose of these tests was to evaluate the entertainment value of the game and the value of the tags produced in the game. Playability aspects, such as learnability and efficiency were not one of the main objectives for the test. This is because the game used for testing is a prototype and the user interface will still change during development. Though, to rule out major usability flaws and problems as a possible reason for dissatisfaction from the player, playability was still taken into account when evaluating the data from the tests. For example the players might have gotten frustrated over the way cards were chosen, which would affect their experience. When evaluating the data, it was important to try and distinguish these usability problems from how the players otherwise felt about the game.
5.2 Metrics and data collected

The most important question when evaluating the entertainment value is how the players felt about the game. Did they have fun when playing the game? Even though they might have gotten frustrated or bored for a while during the game, how was the overall experience. Did the users leave with a desire to try the game again?

The metrics for these tests were chosen to correspond to the goal of the tests, which was two parted. One of the goals of the tests was to test the entertainment value of the game and the other one was to evaluate the value of the tags produced. To obtain the metrics for the tests, the Goal-Question-Metric approach was used and the results from it can be seen in Table 1 and 2.

The tests produced four kinds of data. There was video material collected from the actual playing of the game and the interview that took place after the playing. The participants were also asked to fill in a questionnaire (See Appendix 1), with questions about how they felt about the game and about playing it. The questionnaire questions were answered on a 7-point likert-scale. This means that the participants answered statements about the game, on a 7-point scale from "strongly agree" to "strongly disagree". The tester also took notes during game play and the interview to capture data that might not have been visible or clear from the video material. An example of this could be the body language of the participant or what strategy he/she was using in the game. Additionally, the game saved a lot of logging data during the game. It logged when a note was added to a photo, when a player found a pair and when a game started and ended.

For the testing of the entertainment value, the player’s feelings about the game, the flow and involvement of the game were evaluated. Also the playability of the game was tested, even thought it was not a major goal for this study. Playability was added to make sure bad playability was not a reason for the players possible dissatisfaction with the game. The questions for the evaluation can be seen in Table 1.
<table>
<thead>
<tr>
<th>GOAL</th>
<th>Purpose</th>
<th>Viewpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>Issue</td>
<td>Evaluate the entertainment value of the PhotoTag-game from a users point of view.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>QUESTION</th>
<th>How did the player feel about the game?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metrics</td>
<td>Questionnaire</td>
</tr>
<tr>
<td></td>
<td>Was it fun? Questions 1-2 and 5-7.</td>
</tr>
<tr>
<td></td>
<td>Was it boring? Questions 10 and 19.</td>
</tr>
<tr>
<td></td>
<td>Was it frustrating? Questions 8-9 and 18.</td>
</tr>
<tr>
<td></td>
<td>Would the player want to play again? Questions 3, 4, 25 and interview question 7.</td>
</tr>
<tr>
<td></td>
<td>Time to find the first pair.</td>
</tr>
<tr>
<td></td>
<td>Average time to find a pair.</td>
</tr>
<tr>
<td></td>
<td>Data from the user thinking aloud.</td>
</tr>
<tr>
<td></td>
<td>Notes taken during the test.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>QUESTION</th>
<th>How was the flow of the game?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metrics</td>
<td>Questionnaire, questions 19-20.</td>
</tr>
<tr>
<td></td>
<td>Average time to find a pair.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>QUESTION</th>
<th>Did the player feel involved in the game?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metrics</td>
<td>Questionnaire, questions 21-22.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>QUESTION</th>
<th>How was the playability of the game?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metrics</td>
<td>Questionnaire, questions 11-17.</td>
</tr>
<tr>
<td></td>
<td>Data from the user thinking aloud.</td>
</tr>
<tr>
<td></td>
<td>Notes taken during the test.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>QUESTION</th>
<th>Do the player ordinary like these kinds of games?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metrics</td>
<td>Questionnaire, questions 23-24.</td>
</tr>
</tbody>
</table>

Table 1: A GQM-table for the evaluation of the game.

When evaluating UX, an important question is: how did the user feel about the game? This is also the first question in the GQM-model. Because this is a hard question to answer it was divided into sub questions: Was it fun? Was it boring? Was it frustrating? Would the player want to play again? These questions were
answered with questions 1-10, 18-19 and 25 in the questionnaire (See Appendix 1) and with an interview question (See Appendix 3).

The data from the players thinking aloud and notes taken during the game were also used to answer the question about how the players felt about the game. Included as objective metrics were also the average time to find a pair and the time to find the first pair in a game. The game can get boring if the time to find the first pair is very long. This means the player does not get any rewards and the playing might start to feel like work. The flow of the game is also influenced by the average time it takes to find a pair. If it is too long, the game might be too difficult and therefore boring.

The second and third question have to do with the concepts that have been showed to indicate the UX in games. Mostly the flow and involvement when playing the game were considered. These questions were answered in the questionnaire with the questions 19 to 22. The flow of the game was also measured with the average time to find a pair.

The fourth question in the GQM-model is about the playability of the game. To eliminate bad playability as a reason for player’s feelings about the game, questions about the playability and controls were asked in the questionnaire with questions 11 to 17. This was also analysed using the data from the players thinking aloud and the notes taken during game play.

The fifth question in the model is, do the player ordinary play these kinds of games. This was added to determine if the participants in the tests are from the target user group. The metrics for this question was the questionnaire questions 23 and 24.

The other part of the study was to evaluate the quality of the tags produced. The metrics used for this can be seen in Table 2.
<table>
<thead>
<tr>
<th>GOAL</th>
<th>Purpose</th>
<th>Issue</th>
<th>Object (process)</th>
<th>Viewpoint</th>
<th>Evaluate the quality of the tags from a users point of view.</th>
</tr>
</thead>
<tbody>
<tr>
<td>QUESTION</td>
<td>Do the tags represent the photos they are associated with?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metrics</td>
<td>Interview</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QUESTION</td>
<td>Comparing the tags to the photos.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metrics</td>
<td>Interview</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QUESTION</td>
<td>Would the player want to add similar tags to their photos?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metrics</td>
<td>Interview</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QUESTION</td>
<td>Would the players want to pay for these kinds of tags?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metrics</td>
<td>Interview</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: A GQM-table for the evaluation of the tags.

