Data Gathering Methods for Evaluating Playability of Pervasive Mobile Games

Hannamari Saarenpää

University of Tampere
Department of Computer Sciences
Interactive Technology
Master’s Thesis
Supervisor: Saila Ovaska
March 2008
This thesis presents new game genre, pervasive games, that mix reality and virtual reality by using context in their gameplay. In this thesis I discuss the data gathering methods that are commonly used when evaluating games, and how adequate they are for pervasive game evaluation. I use two case studies, *Coup* and *Insectopia*, as well as thirteen other pervasive game evaluation studies to demonstrate what kind of information can be collected with different data gathering methods, and discuss their benefits and disadvantages for the evaluation.

The results of the example game evaluations show that it is seldom sufficient to use just one method when gathering data for evaluation. Instead, it is recommendable to use at least two methods, one for quantitative and one for qualitative data. Evaluation should be part of the whole game development process from the game idea to the final game product. Different data gathering methods proved to be valuable in different phases of the game development process. Whereas game scenarios and target group discussions can be good for getting information about the game idea and the acceptance of the game, questionnaires, interviews, and log files are good sources to get knowledge about the final game product. The results also show that it is hard to evaluate pervasive aspects of a game since even the evaluator may not know the issues to concentrate on when planning an evaluation. Also it is hard for participants to imagine a pervasive game without any previous experiences of them. Improved practices and new methods for evaluating pervasive games are needed.

**Keywords and -phrases:** Pervasive games, Playability, Evaluation, Data gathering methods, Player experience, Mobile games
Acknowledgements

I would like to thank the following people for their help in making this happen. My supervisor Saila Ovaska from the University of Tampere, who continued to help and support me through the whole process. She gave me a lot of valuable feedback and pushed me forward with this. Elina Ollila and Jussi Holopainen from Nokia Research Center for providing me this topic and guiding me through this process. I also have had a lot of inspiring discussions with Hannu Korhonen from Nokia Research Center. IPerG project has also supported my work by providing access to their papers and publications.
Content

1. Introduction ............................................................................................................ 1
2. Pervasive games ..................................................................................................... 3
   2.1. What are pervasive games? ......................................................................... 3
   2.2. Breaking the magic circle ............................................................................. 5
       2.2.1. Spatial expansion ............................................................................... 6
       2.2.2. Temporal expansion .......................................................................... 7
       2.2.3. Social expansion ................................................................................ 7
   2.3. Pervasive game sub-genres ......................................................................... 8
       2.3.1. Alternate reality games (ARGs) ......................................................... 9
       2.3.2. Pervasive live action role playing games (pervasive larps) ....... 9
       2.3.3. Massively multiplayer mobile games ............................................ 10
       2.3.4. Online-onstreet games ...................................................................... 15
       2.3.5. Proximity games ............................................................................... 16
       2.3.6. Event games ..................................................................................... 19
       2.3.7. Crossmedia games .......................................................................... 19
       2.3.8. Summary .......................................................................................... 21
   2.4. Pervasive mobile games ............................................................................. 24
   2.5. Case studies .................................................................................................. 25
       2.5.1. Insectopia .......................................................................................... 25
       2.5.2. Coup .................................................................................................. 27
   2.6. Research questions ...................................................................................... 29
3. Game evaluation .................................................................................................. 30
   3.1. Why to evaluate games? ............................................................................. 30
   3.2. Components of playability ....................................................................... 31
   3.3. Playability vs. usability ............................................................................. 36
   3.4. The game evaluation process .................................................................... 39
   3.5. Selecting and recruiting participants ........................................................ 40
   3.6. The challenges of game evaluation ........................................................... 43
       3.6.1. Getting enough participants to play the game ......................... 43
       3.6.2. Ethical issues and privacy ............................................................. 44
       3.6.3. Technical problems ......................................................................... 45
4. Literature overview to data gathering methods ............................................. 46
   4.1. Heuristics .................................................................................................. 46
   4.2. Questionnaires .......................................................................................... 50
   4.3. Interviews .................................................................................................. 51
   4.4. Gameplay diaries ....................................................................................... 52
   4.5. Observation ............................................................................................... 53
5. Analysis of the results – Different methods for evaluating playability ...56

5.1. What playability problems can be found using different methods? ...56

5.1.1. Heuristics.......................................................................................... 58
5.1.2. Questionnaires.................................................................................. 59
5.1.3. Interviews.......................................................................................... 60
5.1.4. Gameplay diaries............................................................................... 60
5.1.5. Field observation............................................................................... 61
5.1.6. Log data............................................................................................ 63
5.1.7. Focus groups.................................................................................... 64

5.2. Methods in different evaluation phases............................................. 65
5.2.1. Game scenario.................................................................................. 68
5.2.2. Paper prototype............................................................................... 69
5.2.3. Beta version...................................................................................... 70
5.2.4. Final game product........................................................................... 71

5.3. The validity of information gained using different methods.......... 71

6. Discussion............................................................................................................. 75

6.1. How to evaluate playability of pervasive games? ......................... 75

6.1.1. Gameplay.......................................................................................... 75
6.1.2. Usability............................................................................................ 76
6.1.3. Pervasiveness.................................................................................... 76
6.1.4. Mobility............................................................................................. 77
6.1.5. Multiplayer aspects.......................................................................... 78

6.2. Choosing data gathering methods....................................................... 79

6.3. Things to remember when planning an evaluation..........................81

6.4. Limitations of this study and further research.................................83

7. Conclusion............................................................................................................ 84

References...................................................................................................................... 86

Appendix A The questionnaire............................................................................ 95
Appendix B The interview outline .................................................................... 101
Appendix C The pervasive game flow heuristics [Jegers, 2006]..................103
Appendix D The usability heuristics [Nielsen, 1994, p. 30].........................105
Appendix E Events that were recorded in Coup............................................ 106
1. Introduction

How would you feel about being an insect when you are walking around the streets and someone is trying to catch you? Or the opposite, how would you feel about being the collector and following people and trying to "catch them"? This is the reality in a mobile game Insectopia which is one of the new types of games, pervasive games. In Insectopia each Bluetooth ID is an insect and a player is trying to catch them when they are in certain proximity. Each Bluetooth connection is an individual insect, so in a way if you are carrying around a device which has a Bluetooth connection open you are an insect.