After game play, the participants in the tests were shown the tags they produced during the game. They were then asked to look at the tags and compare them to the photos. They were asked if the tags represent the photos and if they would find the photos by doing a search with those tags (See Appendix 3).

The participants were also asked if the tags had any personal value for them. Would they want to add similar tags to their photos and would they even be prepared to pay for those kinds of tags? If they answered no, they were asked to elaborate. There might be many reasons, for example: they personally do not see any need for the tags, the tags are not good enough or they use a better system.

The tags were also evaluated by comparing the tags to the photos. Did they have anything in common? Did the tags say something about an object or event in the photo? Were there a lot of unnecessary tags or a lot of prepositions added?

Except for the metrics already presented, the participants were, on the questionnaire asked if they take a lot of photos, organise their photos and if they upload their photos to some cloud service. This was done to get a an understanding about the participants background, whether they are used to adding tags and handling photos.

In the interview the participants were asked if they would be ready to provide their photos to be used in the game. They were also asked if they would be ready to give
the tags, that they produced, to the owner of the photos. This was done to make sure that the players of the game are comfortable with the idea of producing tags to other peoples photos.

5.3 Test participants

The test participants were divided into two groups, which can be seen in Table 3. For each group a certain amount of people were invited and each test was done individually.

<table>
<thead>
<tr>
<th></th>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photos</td>
<td>Unknown photos</td>
<td>Acquaintance’s photos</td>
</tr>
<tr>
<td>Participants</td>
<td>5</td>
<td>3x2</td>
</tr>
<tr>
<td>Play</td>
<td>Individual play</td>
<td>Individual play</td>
</tr>
<tr>
<td>Levels</td>
<td>Easy, Medium, Hard and optional Insane</td>
<td>Easy, Medium, Hard and optional Insane</td>
</tr>
<tr>
<td>Tag evaluation</td>
<td>Tags evaluated by themselves</td>
<td>Tags evaluated by themselves and their peer</td>
</tr>
</tbody>
</table>

Table 3: Table showing the setup for the two groups used for testing.

The first group played the game with non-familiar photos and the result from these tests measured the plain entertainment value of the game. For this group five people were invited.

For the second group six (3x2) people were invited and the participation was done in pairs. One of the participants provided the photos and the other one played the game with these photos. The pair was familiar with one another, the one providing the photos was a family member to the player. This made the photos used in the test somewhat familiar. The player might have seen the photos, or the objects in the photo might be familiar, but the photos were still not taken by the player. The photos were acquired before the actual test so that they were ready in the game when the player arrived. The test layout was otherwise the same as for the first group. The participant was also asked to evaluate his/hers own tags and the tags were saved for later on. On a later occasion the tags were shown to the one providing the photos. He/She was then asked to evaluate the tags with a semi structured interview.
Test participants for the first group were only asked demographic questions at the recruiting stage, such as age, sex, occupation. If there would have been more available participants than needed, they would have been chosen from different age groups, occupations and an even number of males and females. There were exactly the right amount of available participants and they did not represent different age groups or occupations, but there was a good combination of males and females.

The participants for the second group were recruited through the ones providing the photos. Participants, that take a lot of photos and preferably also organize or tag them in some way, were sought. They needed to have some close friend or relative that would also agree to come and test the game. They had to provide at least 95 photos for the test. Their demographic details were also gathered at the same time as the evaluation of the tags.

5.4 Settings for the game

The easy level had 10 pairs of cards, which were decided on advance. For the first group, the photos used were generic photos of landscapes, nature etc. The medium level was 27 pairs of cards, which were also decided on advance. For the first group half of the photos were generic photos of landscapes, nature etc., and half were photos that could be taken from anyones photo album, including people, happenings etc. The hard level had 56 pairs of cards and for the first group it contained a mixture of photos. For the first group the insane game contained 100 different photos of the Eiffel tower.

The sizes of the board were the same for the second group, but the photos were provided by the participants pair. These photos were randomly divided among the boards.

For the first group (playing with random photos) the testing was done at a laboratory. For the second group (playing with familiar photos), the testing was for some done at the laboratory and for others at other locations that were convenient for the participants.

5.5 Procedure for the tests

To ensure that all events were recorded, the tests were filmed. Primarily the sound was used from the film, but to make it possible to link the sound to an event, the
camera was angled to film the screen and if possible the hands and back of the player. This material was not and will not be used for anything else than this thesis. It will not be shown to anyone outside the scope of the study and will be destroyed when the study is finished.

When the test player arrived at the testing venue, he/she was first given a brief explanation about what we were testing (See Appendix 4). It was explained that the game is a prototype of a game that we are testing for a customer. It was explained that the player will first get to play the game and then afterwards be asked to fill in a questionnaire.

The player was asked to think-aloud during the playing. This included instructions on how to think-aloud and a small example on how it is done. He/She was also told that if it feels hard to think-aloud and at the same time concentrate on the game, he/she should rather concentrate on the game than the thinking aloud. It was made clear to the player that it is the game we are testing and not the player. The player was told that it is fine to ask questions, but the tester might not be able to answer all of them.

The player was made aware of the camera and that the material will be used only for this study and that the camera records only the sound, screen, hands and, if possible, the back of the player. The player was told that he/she can end the test at any time. Finally the player was asked if everything was understood and if he/she has questions before starting.

The controls and rules of the game were taught to the player in the first round of the game. This differs from classical usability testing, but was here done to aid the player in getting familiar with the game and to more quickly get the point of the game. This helped to minimise the learnability aspect of the usability in the game and to reduce the risk that the player got frustrated because he/she did not know all possible controls. If the player felt like he/she want to practice some more, he/she was given the possibility to play another easy level for practice. The actual test began when the player played a medium level game and at this point he/she was reminded to think-aloud.

If the player finished the medium level, he/she was asked to play another game. The player was offered to play a hard level, but if he/she felt he/she wanted a lot more challenge, the player could also try the insane level. After the third game, if there was still time and the player wanted to play more, he/she could. The tester stopped the test, when one hour of the test had passed, if the player had not at this point
wanted to stop playing.

If the player seemed to struggle with the concept of thinking aloud, the tester tried to instead ask questions about why the player did certain actions, what strategy he/she was using and what feelings the player got from the game. Also if the player seemed to be very concentrated on the game, the tester would refrain from interfering or asking questions between games. During the playing, the tester took notes on how the player seemed to enjoy the game, mostly body language. This was important because the camera did not film the face and body of the player.

When the game play was over, the player was asked to fill in a questionnaire with questions about how he/she felt about the game (See Appendix 1). After everything was filled in, the player was shown the tags that was entered during the game. The tags and photos were shown in a table, with the photo on the left and the related tags on the right. The table was automatically produced by the system. The tester conducted a semi structured interview about the value of the tags (See Appendix 3).

The tests looked the same for the two groups, the only thing that differs was the photos used in the test. On a later occasion the ones providing the photos for the second group was interviewed. They were asked to evaluate the value of the tags in a similar interview as the ones that were done after game play for the players (See Appendix 5).

The time reserved for the test sessions when playing the game, was about one hour per player. 15 minutes were reserved for the introduction and 15 minutes for the questionnaire and interview. This left about 30 minutes for game play. In the pilot test, this was approximately the time the test players wanted to play the game.

Test players were asked to reserve 1,5 hours for the test session, so that there was room for some extra time if needed. Some of the players used all this time, because they wanted to finish the level they were on.

The later tag evaluation sessions, with the participants providing the photos, was reserved 30 minutes per participant.

5.6 How the data was analysed

There was a lot of data gathered from the tests: the questionnaire answers, camera material, data from the logs and notes taken during the tests. The camera material
was transcribed into written form, including what the user was doing while expressing something, if this was not clear from context. This was done for two of the first group’s tests. They were chosen on the basis on which players were best able to think-aloud and to express their feelings. The camera material from the other tests was also analysed, but not transcribed, to either support or contradict findings from the ones transcribed.

From the second group, the camera material from one test was transcribed and rest of the material was analysed to either support or contradict findings in the material transcribed. This in the similar way as the first group.

The answers from the questionnaires were also compared to the findings from the camera material. The camera material gave indications to why a player reported that the game was fun or boring in the questionnaire. Sometimes it also seemed from the camera material that the user did not like the game but he/she reported on the questionnaire that he/she really enjoyed the game and would want to play again. In these cases, the filmed material was gone over ones more and the player’s report was doubted. The player might have reported something else than he/she actually thought, to please the tester. The notes taken at the test session also gave an indication to which answer was more probable.

The results from the two groups were compared to each other and this gave an indication to how much the photos used in the game influenced the playing experience. The results from the tests are discussed in the next section.
6 Results

The tests for this study were conducted according to plan. There were two groups of players, one playing with random photos and another with photos provided by somebody close to them. The details of the participants will be more closely explained in Section 6.1.

During game play, the participants were asked to think-aloud and to express any feeling they had about the game. The playing was filmed and the tester wrote down notes on reactions or events that might not be visible from the camera material. After game play the participants also answered a questionnaire and some questions about the tags that they had added.

The main results and observations from the tests are divided into two sections. The entertainment value of the game is presented in Section 6.2 and the value of tags produced during game play is presented in Section 6.3.

6.1 Participants

There were two groups of participants chosen for the tests. One group with five participants, that played the game with random photos and a second group with six participants, where three of them played the game with photos provided by the other three. The participants have been given names that have nothing to do with their real names but correspond to their sex.

In the first group there were two males: Cain and Dan, and three females: Ada, Bea and Eve, with ages ranging from 20 to 27. Ada, Cain and Eve reported that they play computer games more than once a week. All, except Cain, reported that they often play casual games, like Solitaire or Angry Birds. Bea and Dan reported that they take photos at least once a week and Bea, Dan and Eve answered that they organise their photos in some way. All of the participants in the first group were technically competent and they were either students or software developers.

In the second group there were six participants, in three pairs. There was Fanny, who played the game with her sister’s (Ina) photos; Gina, who played with her daughter’s (Jenny) photos; and Holly, who played with her son’s (Chris) photos. The ages of Fanny, Gina and Holly ranged from 43 to 58. The ages for the ones providing the photos ranged from 27 to 34.

From the ones playing the game in the second group, Fanny and Holly reported that
they take a lot of photos. All three reported that they organise their photos in some way and Holly added that she also adds tags or labels to them. None of the ones playing the game in the second group were technically competent, but they were used to using a computer, either at home or at work. Holly reported that she often play computer games and casual games, whereas Fanny and Gina reported that they do not. From the ones providing the photos, Ina was the only one that reported that she adds tags or labels to her photos, but they all organise their photos in some way. Chris and Ina reported that they take photos often.

6.2 Entertainment value

The first part of the study was to evaluate the entertainment value of the game. This was divided into five parts concerning the players’ feelings about the game, the flow of the game, the players’ involvement in the game, the playability of the game and the players’ game preferences.

In the next sections the results of the tests for the entertainment value are presented. First the playability of the game is discussed, then the players’ feelings about the game, the difficulty level, frustration in the game and replay value of the game.

6.2.1 Playability

Even though playability was not a major objective for the evaluation, it was clearly a problem in the game. All players expressed that the playability or user interface of the game was in some way bad. Some players were able to play the game without bigger problems, but suggested ways the user interface could be improved. Other players struggled a lot with the user interface during the game. The main problems were the rules of the game, the fact that the players needed to click a lot in the game and the adding of tags.

The rules of the game were clear to most players, but from the second group two players struggled with them. Gina answered that she did not understand the rules and during the game she struggled with the flow of the game. She seemed nervous and kept saying that she is bad at games like this. To both Gina and Fanny it was very unclear that one round in the game is two cards looked at. That after the second card is seen, the system will close both cards. Fanny said: "confusing when you have to take two cards at a time" (original: "hämmentävää kun pitää paria ottaa kerralla"). The younger players thought this function was clear, as this is the
way a normal memorygame works.