Pervasive games combine virtual game worlds with elements of the real world. One way to define them is to say that they are “context-aware games combining physical presence with virtual experience” [Montola et al., 2005]. The main difference between traditional games and pervasive games is that pervasive games try to break the magic circle of games. Magic circle is the imaginary boundary between the game world and reality where game rules apply [Salen and Zimmerman, 2003, p. 95-96]. Pervasive games try to blend games into our everyday lives by breaking this magic circle either spatially, temporally, or socially.

Pervasive games differ a lot from traditional games, and they require a new playing style. Whereas in traditional games the playing time and place are known, in pervasive games they are uncertain and changing. Pervasive games can even involve non-players in the game without them knowing of it. This may cause ethical concerns.

The special nature of pervasive games makes the evaluation of them more challenging than other types of games. In many situations lab testing or direct observation is not possible because the essential part of the game is how it blends with players’ everyday lives. Also many pervasive games are location-based which means that the location of a player affects the game. For example, it is possible that the game character in the game world moves according to the movements of the player in the real world. This means that pervasive games cannot be evaluated in the same manner as other games.
The focus in game evaluation is playability. Playability is part of a player experience and it always includes usability and gameplay issues. In addition, it includes game specific issues, for example multiplayer or mobile issues. In this thesis the data gathering methods that can be used for evaluating the playability of pervasive mobile games are in focus. The reason why this thesis concentrated solely on pervasive mobile games is that pervasive games are usually played with a mobile device. Mobile devices and especially mobile phones are pervasive by nature [Niewdorp, 2005] so they support pervasive games well.

The main research questions are: what makes pervasive games different from traditional games, how does it affect evaluation, and what are the data gathering methods that are suitable for evaluating playability of pervasive games in different game development phases? There are many studies about games, fun, and even game experience (e.g. [Järvinen et al., 2002; Sweetser and Wyeth, 2005]) but still only few studies concentrate solely on game evaluation. At the moment there are no proper studies about how pervasive games should be evaluated in order to get valid information about the aspects that are characteristic of pervasive games and affect playability. To learn about the playability evaluation of pervasive games we evaluated two pervasive mobile games, Coup and Insectopia, and in addition to this, studied papers of thirteen other pervasive games. Coup and Insectopia were developed as part of the Integrated Project on Pervasive Gaming (IPerG, http://iperg.sics.se).

The hypothesis was that playability evaluation of pervasive games would require new methods and practices to be used than what have been used in game evaluations so far. This proved to be true. It is possible to evaluate the playability of pervasive games with existing data gathering methods but they need to be modified to fit the needs of pervasive games. It was shown that in pervasive game evaluation it is crucial to evaluate the game throughout the game development process, and use several data gathering methods.

The structure of this thesis is following. Chapter 2 introduces the theoretical background of pervasive games and the games that are used as examples including the case study games, Coup and Insectopia. Chapter 3 is about game evaluation process and playability. Chapter 4 describes the data gathering methods that are used in game evaluation. In Chapter 5 the results are presented; what methods were used in the example game evaluations and how they worked. The last two chapters are for discussion and conclusions. The results are discussed in the context of playability and also some guidelines are given to future evaluators.
2. Pervasive games

In 1997 a psychological thriller film called *The Game* was published. The idea of the film was that a wealthy businessman got prepaid access to a weird game that in many ways integrated into his life, as a gift for his birthday. When the story goes further and the businessman is deeper in the game, he cannot really tell anymore what is real and what is part of the game. Instead he starts to think about conspiracy against him. This film is an early and intensified example of one kind of pervasive game, alternate reality game, where the boundary of the reality and the game is blurred.

Pervasive games are heavily based on ubiquitous computing. Ubiquitous computing means that computers become part of the everyday objects and integrate into the environment. This term is close to the term pervasive computing (everywhere present) which has given the name for pervasive games. A class of ubiquitous games exists as well, but what separates them from pervasive games is that they usually are more part of the real world. Whereas in pervasive games real world elements are brought to the virtual game world, ubiquitous games bring game features to the real world.

Pervasive games actually would not be games if Huizinga’s [1955] definition of play and games was followed precisely. Pervasive games do not happen at a specific time or in a specific place. They also differ from traditional games because it is not always clear who are involved in a game. It is even possible that participants themselves do not know that they are being part of a game.

In this chapter pervasive games are defined and the expansion styles of games, the different possible ways that they can expand the traditional magic circle are described.