Three of the players answered that the controls were not easy to use and two players answered that the controls did not respond as they expected. The game clearly had an efficiency problem, for example Bea said: "It’s a fun game, but the controls ruin it, so that you have to click a lot". Others also complained about the amount of clicking that had to be done in the game. Often if the players did a miss-click on the second card, they got frustrated that they had to open up both cards again. This meant they needed to do a lot of unnecessary clicks when they already knew which ones were a pair.

The game also had learnability problems. Almost all of the players tried to close the big photo (put the card back onto the board) by clicking outside the photo. Many players also did this many times during the game. Dan said: "If you want to close [the photo], it would be easier to click outside or throw to the side" (original: "jos haluaa sulkea, niin että klikkaisi ulkopuolella tai heittäisi sivuun"). Fanny also reported that the fact that the photo pops up on top of the others is confusing and makes her forget where the card was on the board: "Confuses a bit when the picture comes up like that" (original: "hämää vähän kun kuva tulee esille noin").

Five out of eight players reported that the adding of tags was somewhat difficult. One player could not say either way. The adding of tags had at least one learnability and memorability problem that Fanny struggled with. She got the "search" and "add note" fields confused often. She tried adding tags by entering them into the "search"-field and could later not understand why the photos could not be found using that tag.

Players also made comments about the search functions, because there was a flaw in the game, which made the searched cards sometimes remain highlighted until another search was done and sometimes the highlights would disappear when a card was opened.

6.2.2 How the players felt about the game

The feelings after the game varied a lot between players. From the first group, four players answered in the questionnaire that they at least somewhat liked and enjoyed the game. None expressed that they wanted to spend a lot of time playing the game, but they would like to play the game again. Their reactions during the game were not overly enthusiastic.
From the second group, with players playing with familiar photos, Holly reported that she really liked and enjoyed the game. On the questionnaire she answered that she would like to play the game again and if possible also spend a lot of time with the game. She confirmed this verbally. She said: "these are exactly the kinds of games that I like to play and I would be really hooked on the game" (original: "tämä on just sellaisia pelejä mitä mä tykkään pelata ja peli koukkuttaisi minua todella paljon").

In both groups there was one player that clearly did not like the game: Cain, from the first group, and Gina from the second group. Cain found the game too easy and got bored during the game. The insane level gave him a better challenge, that he first got exited about, but halfway through the game he expressed that the playing started to feel numbing (original: "Menee vähän puuduttavaksi") and stopped playing. Gina clearly had problems with the user interface and also got frustrated when she did not understand how the game worked. She said "I would not play this at home" (original: "en kyllä kotona pelaisi").

As can be seen in Figure 12, the time to find the first pair in the game can sometimes be quite long, especially in the more difficult levels. It has to be taken into account that it sometimes happens that the first pair is found by chance. The figure shows that it can take up to ten minutes to find the first pair, which is quite a long time and many players might get bored of the game during this time, because they get no positive reinforcements. Additionally this gives the game a feeling of work, which some players also implied. For example Cain said that the game felt like a mechanical performance (original: "siitä tulee aika mekaaninen suoritus"). Fanny also said that the game felt a lot more fun immediately when "something started to happen" (original: "heti huomaa että kun jotain alkaa tapahtumaan, niin paljon kivempaa").
6.2.3 Difficulty level

For a game to be entertaining the difficulty level and learning curve of the game needs to be appropriate. The players in the first group were somewhat pleased with the difficulty level of this game. The players found the easy and medium level relatively easy, but the hard and insane level provided enough challenge. Bea expressed that the medium and hard level were relatively easy, but during the insane level said "are you kidding me" and "this is a bit too difficult for a casual game now". The difference between the hard and insane level might have been a bit too steep.

From the second group, Gina and Fanny found the game a bit too difficult. Gina played the easier levels a couple of times, because she felt those provided her enough of a challenge. Fanny did not explicitly express that the game was too difficult for her, but she also struggled with the easier levels.

This difference between the two groups can also be seen in Figure 13, which shows the average time it took the players to find a pair. The different colours in the figure represent the different levels in the game and only the games that were played to the end are displayed. The times are clearly longer for the second group. This indicates that the game felt more difficult for the players in the second group and
one reason for this might be the difference in technical competence between the two groups. The players in the first group were also younger and more used to playing these kinds of games.

Figure 13: This chart shows the average time to find pairs in the game.

In Figure 13 can also be seen that the time to find a pair did not really increase that much with the difficulty level. The players played an easy level to learn how the game works and to train at first, but they were still getting familiar with the game in the medium level. This might explain why the average time to find a pair is almost the same in the hard and medium level. For the players in group 2 the chart shows clearly that they got a lot better at the game in the hard level. During the playing it could also be observed that the players were a lot more comfortable with the game in the hard level. Then the players had already figured out the controls and how to play the game the fastest.

6.2.4 Frustration

All players seemed at some point frustrated during the game, even though all did not report this. This frustration during the game was not necessarily a bad factor and it did not seem to hinder the playing experience. It might have enhanced the experience by bringing forth positive feelings when succeeding in the game. This
might also be a reason for why the players did not report being frustrated during the game, as the game was still enjoyable. There were reactions like "no.", groaning and sighs from the players during the game, but also "yes.", "ah, there it was." and "yes, these two I have already seen".

The players that liked the game the least reported being frustrated after and during the game. For example Fanny thought aloud: "now my nerves are going" (original: "nyt alkaa hermot menee") during the game and Gina groaned and seemed very frustrated at times when she thought she knew where a pair could be found, but remembered wrong.

The reasons for frustrations seemed to be the adding of tags, the user interface and the challenge in the game. The challenge in the game seemed to be at a right level for most players, but at least Gina felt the game was to difficult.