2.1. What are pervasive games?
Games that can be defined as pervasive games can vary a lot from one another. The first pervasive games were published in around the year 2000 (*The Nokia Game* 1999, *Can You See Me Now?* 2001, *The Beast* 2001, and *Botfighters* 2001; for
further information see Links to Games, p. 93-94). One of the first pervasive games was The Beast (also known as The A.I. Game). It started from Steven Spielberg’s movie Artificial Intelligence: A.I. in 2001 where credit was given at the end of the film to a fictional person “Jeanine Salla, Sentient Machine Therapist”. Some of the people that spotted this fictional person from the end of the movie ‘googled’ her and found many websites that were all dated in the year 2142 (the same year that was in the movie) and a whole game started (e.g. [McGonigal, 2003]). The point of the game was to find out who was the murderer of a fictional person named Evan Chan. The players collaborated with each other online where they had formed an online group Cloudmakers. More than 7000 people were involved in this group [McGonigal, 2003]. The game was a huge success even though the players had to find out themselves what to do next. The players finally solved the mystery one year later, after solving many puzzles and tasks that were only possible to solve by collaborating with other players. Also later there have been collaboration attempts in Cloudmakers to try to solve even actual crimes like the terrorist attack on the 11th of September 2001. This is a proof of how strong sense of commitment and collaboration people got from The Beast game. This kind of collaborative problem solving is one possible game element in pervasive games.

The facts that pervasive games can differ so much from one another and that they are such a new phenomenon are probably the reasons why there is no one specific definition for pervasive games. The researchers do not have a joint understanding of the definition. One definition that has been used for example in IPerG is from Montola [2005, p. 3]; a game “[...] that has one or more salient features that expand the contractual magic circle of play socially, spatially or temporally.” Some of the researchers in this area claim that the definition should be more based on technology, which is in many cases important in pervasive games. Also one might disagree with the use of magic circle because it would mean that games really exist in ‘a little box’ that could then be expanded (see e.g. [Brown, 2007, p. 49]). Still I will use Montola’s definition in this thesis because it is probably the most used definition.

Distinctive feature in all pervasive games is that they break or at least expand the magic circle. Nieuwdorp [2005] describes that pervasive games waver between reality and fantasy. She also says that whoever plays a game must accept the transformation between the real world and the game world. This means that pervasive games are context-aware. They include real-life elements into gameplay. For example non-players could be used as a game source; real-life weather conditions could be used to set the weather in a game, the time could go on in a game according to the real time, and so on. The
possibilities are almost endless. This is a bit contrary to what was said earlier that a pervasive game is always present for a player. Still they basically mean the same thing. Pervasive games have the ability to adapt to the player’s everyday life [Montola et al., 2006]. Players, who play pervasive games, have to learn to take games in an unusual way. The rules that they have used to act according to in their real life can be turned upside down in pervasive games. Things are not always what they seem to be.

2.2. Breaking the magic circle

The term magic circle was first introduced in Johan Huizinga’s famous book Homo ludens [1955] and after that many other researchers have used it to describe “a certain play-ground within which special rules obtain” [p. 10]. This term can be used to describe all playgrounds, temporary magical worlds. Magic circle refers to the magical game world within the real world. It is a boundary between reality and make-believe [Adams and Rollings, 2006, p. 6-8]. This is presented in Figure 1. Pervasive games are special by nature in that they aspire to break this magic circle of games.

The magic circle has always some limitations of time and space [Klabbers, 2006, p. 3]. Play happens in a certain place at a certain time. It also requires that all the participants accept this magic circle of play and its rules and limitations and treat it such [Klabbers, 2006, p. 3; Salen and Zimmerman, 2003, p. 97-98]. This kind of state of mind can be called lusory attitude [Salen and Zimmerman, 2003]. The term comes from the word ludology that means the field of game studies that see games foremost as systems. Magic circle is the frame of a game and it is a closed space where the game takes place.

*The magic circle can define a powerful space, investing its authority in the actions of players and creating new and complex meanings that are possible in the space of play. But it is also remarkably fragile as well, requiring constant maintenance to keep it intact.*

[Salen and Zimmerman, 2003, p. 98]
Within the magic circle, the players agree to attach a temporary, artificial significance to situations and events in the game. The magic circle comes into existence when the players join the game, in effect, when they agree to abide by the rules. It disappears again when they abandon the game or the game ends.

[Adams and Rollings, 2006, p. 8]

Many researchers agree with the existence of magic circle, or pretended reality as Adams and Rollings [2006] call it. Koster [2005, p. 38] disagrees with many other researchers about the definition and argues that magic circle is too limiting for a game to be long lasting. He says that a game needs to integrate with human psychology, physics, and so on in order to be truly challenging and replayable (which means that a game is still enjoyable and exciting after the first play session). This is why he does not think that games exist in their own magic circle but rather are part of reality. This kind of approach is very close to what pervasive games are after. Games should be part of reality; they should blend in players’ everyday lives but not disturb their daily activities. Magic circle is quite limiting and it also narrows the play world, but still I think it is useful for representing the current situation of games. Its function is rather to illustrate the situation than to be precisely accurate.

Pervasive games can break the magic circle spatially, temporally, or socially. Games that are spatially (or socially) expanded are easier to define as pervasive games since they are so different from traditional games. Instead, the games that expand the magic circle temporally can be quite difficult to define as pervasive games. In order for a game to be pervasive it has to have at least one of these expansion styles but some games can have all three. Next these expansions are described more closely.

2.2.1. **Spatial expansion**
Spatial expansion means that a game is basically playable anywhere. It means that the game does not depend on certain place, equipment or organisation; in other words the game should be decontextable [Linderoth et al., 2005]. A game has a spatial expansion when it is taken out of “the ordinary playground”. Playground means the physical area where a game takes place. It is usually more or less fixed. Pervasive games that break the magic circle spatially expand this playground of a game. For example the whole city can be a gaming area (like in *Botfighters* and *Songs of North*). Montola [2005] has defined spatial expansion: “Spatial expansion indicates that the socially constructed location of the game is unclear or unlimited.” Still it is not enough that a game is playable
everywhere in order to be pervasive. If that would be the case, then all mobile games and hand-console games would be pervasive. A pervasive game that is spatially expanded has to use space as its game mechanic; it has to be built into its rules. For example in *Insectopia*, one of the case study games, a player has to catch insects that are actually unique Bluetooth IDs by moving around in the physical environment.

### 2.2.2. Temporal expansion

Temporal expansion means that a game is not bound in any specific time. Players can join the game and leave the game whenever they feel like it without destroying the game. Linderoth *et al.* [2005] divided temporal expansion into four different subcategories: negotiable openings and endings, negotiable game time, activity blending, and dormancy (which means that the game can be played passively until the game calls for the player’s attention).

The game sessions of pervasive games are often mixed with ordinary life. The problem with this is that a “game might require attention at worst possible times or require generally too much attention” [Montola, 2005]. Temporal expansion demands extra attention from game developers because all players may not be present in the game at the same time. This can cause that there are times in the game when only one player is present. If the game is a multiplayer game where the gameplay is based on the competition or collaboration of the players, these ‘alone’ times have to be taken into consideration. This usually happens by providing some tasks that a player can do alone in the game.

### 2.2.3. Social expansion

Social expansion can mean more than just one thing. Linderoth *et al.* [2005] have divided social expansion styles into four groups: actor detachment, identity detachment, accessibility, and non-player relations. Actor detachment means the ability of a game to be playable by divergent actors. In order to have high actor detachment it has to be playable for both men and women, children and adults, etc. Identity detachment means the game’s ability to attract different people despite of their age, gender, or background. Accessibility refers to how easy it is to access the game. There can be either technological or social obstacles in getting into the game. Non-player relations [Linderoth *et al.*, 2005] refer to how a game exploits outsiders in a game. This is a very specific feature in pervasive games that is not used in other types of games. The possibility to use non-players as game elements either so that they are aware of it (*Hot Potato*) or not (*Insectopia*) is unique and can cause ethical concerns. Niemi *et al.* [2005] have studied attitudes towards games that in some way affect non-players. Half of
the adult participants (aged 28-46) thought that it is too much if a game registers
or records some information without permission whereas young (aged 15-16)
participants were much more accepting. This would suggest that pervasive
.games that expand the magic circle socially are more suitable for young people
who are more open-minded towards them.

2.3. Pervasive game sub-genres
Pervasive games are very heterogeneous. Figure 2 lists pervasive game sub-genres that Montola et al. [2006] have proposed. This is not the only
classification that there is (see e.g. [Magerkurth et al., 2005]) but it is the most
suitable for the purposes of this thesis. It should be noted that one game can be
part of more than just one genre.

I will introduce fifteen pervasive games that I have studied in more detail. I
refer to these fifteen games later on as the example games. All the games are
described under the sub-genre where they fit the best. The descriptions are
based on my own experiences and publications that are written about the
games. My focus in this paper is pervasive mobile games, so all these example
games use mobile devices in their gameplay. Table 1 lists the example games
and their current webpages. In this thesis also some other pervasive games are
mentioned, but either they are not mobile games or there are no studies of
them, so they are not included into these example games. Further links to
games appear after the list of references on pages 93-94.
Table 1 The example games and their current webpages (checked 29.11.2007).

<table>
<thead>
<tr>
<th>The game</th>
<th>The webpage of the game</th>
<th>The sub-genre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Botfighters</td>
<td><a href="http://www.gamespot.com/mobile/action/botfighter/index.html">http://www.gamespot.com/mobile/action/botfighter/index.html</a></td>
<td>Massively multiplayer mobile game</td>
</tr>
<tr>
<td>Coup</td>
<td><a href="http://coupgame.org">http://coupgame.org</a></td>
<td>Massively multiplayer mobile game</td>
</tr>
<tr>
<td>Day of the Figurines</td>
<td><a href="http://www.dayofthefigurines.co.uk/">http://www.dayofthefigurines.co.uk/</a></td>
<td>Massively multiplayer mobile game</td>
</tr>
<tr>
<td>Epidemic Menace</td>
<td><a href="http://iperg.fit.fraunhofer.de/">http://iperg.fit.fraunhofer.de/</a></td>
<td>Crossmedia game</td>
</tr>
<tr>
<td>Feeding Yoshi</td>
<td><a href="http://www.yoshigame.com/">http://www.yoshigame.com/</a></td>
<td>Proximity game</td>
</tr>
<tr>
<td>Garden of Earthly Delights</td>
<td><a href="http://mmroserver.hgo.se">http://mmroserver.hgo.se</a></td>
<td>Crossmedia game</td>
</tr>
<tr>
<td>HotPotato</td>
<td><a href="http://www.sics.se/hotpotato/index.html">http://www.sics.se/hotpotato/index.html</a></td>
<td>Proximity game</td>
</tr>
<tr>
<td>Insectopia</td>
<td><a href="http://www.insectopia.org">http://www.insectopia.org</a></td>
<td>Proximity game</td>
</tr>
<tr>
<td>Pirates</td>
<td>-</td>
<td>Massively multiplayer mobile game</td>
</tr>
<tr>
<td>Savannah</td>
<td><a href="http://www.futurelab.org.uk/projects/savannah">http://www.futurelab.org.uk/projects/savannah</a></td>
<td>Massively multiplayer mobile game</td>
</tr>
<tr>
<td>Songs of North</td>
<td><a href="http://www.uta.fi/hyper/projektit/mogame/laulut_esite.pdf">http://www.uta.fi/hyper/projektit/mogame/laulut_esite.pdf</a></td>
<td>Massively multiplayer mobile game</td>
</tr>
<tr>
<td>Spy Blob</td>
<td>-</td>
<td>Proximity game</td>
</tr>
<tr>
<td>Uncle Roy All Around You</td>
<td><a href="http://www.blasttheory.co.uk/bt/work_uncleroy.html">http://www.blasttheory.co.uk/bt/work_uncleroy.html</a></td>
<td>Online-onstreet game</td>
</tr>
<tr>
<td>Your Story</td>
<td>-</td>
<td>Proximity game</td>
</tr>
<tr>
<td>Yum Yum Sheep</td>
<td>-</td>
<td>Proximity game</td>
</tr>
</tbody>
</table>