6.2.5 Writing tags

Four of the players expressed that it was to some degree frustrating to add tags to the photos. This could also be observed with six players trying to play the game without adding any tags. Gina said: "I will not write any more, it is so boring" (original: "en rupee kirjoittamaan, se on niin tylsää"), but quickly got frustrated when she could not remember where the cards were without tags. Two players, Bea and Ada, played the last games without using any tags and seemed less frustrated using this method. The others started adding tags again after a while, seeing it made the game easier. Cain felt that because it was possible to add tags, it felt natural to do so: "the tagging comes naturally, when it is possible [to do so]" (original: "tagaaminen tulee vähän luonnostaan, kun se on mahdollista [tehdä niin]").

Some of the players also complained about the adding of tags taking a long time and tried to find ways to speed up the process. For the older players it seemed a bit difficult to move freely between the mouse and keyboard, as they were not as used in doing so. The writing of tags also took longer for the older players as they needed to watch the keyboard and what they were typing.

6.2.6 Replay value

Four out of seven players expressed that they would like to play the game again, but for different reasons. From the first group, Eve expressed that she would like to play the game again to improve her skills and maybe try different strategies: "I don’t
know for how long I would play, but I would want to try it again, just because I now know how to do it better" (original: "en tiedä miten pitkään jaksaisin pelata, mutta tekisi mieli kokeilla uusiksi, ihan vaan sen takia että osaisin paremmin"). From the second group, Fanny said she wanted to play the game again, if it would always contain photos of loved ones. Holly really enjoyed the game and expressed that she would like to play it a lot if she had the time.

From the first group, three of the players would pay for the game. They would pay something between 50 cents and one Euro, if the user interface was improved and the game could be played on for example a tablet or mobile phone. From the second group, Holly would pay up to 10 Euros for the game, if it could be downloaded to the computer. She added that she would not pay anything for an internet game. Fanny said she might pay 8 Euros for the game, if the photos would change often enough and if it would always contain photos from friends and family.

6.2.7 Significance of photos

The nature of the photos used in the game naturally influenced the gaming experience. Photos of family and friends created positive feelings and caught the attention of the players.

The second group that played with familiar photos clearly put more attention to the photos than the first group. They sometimes stopped playing to watch the photos and to say something about the content. Holly also reported that there were times in the game when she was doing nothing, because she stayed to look at the photos (original: "Jäin katselemaan kuvia"). All of the players in the second group commented on the photos in some positive manner, for example: "Hoho, here are so good photos that I haven’t seen before" (original: "Hoho, tääellä on niin hyviä kuvia mitä en ole nänyt") and "Lovely picture of Irina" (original: "Ihana kuva Irinasta").

Together with the fact that the participants in the second group were less used to playing, this was a reason for the times to find pairs being longer in the second group. The times can be seen in Figure 13 and 12. The players from the second group also seemed to remember more clearly what tags they had added to the photos, because the photos were connected to feelings and also often contained names of the persons in the photos.
6.3 Value of tags

The first group created almost 900 unique tags for the photos and the second group about 300 tags. The second group’s tags could not be compared to each other, because the photos were different, but for the first group many of the players added the same or similar tags for the photos. For example photos with a cat were often given the tag "cat" by players. The players also added only one tag to each photo most of the time in the easier games. In the insane game, more tags were needed to find the photos.

More details on what kinds of tags the players added are given in the Section 6.3.1. In the Sections 6.3.2 and 6.3.3 the tags are compared to the photos and the value of the tags for the players is shown.

6.3.1 Adding of tags

The way the players added tags and named the photos varied a lot. Some used code-words to try and speed up the process of adding tags, others entered half sentences. In the insane level, with only photos of Eiffel towers, it was a challenge to come up with individual tags for each photo. Therefore some players started to use words to group the photos together, for example "day", "top" or "down". The player could then check all photos that was "added" to this category, when a similar photo came up. This meant a lot of clicking through the photos, but sometime the player would remember the position of the right card, when there were only some cards to choose from. This also meant that the challenge to remember roughly where the photos are situated on the board got relevant again.

Dan used a code system with the tags. He only added the first or first two characters in the word or words he would use to describe the photo. For example "santa" was "sa", "dog" was "d" etc. He quickly encountered the problem that many photos had a tag with for example "a", but still found this faster than actually writing the full tags. In the insane level he tried using more words, so that "bwsifu" meant "black white sun light fullview under". With this method he found that he could not remember in which order he had added the characters and he did not finish the insane level.

Tags were added in two languages, Finnish and English, depending on which language the test was conducted.
The tags produced in the game developed by Ahn and Dabbish [Ahn and Dabbish 2004], that was presented in Section 2, differs from the tags produced by Phototag. This is because in Ahn and Dabbish’s game the players need to think about how the other person will tag the photo. This means bad tags like offensive words, prepositions, code words etc. would not be created from Ahn and Dabbish’s game. This is because the chance that both players use a bad tag is small unless it really says something about the photo.

6.3.2 Tags correlation to the photos

After the player had played the game, he/she was shown the tags he/she had entered during the game. The photos were shown on the computer screen, with the tags produced by the player beside them. The player was then asked if he/she felt that the tags represented the photos and if he/she felt the photo could be found with its tags. In the second group, the tags were also later on shown to the ones providing the photos. They were then asked the same questions.

From both groups the players answered that yes, most of the tags represent the pictures. One added that she picked words by which she would immediately know which word she had added when looking at the photo. One commented that: "To some extent, I would probably have tagged them like that myself" (original: "jossain määrin olisin varmaan tollein ne nimennyt itse"). The players also found tags that were not so descriptive, but something the player had added that he/she would remember. One said: "one easily adds tags that only mean something to oneself" (original: "helposti menee niin että tagit on sellaisia mitä merkkaa vaan itselleen jotain"). An example of this can be seen in Figure 14, which shows a photo used in the game with a man making some kind of food in pots. For the Finnish words the English translation is added in parenthesis. At least two players tagged this photo with the tag "kebab", which does not say anything accurate about the photo. The man is not making kebab and there is not any kebab in the photo.
Two of the players also pointed out typos they had made and that their tags for the photos of the Eiffel tower were not that good. Photos that were shown in the wrong orientation, were often given tags like "vino" (crooked) or "kallelaan" (askew), which does not say anything about the content in the photo.