2.3.1. Alternate reality games (ARGs)

Alternate reality games usually involve hundreds of players to solve some puzzles [Montola et al., 2006]. In these games the essential part of gameplay is problem solving that can happen either in real-life, in the Internet, or using other media sources. ARGs are not tied to one media source. Also quite characteristic for ARGs is that they do not have chronologically coherent narrative. People proceed in a game finding new clues and storylines in quite random order, depending on what their focus is. Many times ARGs rely on ‘this is not a game’ (TINAG) aesthetic (see e.g. [McGonigal, 2003]) which means that a game does not behave like a game, and no one states that it is a game when it starts [Montola et al., 2006]. This also means that ARGs do not have any fixed rules or overly designed gameplay but players create these themselves. Examples of ARGs are The Beast, The Nokia Game, and I Love Bees which all were made to support a marketing campaign. These games are not introduced in this paper in more detail because they are not mobile games. For that same reason they were not presented in Table 1.

2.3.2. Pervasive live action role playing games (pervasive larps)

Pervasive live action role-playing games expand the boundaries of traditional larps in some way. For example the IPerG demonstration Momentum (e.g.
[Jonsson et al., 2007]) expanded the magic circle temporally by lasting five weeks. Where in traditional larps the game is limited to its players, pervasive larps can also interact with non-players as well as any object from the surrounding world. In another pervasive larp Prosopopeia also the ‘this is not a game’ aesthetic that is common to ARGs was used. Prosopopeia (e.g. [Jonsson et al., 2006]) was a ghost story where all the players were acting real but now deceased people that were all friends of a woman who was haunting. Together these players had to find out why the poor woman had not had peace but instead was captured between life and death.

Pervasive larps are quite complicated to arrange and usually they need many game masters to guide the players through the game. They also usually bring up many ethical considerations. I will not discuss pervasive larps in this thesis because they are not mobile games.

2.3.3. Massively multiplayer mobile games
Massively multiplayer mobile games typically have a persistent game world in a mobile phone [Montola et al., 2006]. Basic mechanic can be for example moving around in the physical space of the game world and interacting with both virtual objects and other players using a mobile phone.

Some of the games presented under this category are location-based, meaning that the location of a player affects the game. These games could have been presented also under the proximity category, but since the proximity is not the core element in these games they are included into this category.

Coup
Coup (Figure 3) is one of my case study games. It is a multiplayer game of high chivalry, quarrelling aristocracy, and gruesome backstabbing [Coup, 2006]. The main goal is to stay as high up in the feudal hierarchy as long as possible and at the same time gain renown and gold in tournaments. The players themselves form the hierarchy; if you invite new players to the game you will automatically be their lord, at least for a while. In order to climb up the hierarchy you can either try to overthrow your lord by force or try to persuade the higher-ranking players to move you up in the hierarchy [Coup, 2006]. The most coveted position, the Emperor, has ultimate power over the kingdom. Unfortunately, Emperors usually do not stay as Emperors for long as all the other noblemen try to get that position.
Communication between players happens by sending in-game messages. Also the communication between the game and players happens via messages.

The length of the game depends on the season length that is chosen by the person who starts the game. Season length means the actual time that it takes for season (spring, summer, fall, and winter) to change in the game. This has a huge impact on the gameplay. If the season length is short, the game lasts only a couple of hours and then it of course requires more attention during that period of time. When the season length is long, the game can last several days. Players have more freedom to choose when they want to play the game, because all the actions happen so slowly in the game.

**Day of the Figurines**

*Day of the Figurines* (Figure 4) is partly a board game, partly a secret society. The game is set in a fictional town that is littered, dark and underpinned with steady decay.
The game duration is 24 days and each day represents one hour in the life of a small English town in the board. The whole game is constructed around sending and receiving text messages. All the information that a player gets is received via text messages and also all the actions that a player wants to make happen by sending a text message. The game requires more effort than some other pervasive games [Crabtree et al., 2007].

1. The game board is always situated in some public place like a gallery. Players must first visit that place to see the game board.

2. Players must then select a figurine each to represent them in the game world.

3. Finally, players must answer some questions that define the characteristics of their figurine, and give the figurine a name. After giving this information to the game operator the players get an instruction card and a map of the fictional city.

**Savannah**

*Savannah* (Figure 5) is a collaborative location-based game for children to help them learn about the ecology of the African savannah. In the game all players are lions that live in a pack of six meaning that children play the game in groups of six [Benford et al., 2005b].

The idea of the game is that the players have to explore the imaginary savannah and discover essential resources like food that lions need to survive. The game field is an empty school playing field, which transforms into savannah in the interface of personal digital assistants (PDA, i.e. a handheld computer) that every player has [Benford et al., 2005b].
Pirates!

*Pirates!* [Björk et al., 2001] is a multiplayer game where each player is a captain of a ship. Players need to solve missions, explore new islands in search for trading goods, and fight against other players in sea battles. All the islands are unique, they have different terrain, but also the treasures and threats are different. The goal is to explore islands and complete missions that the viceroy has set up. If a mission is successfully completed a player gets an award that will get her higher in the ranking list, but she also gets a better ship.

When a player joins the game for the first time she is given a newbie task, which means that while a player fulfils this task she is safe from all the dangers of the game [Björk et al., 2001]. The reason for this is to encourage the player to continue playing the game and give her time to get to know the game better.

*Pirates!* is location-based: the players’ physical presence affects the game [Björk et al., 2001]. Sailing in the game is equivalent to walking in real life between different places. Proximity is needed when a player wants to have a battle with another player in the game. In order to do this, a player has to approach another player in the real world. In order to avoid conflict a player has to get away from the other player’s proximity. All the moving in the game is done by actually moving in real life. The players’ position in reality is in relation to their position in the virtual game world.

Songs of North

*Songs of North* (Figure 6) is a location aware mixed reality game that was designed to demonstrate possible technical and game design solutions for pervasive games in the Hypermedia Laboratory project called Wireless Gaming Solutions for Future (MOGAME) [Lankoski et al., 2004].

In this game the player is a shaman, who can contact with the spirit world by using a shaman drum. The spirit world is invisible but the player can hear sounds from there. Also the bone at the shaman drum represents things in the spirit world. The drum also acts as a map for a player and it should be projected over a real map. The player who wants to move in the game world has to move in the real world. The game itself is constructed around tasks that can be solved
alone or by collaborating with other players. Collaboration is very strongly encouraged and some tasks are almost impossible to solve without collaborating with some other player(s).

**Botfighters**

*Botfighters* is a location-based game that was launched in April 2001 in Sweden by a Swedish company named It’s Alive!. After that it has also been launched and played in Finland, Ireland, China, and Russia [Söderlund, 2005]. The goal in this game is quite simple, to search and destroy bots that are actually other players. By doing this the player earns credits and gets higher in the high-score list. The introduction text to the game that was presented in the *Botfighters* webpage leads players to the future which is not that bright and where players have to fight in order to sustain their freedom.

> The year is 2105. The world-spanning Global Nation controls 99\% of our planet’s resources. It is a sprawling bureaucracy, bloated and corrupt. The bureaucrats wield total power over 17 billion people, and only a few dare to oppose.

But still a war is being waged; a war where rebels fight for the freedom to control their own lives and where corporate loyalists strive to uphold the system that once saved our planet from extinction. You start as a newly graduated bot pilot, and then the action begins as you join the ferocious battles of the botfighters. [Botfighters, 2005]

---

Figure 7 Pictures from *Botfighters* [Söderlund, 2005].
Botfighters uses a mobile phone to present the location of bots and battles. A web page is used to build and update player’s robot (Figure 7). The overall game mechanism is based on text messages that are used for executing every action in the game (searching for bots, targeting, and shooting) [Sotamaa, 2002]. Botfighters was one of the first commercial pervasive mobile games that existed and it was a huge success story. In 2005 it had about 40 000 registered players worldwide [Söderlund, 2005].

2.3.4. Online-onstreet games
Online-onstreet games use both the online access to virtual environments and a physical space [Montola et al., 2006]. One of the first pervasive games Can You See Me Now was an online-onstreet game as was its follower Uncle Roy All around You.

Typically in these games there are two quite separate player roles. Some players are street players who act according to the instructions of online players.

Uncle Roy All around You
Uncle Roy All around You (Figure 8) is a game where some players have to walk around the city to find a shadowy and mysterious character and other players monitor their success from an office [Benford et al., 2006b; Magerkurth et al., 2005].

![Figure 8 Images of a street player and her view (left), and an online player’s views (right) [Uncle Roy All around You, 2003a].](image)

All the street players have a PDA and their task is to follow textual clues that are based on their current location [Uncle Roy All around You, 2003b]. Online players can follow their progress and communicate with them via text messages. Street players have to perform some tasks like retrieve a postcard, get into a waiting vehicle, and so on. There is not a big price waiting in the end and this game is not about competing against other players. Instead, the goal of this game is to create new experiences to the players.
2.3.5. Proximity games

Proximity games have core game mechanic somehow based on proximity using Bluetooth or radio frequency identification (RFID) technologies [Montola et al., 2006]. It is estimated that proximity games are going to become more common when mobile phones get better and more people have Bluetooth connection in their mobile phones [Montola et al., 2006]. Games like Hot Potato, Yum Yum Sheep and Insectopia [Niemi et al., 2005] are based on open Bluetooth connections of non-players. The games that are presented under this category use proximity as an essential game mechanic.

Insectopia

"Insectopia is a new kind of cell phone game where the real world spills into the game world" [Insectopia, 2006]. The game (Figure 9) is a mobile phone game where the goal is to search and catch as many insects (unique Bluetooth IDs) as possible. Insectopia is the second case study game in this thesis.

There are two ways to play this game; one is to try to catch as many insects as possible (quantitative way of playing). The other way is to try to catch only few insects but to select those that are rare and therefore more valuable (qualitative way of playing). One important thing in this game is also to keep those insects that the player already has caught in a good shape and alive [Insectopia, 2006; Peitz et al., 2007].

![Figure 9 Screenshots of the Insectopia game][Peitz et al., 2007].

Insectopia is a solo game but also makes it possible to search insects together with a friend. This joint search happens by sending a request for a friend who is also playing the game. Then this friend can either accept or reject this request. This is the only communication form that is supported by the game. Catching is faster when playing together with a friend; the players can catch all the insects they see. In solo game there is a waiting period after each catch when you cannot catch anything, only search.
**Hot Potato**

*Hot Potato* is a proximity game that is played with a mobile phone. The idea of *Hot Potato* (Figure 10) is to gain points by handling as many ‘potatoes’ for as long as possible [Niemi et al., 2005].