The ones that provided the photos answered that the tags represent the photos to some degree and that they would probably find the photos with the tags. Most photos were tagged with words that described the most dominant features in the photo or text that was visible in the photo. However, they also found tags that were wrong, for example Ina found photos that were tagged with names, where the baby or dog in the photo actually was named something else than the tag. Chris also reported that there were some tags he would not even think of searching for, for example one photo was tagged "taksi" (taxi), but the photo showed four men on some kind of grassland.

The tags added by the players in the second group were clearly of another nature than the first group’s tags. Because the people in the photos were familiar, they were almost always tagged with the name of the person/persons in the photo and some other detail. Holly also added very personal tags, that only meant something to her. To one photo she added the tag "toto" and explained that it was because she remembered that the band Toto had stayed at the hotel at the same time as them.
When comparing the tags to the photos it is clear that there are a lot of good tags and some bad tags. Some of the photos could immediately be found with those tags. One example can be seen in Figure 15. The photo shows a river with a bridge in a snowy landscape with ducks. The tags the players had added to the photo were: joki (river), snow, river, snowy, bridge and talvi (winter).

![Tags: joki (river), snow, river, snowy, bridge, talvi (winter)](image)

Figure 15: This figure shows a photo used in the game and on the right the tags the players have added to the photo.

In Figure 16 can be seen another example, where the photo shows in black and white the Eiffel tower from beneath, with some lights on top of the tower and it seems like it has been raining. The tags added by the players to this photo are: alavalo (downlight), bwslfu, bwsluf, lähi (close), kontrasti (contrast), mustavalko (black and white). The tags "bwslfu" and "bwsluf" are code words Dan added as tags.
The photo in Figure 16 is a tricky photo to tag consciously even, but some of the tags do say something about the photo, even if they alone would not describe what is in the photo. The original tags that the owner had added to the photo are: Eiffel tower, paris, france, europe, travel and rain, canon, 5d, cmozzphoto.com. These tags also say something about the context of the photo, which the players of the game can not be aware of.

Another good example can be seen in Figure 17. This photo was tagged by the players with the words: hdri, pieni (small), puisto (park), ihmisiä (people) and puistohdr (parkhdr). The original photo was by the owner tagged with: HDR, nikond40, photomatix, france, french, paris, eiffel, tower, people, weekend, picnic, park, fun, tall, art, paint, sky, center, europe, european, EU, clouds, sunny. Some of the tags added by the players are the same as the ones the owner of the photo have added. This shows that the game can produce valuable tags.
6.3.3 The value of the tags to the participants

In the interview the players were asked if they would add the tags they produced to the photos, if the photos were their own. Two of the players answered clearly no. Gina and Daria felt that they could produce better tags themselves manually and Dan found the tags that he added were not good enough. Eve said she would add the tags, but not the ones produced during the insane game.

The ones, who provided the photos for the second group, were also asked the same question. Jenny expressed that she might add the tags and Ina answered that she would add the tags directly, because there is no harm in adding them. Chris said that he would add the tags for his personal use, but not if others could see them and search with them. He also added that he would like to get more tags for the photos from his travels, but not for the photos with his daughter, because of privacy issues.

The players were asked if they would pay for the tags that were produced during game play. From the first group, only Dan expressed that he would be ready to pay for the tags, if they first went through some kind of quality check. The reasons for
not wanting to pay for the tags were that they felt like they do not need tags at the moment or that they can add them themselves.

From the second group, Holly said she might pay for tags, if that meant she would get a lot of tags. Jenny answered that she does not need tags and Chris expressed that he finds paying for something in social media a bad practise. He also added that he could add tags himself for free. Ina said that she could pay a couple of Euros for the tags, if the payment was made easy.
7 Discussion and conclusions

The user experience in the game Phototag has been tested using usability testing methods. The tests showed a lot about the game, for example that the prototype has some serious playability problems, which affects the user experience. The user interface of the game is hard to use and not intuitive.

It was shown that the game has learnability, memorability and efficiency problems, where efficiency is one of the biggest. When the player is watching a photo, the photo needs to be closed before clicking on another photo and both photos are automatically closed after two photos have been seen. This creates a lot of unnecessary clicking when the player might already know which cards are a pair. The players that were less technically competent had the biggest problems with the user interface and easily got confused by it.

The tagging of photos also slowed the game down and players complained that it made the game feel mechanical. They also reported that tagging the photos made the game too easy and removed the challenge of remembering the positions of the cards. Even though the players found the adding of tags strenuous and time consuming, they did see how it helped to finish the game. This could be seen with many trying to play the game without tags, but returning to adding tags, because the game was much harder without them.

Even though some of the players got frustrated during the game, many still found the game somewhat enjoyable. The players that reported that they enjoy playing casual games a lot also enjoyed this game the most. These players also understood the idea of the game quickly and how to use the user interface. Holly and Ella both play a lot of casual games and they were the ones to report that they enjoyed the game. There was also a correlation for those that seldom play casual games. Cain and Gina, which from both groups liked the game the least, answered that they do not normally play these kinds of games. Gain and Gina also had a long average time to find pairs in the game. In Figure 13 can be seen that Cain had the longest time to find a pair in both the medium and hard level, and Gina had the longest time in the medium level. Gina also had a long average time in the hard level, even though Holly had a longer time.

The difference between the photos used in the game had a clear impact on the game. Photos that were taken by somebody close to the player, made the game much more interesting. Even though the photos made the game more enjoyable for the player,
the photos does not necessarily add any gaming value to the game and makes it harder for the players to concentrate on the gaming aspect.

Many of the players expressed that they would want to play the game on another platform, for example a tablet or mobile phone. Phototag is a more casual game, for which few use a computer to play anymore. Casual games are mostly played on mobile phones and tablets, because they are often accessible when there is free time. This would add additional complications to the adding of tags, because writing on a phone or tablet is not always easy.