![Figure 10 Screenshots from Hot Potato](https://example.com/hot_potato_screenshots.png)

The game is played with a mobile phone that is Bluetooth enabled. The game idea is based on proximity. A player has to throw a potato forward to a non-player via Bluetooth to cool the potato down [Hot Potato, 2005; Niemi et al., 2005]. All the potatoes get hotter and hotter when the player is holding them, and eventually they will explode and the player will lose a life. Before this happens the player has to throw the potato to some non-player who is in proximity. Then the player has to wait a while until she thinks the potato has cooled down enough and pick it up again. In order to do this the non-player who has the potato has to be in the proximity of the player.

The background story of *Hot Potato* is presented below.

One day - a day much like any other - the Mexican dirt farmer Francisco was toiling away in the suffocating desert heat. He was digging a well to lay on water for his fields.

Whilst digging he stumbled upon a buried treasure. Deep down the hole, inside a piece of cloth, he found five of the most peculiar potatoes ever revealed to man. They were very precious indeed. The potatoes had faces and feet and hairstyles! They were scorching hot, heating in the palm of his hand. They were getting warmer and warmer. Francisco soon came to understand that he had to cool them down somehow or else they would explode and blow him into smithereens!

[Hot Potato, 2005]
The game is quite challenging since it demands that the player keeps track of the non-players who have the potato when it is sent to cool down.

**Yum Yum Sheep**

*Yum Yum Sheep* is a mobile phone game based on Bluetooth connections. Every Bluetooth connection is a sheep and the player is a monster who tries to catch those sheep one by one [Niemi et al., 2005]. The basic idea is pretty similar to *Insectopia*; each Bluetooth ID represents one sheep in *Yum Yum Sheep* whereas in *Insectopia* they represent insects. All the sheep that are in the proximity of the player are shown on the screen. The goal is that only one sheep is in the proximity at a time because if there are more than one sheep in the proximity they will attack the player.

**Your Story**

*Your Story* is a single-player game that is heavily based on narrative. A player reveals a mystery piece by piece as she moves forward in the game. In some points of the story the player has to send a Bluetooth message to a non-player and ask for help [Niemi et al., 2005]. If that person is willing to help the player can proceed in the story.

**Spy Blob**

*Spy Blob* is a controversial game idea that came quite close to stalking. The basic idea is that a player selects one Bluetooth connection and stays within contact of that same connection for a fixed time [Niemi et al., 2005]. This usually requires that the player follows the non-player that she has selected as a target. Because of the nature of the idea, the game was restricted to be played within a limited area. The people in that area were informed about the game so that they could shut down their Bluetooth connections if they did not want the players to follow them.

**Feeding Yoshi**

*Feeding Yoshi* is a multiplayer game that breaks the magic circle temporally lasting several days. The test game that was described in Bell et al. [2006] lasted a week. In Figure 11 there are two pictures of the user interface of the game.
The goal in *Feeding Yoshi* is to feed Yoshis that can be found all over the city with fruits that they desire. Players have to grow fruits in plantations from seeds that they get from Yoshis. Plantations can also be found all over the city. Both Yoshis and plantations are wireless access points (Yoshis protected networks and plantations open networks) that the players’ PDAs scan when they are walking around the city.

### 2.3.6. Event games

Event games last only a short, defined duration at a time. These games can give once-in-a-lifetime experiences to a player and they are not necessarily trying to be replayable [Montola *et al.*, 2006]. Setting up an event game like *Momentum* [Jonsson *et al.*, 2007] (Section 2.3.2) can take a lot of time and effort, and together with the technology that is needed for the game the costs can get quite high. Still the experience that a player gets from these kinds of games is so intense and enjoyable that people are willing to pay to be part of the game. None of the example games (Table 1) are solely event games.

### 2.3.7. Crossmedia games

Crossmedia games are games that use simultaneously multiple media sources and devices [Montola *et al.*, 2006]. Many crossmedia games use traditional media sources like TV, radio, print media, and mobile equipment in the gameplay.
**Epidemic Menace**

*Epidemic Menace* (Figure 12) is a crossmedia game where a virus mutation threatening the humankind has been released on a university campus and players’ goal is to identify where the viruses are and catch them [Ohlenburg *et al.*, 2006].

![Figure 12 Pictures of the game session of Epidemic Menace [Ohlenburg *et al.*, 2006].](image)

All the players become medical experts in the game and they work in collaborative teams to catch all the viruses [Ohlenburg *et al.*, 2006]. This must happen before the viruses get outside the campus area. Players also need to figure out how this terrible disaster could have happened. *Epidemic Menace* is a collaborative game played in teams where players have two different kinds of roles, stationary or mobile [Ohlenburg *et al.*, 2006]. In stationary play mode a player is in a game team room observing and analyzing viruses and giving instructions for mobile team players. Players who play the game in the mobile play mode are outdoors capturing viruses.

**Garden of Earthly Delights**

*Garden of Earthly Delights* is a fantasy role-playing game that can be played both with a PC and a mobile phone. It is based upon a persistent, virtual, multiplayer 3D environment that runs on networked PCs [Koivistö and Eladhari, 2006]. Still there is an extension to traditional PC role-playing games as this game also uses
location as a game mechanic. Players can play the game via mobile phones anytime and anywhere they choose [Koivisto and Eladhari, 2006]. Figure 13 shows screenshots of these two playing platforms.