During the test a lot of tags were created and most of the tags described the photos well. This shows that the game would work for tagging of photos, as long as bad tags are removed. The tags would need to go through some sort of quality checks, for example only the most common tags for a photo would be approved. The tags could also go through a "blacklist" to remove for example offensive words and prepositions. In languages like Finnish, where the prepositions are a part of the word, it adds another challenge to the quality checks of the tags.

The participants saw little value in the tags produced. Only one participant reported that she would be ready to pay something for tags produced during the game. Statistically, this is of course a too small test group to say that the tags would not have value to anyone.

For the future development of the game, the playability of the game needs to fixed. The reason for players not liking the game might only have been a reaction to bad playability. It might be worth considering to change the flow of the game from a normal memory game into a more suitable one for a computer game. The current flow is slow and hinders the players from quickly moving forward. The tagging of photos should also be made easier and more natural, so that the players do not see it as a hinder in the game.

The tests showed that the game can be used for crowdsourcing tagging of photos, because it produces a lot of tags when enough people play it and the tags goes through some quality checks. But in its current form the game does not work as an entertaining game and the presented fundamental problems should be fixed before developing the game further.
References


## Appendix 1. Questionnaire

<table>
<thead>
<tr>
<th></th>
<th>Strongly agree</th>
<th>Agree somewhat</th>
<th>Agree</th>
<th>Undecided</th>
<th>Disagree somewhat</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I enjoyed playing the game</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>2. I liked the game</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>3. I would play the game again</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>4. I would like to spend a lot more time playing this game</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>5. The game felt unfair</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>6. The game was too difficult</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>7. The game was too easy</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>8. I was frustrated at the end of the game</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>9. I was frustrated when playing the game</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>10. I got bored when playing the game</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>11. The rules of the game were easy to learn</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>12. I understood the rules of the game</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>13. The controls were easy to use</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>14. The controls responded as I expected</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td></td>
<td>Strongly agree</td>
<td>Agree</td>
<td>Agree somewhat</td>
<td>Undecided</td>
<td>Disagree somewhat</td>
<td>Disagree</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>---</td>
<td>----------------</td>
<td>-------</td>
<td>----------------</td>
<td>-----------</td>
<td>-------------------</td>
<td>----------</td>
<td>------------------</td>
</tr>
<tr>
<td>15. I knew all the actions that could be performed in the game</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>16. I knew how to win the game</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>17. I found the adding of the tags difficult</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>18. I found the adding of the tags frustrating</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>19. There were times that I was doing nothing in the game</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>20. It was easy to concentrate on the game</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>21. I felt involved in the game</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>22. I forgot the time when playing</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>23. I usually do not choose this type of games</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>24. I do not like this type of game</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>25. How much would you pay for the game</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>26. Additional comments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Strongly agree</td>
<td>Agree</td>
<td>Agree somewhat</td>
<td>Undecided</td>
<td>Disagree somewhat</td>
<td>Disagree</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>---</td>
<td>----------------</td>
<td>-------</td>
<td>----------------</td>
<td>-----------</td>
<td>------------------</td>
<td>----------</td>
<td>------------------</td>
</tr>
<tr>
<td>1. I play computer games more than once a week</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>2. I often play solitaire, Angry Birds or other similar small games</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>3. I take photos almost every week</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>4. I upload my photo to the internet, using for example: flickr, Facebook, google+ or personal blog</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>5. At least a few times a year I organize my photos in some way</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>6. I add tags or labels to my photos</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>7. I use some photo management software to handle my photos</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>8. Sex:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Age:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Occupation:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 2. Finnish translation of the questionnaire

<table>
<thead>
<tr>
<th>Question</th>
<th>Vahvasti samaa mieltä</th>
<th>Samaa mieltä</th>
<th>Jossain määrin samaa mieltä</th>
<th>En pysty sanomaan</th>
<th>Jossain määrin eri mieltä</th>
<th>Eri mieltä</th>
<th>Vahvasti eri mieltä</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Nautin pelin pelaamisesta</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>2. Pidin pelistä</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>3. Pelaisin peliä uudestaan</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>4. Haluaisin käyttää paljon aikaa tämän pelin pelaamiseen</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>5. Peli tuntui epäreilulta</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>6. Peli oli liian vaikea</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>7. Peli oli liian helppo</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>8. Olin turhautunut pelin lopussa</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>9. Olin turhautunut pelin aikana</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>10. Työlsistyin pelin aikana</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>11. Pelin säännöt oli helppo oppia</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>12. Ymmärsin pelin säännöt</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>13. Pelin kontrollit oli helppo käyttää</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>14. Kontrollit toimivat kuten odotin</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>Tarkastelutaulukko</td>
<td>Vahvasti samaa mieltä</td>
<td>Samaa mieltä</td>
<td>Jossain määrin samaa mieltä</td>
<td>En pysty sanomaan</td>
<td>Jossain määrin eri mieltä</td>
<td>Eri mieltä</td>
<td>Vahvasti eri mieltä</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------------</td>
<td>--------------</td>
<td>-----------------------------</td>
<td>-------------------</td>
<td>---------------------------</td>
<td>-----------</td>
<td>-------------------</td>
</tr>
<tr>
<td>15. Tiesin kaikki mahdolliset toiminnot pelissä</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>16. Tiesin miten voittaa pelin</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>17. Tagien lisääminen tuntui hankalalta</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>18. Tagien lisääminen tuntui turhaustavalta</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>19. Oli aikoja pelissä kun en tehnyt mitään</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>20. Oli helppoa keskittyä peliin</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>21. Tuntui kun olisin syventynyt peliin</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>22. Unohdin ajan kun pelasin peliä</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>23. Tavallisesti en pelaa tämän tyyppisiä pelejä</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>24. Tavallisesti en nauti tämän tyyppisistä peleistä</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
</tr>
<tr>
<td>25. Paljonko maksaisit tätä pelistä?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26. Muita kommentteja</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vahvasti samaa mieltä</td>
<td>Samaa mieltä</td>
<td>Jossain määrin samaa mieltä</td>
<td>En pysty sanomaan</td>
<td>Jossain määrin eri mieltä</td>
<td>Eri mieltä</td>
<td>Vahvasti eri mieltä</td>
</tr>
<tr>
<td>---</td>
<td>----------------------</td>
<td>--------------</td>
<td>-----------------------------</td>
<td>-------------------</td>
<td>--------------------------</td>
<td>-----------</td>
<td>-------------------</td>
</tr>
<tr>
<td>1.</td>
<td>Pelaan tietokonepelejä enemmän kun kerran viikossa</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>2.</td>
<td>Pelaan usein pasianssia, Angry Birds tai sen tyyppisiä pela</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>3.</td>
<td>Otan valokuvia melkein joka viikko</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>4.</td>
<td>Lataan valokuviani netiin, käyttäen esim: flickr, Facebook, google+ tai henkilökohtainen blogi</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>5.</td>
<td>Ainakin pari kertaa vuodessa järjestelen kuviani jollain tavoin.</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>6.</td>
<td>Lisään tageja tai tunnuksia kuviani.</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>7.</td>
<td>Käytän jonkinlaista kuvakäsittelyohjelmaa</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>8.</td>
<td>Sukupuoli:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>Ikä:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>Ammatti:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 3. Interview questions concerning the tags

1. Do you feel that the tags (these words) represent these pictures?

2. If you were to do a search with these tags, do you think you would find these photos?

3. If these photos were your own photos, would you then want to add these tags to the photos? Which ones not?

4a. If not, why not? 4b. If yes, would you be ready to pay for these kind of tags? How much?

5. If you could get these kind of tags for free to your photos, would you be ready to make your photos public? (Why not?)

6. You have added these tags. Would you want the owners of the photos to get these tags? (Why not?)

7. If you got paid for the tags that you produce when playing, would you then be ready to play a lot more?
Appendix 4. Finnish translation of interview questions

1. Kuvaako, sinun mielestä, nämä tagit (nämä sanat) näitä kuvia?

2. Jos teksit haun noilla sanoilla, luuletko että löytäisit tota kuvaa?

3. Jos nämä kuvat olisi sun omia kuvia, olisitko valmis lisäämään nämä tagit kuviiin? Mitkä ei?

4a. Jos ei, miksi ei? 4b. Jos kyllä, niin olisitko valmis maksamaan tämän tyyppisistä tageista? Paljonko?

5. Jos saisit tämän tyyppisiä tageja sun kuviiin, olisitko valmis tarjoamaan sinun kuvat pelin käyttöön? (Miksi ei?)

6. Olet lisännyt tämä tagit. Olisitko valmis antamaan nämä tagit kuvien omistajille? (Miksi ei?)

7. Olisitko valmis pelaamaan enemmän jos saisit maksua tageista, jotka tuotat pelin aikana?
Appendix 5. Interview questions for photo provider

1. Do you feel that the tags (these words) represent these pictures?

2. If you were to do a search with these tags, do you think you would find these photos?

3. Would you want to add these tags to your photos? Which ones not?

4a. If not, why not? 4b. If yes, would you be ready to pay for these kind of tags? How much?

5. If you could get these kind of tags for free to your photos, would you be ready to make your photos public? (Why not?)
Appendix 6. Finnish interview questions for photo provider

1. Kuvaako, sinun mielestä, nämä tagit(nämä sanat) näitä kuvia?

2. Jos tekisit haun noilla sanoilla, luuletko että löytäisit tota kuvaa?


4a. Jos ei, miksi ei? 4b. Jos kyllä, niin olisitko valmis maksamaan tämän tyyppisistä tageista? Paljonko?

5. Jos saisit tämän tyyppisiä tageja sun kuvien, olisitko valmis tarjoamaan sinun kuvat pelin käyttöön? (miksi ei?)
Appendix 7. Test layout

- Greet the participant and let him/her settle down.

- Tell them shortly about the game: "We are testing a game for a client. The game is still under development and they want to know how people feel about the game idea. This is a version of the game Concentrations, where you have some amount of cards and try to find two of a pair. I will explain more precisely in a moment."

- Explain the layout of the test:

1. "First I will ask of you to play an easy game, where I will explain the rules and controls of the game"

2. "After that game, you will be asked to play a second game. For this you will be asked to think aloud. I will explain how that works before we start."

3. "If you complete the second game and still want to continue you can play a hard game. Depending on how much time is left after that, you can play an insane game. All the time preferably thinking aloud"

4. "After the games, I will ask you to fill in a questionnaire with questions about your feelings and perceptions about the game."

5. "You will also be asked to evaluate some data that will be gathered when playing."

6. "After that we are ready and you will be given your compensation for your time."

- The participant will be told about the camera and recording: "The camera will record your actions and your sound. This material will not be shown to anybody outside this project and the material will be destroyed after it has been analyzed."

- The participant is allowed to ask questions: "You can ask me questions if you want, but I might not be allowed or able to answer all of them."

- The participant will be explained how to think aloud: "When you are playing, verbalize your thoughts. Say out loud the things that go through your head. Your feelings and thought patterns are of highest value. A bit like talking aloud to yourself. You can try to imagine that you are alone. I will remind you to keep on
talking if you fall silent. But if it feels to weird or hard to talk at the same time as playing, just tell me.

- Show the participant how to think aloud by doing a simple task (for example plugging in the mouse) and verbalizing my frustration over that the usb-plug is never the right orientation on the first try.

- Ask the participant if everything is understood and if he/she has any questions before starting. Also make it clear that the participant can terminate the test at any time.

- Start the small game and show how to select cards, add notes, search for tags and what the magnifying glass-icon does. Also show the keyboard-shortcuts present.

- When the small game is played, the medium one can be started. Do not disturb the participant when playing, only remind to think aloud, if he/she forgets.

- Large board.

- Possible insane board.

- Questionnaire about the game and demographic questions.

- Show the tags produced.

- Ask about the tags.

- Give the participant the compensation and thank him/her for his/her time.