![Two screenshots of Garden of Earthly Delights in mobile client (left) and one screenshot of the PC client [Becam and Kullgard, 2006].](image)

Garden of Earthly Delights uses both the virtual and the physical world in its gameplay. Players have characters that they control; they socialize with each other and use magic in this fantasy world. There are three possible playing modes for the player. 1. The player can interact with other mobile players based on their physical location. 2. The player can interact in a virtual world with other players via mobile phone but not using location data. 3. The player can play the game totally location-based, where the physical world is “shadowed” by the player’s character in the virtual world and interaction with other players can happen in both worlds [Becam and Kullgard, 2006].

2.3.8. Summary

Pervasive games bring games back to reality. They are somehow connected to the real world. For example, it is usual that the real world is a playground and gameplay is happening in real life: some tasks that the player has to perform in a game are done in real life. This means that some pervasive games include quite much player movement. Pervasive games include usually more interaction between players than some other types of games.

Figure 14 presents the example games in a two-dimensional space of movement and communication. This figure was first presented by Suominen [2003] but I have added the example games into it.

As can be seen from the figure, most of the example games require a lot of movement from the player and also provide a wide and open communication environment. The only games that do not require movement are Day of the Figurines and Coup, which are both mobile games that are not based on location.
Your Story, Yum Yum Sheep, Insectopia, Spy Blob, and Hot Potato are exceptional games in that they do not provide as much communication to the players as many other pervasive games. What is common to these five games is that they all are based on Bluetooth connections. These games may seem to be very interactive and supportive for communication because they are best to be played in a crowd with a lot of open Bluetooth connections. Still the games are mostly single-player games so even if they utilize other people’s Bluetooth connections, players do not interact with them on a deeper level. For example, these games do not demand the players to talk to others.

All pervasive games break the magic circle of games somehow. Table 2 summarizes subsection 2.3 through the fifteen example games. In the left column of this table is the sub-genre where the game belongs. In Epidemic Menace two sub-genres are mentioned instead of just one because it can be categorized into either of these genres.
Table 2 How the example games break the magic circle.

<table>
<thead>
<tr>
<th>Sub-genre of the game</th>
<th>Example games</th>
<th>Breaks the magic circle?</th>
<th>Spatially</th>
<th>Temporally</th>
<th>Socially</th>
<th>How long the game lasts?</th>
<th>How many players?</th>
<th>Is the game asynchronous (A) or synchronous (S)?</th>
<th>Does the game need a puppet master?</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMM</td>
<td>Botfighters</td>
<td>x x x</td>
<td>continuously</td>
<td>AT</td>
<td>A</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMM</td>
<td>Coup</td>
<td>x</td>
<td>~a week</td>
<td>AT</td>
<td>A</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMM</td>
<td>Day of the Figurines</td>
<td>x x</td>
<td>24 days</td>
<td>AT</td>
<td>A</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG (EG)</td>
<td>Epidemic Menace</td>
<td>x</td>
<td>few hours at a time</td>
<td>MP</td>
<td>S</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PG</td>
<td>Feeding Yoshi</td>
<td>x x</td>
<td>continuously</td>
<td>AT</td>
<td>A</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG</td>
<td>Garden of Earthly Delight</td>
<td>x x</td>
<td>not said</td>
<td>AT</td>
<td>A</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PG</td>
<td>Hot Potato</td>
<td>x x x</td>
<td>continuously</td>
<td>SP*</td>
<td>-</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PG</td>
<td>Insectopia</td>
<td>x x x</td>
<td>continuously</td>
<td>SP</td>
<td>-</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMM</td>
<td>Pirates!</td>
<td>x x</td>
<td>few hours at a time</td>
<td>AT</td>
<td>S</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMM</td>
<td>Savannah</td>
<td>x</td>
<td>few hours at a time</td>
<td>MP</td>
<td>S</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PG</td>
<td>Spy Blob</td>
<td>x x x</td>
<td>continuously</td>
<td>SP*</td>
<td>-</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMM</td>
<td>Songs of North</td>
<td>x x x</td>
<td>not said</td>
<td>MP</td>
<td>A</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OO</td>
<td>Uncle Roy All around You</td>
<td>x</td>
<td>few hours at a time</td>
<td>MP</td>
<td>S</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PG</td>
<td>Your Story</td>
<td>x x x</td>
<td>continuously</td>
<td>SP*</td>
<td>-</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PG</td>
<td>Yum Yum Sheep</td>
<td>x x x</td>
<td>continuously</td>
<td>SP</td>
<td>-</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MMM = massively multiplayer mobile game  SP = single player game
PG = proximity game                     AT = alone together
OO = online-onstreet game                MP = multiplayer game
EG = event game                          (* also includes the possibility for joint search)
CG = crossmedia game                      ARG = alternate reality game
PLARP = pervasive live action role playing game

Table 2 also shows how the example games break the magic circle. The length of the game is important when the game breaks the magic circle temporally, so this information is also included in this table. Games that do not break the magic circle temporally (Pirates!, Savannah, Uncle Roy All around You, and Epidemic Menace) continue only few hours at a time while some other games like Coup can last up to a week depending on the session length determined by the player when starting the game. Games that last more than a few hours mean that a player can play the game less intensively, moving in and out of the game when she feels like it. There are also seven games that do not have any fixed length; instead, they are continuously on. In multiplayer games the question whether a game is synchronous or asynchronous is important. Asynchronous means that players can be in the game in different times, but still play together
or against each other. This can be executed for example by letting a player choose the length of the game (like in Coup). In synchronous games all the players are in the game at the same time.

The last column in Table 2 is about whether or not the game needs mastering. Sometimes pervasive games need a certain game master to guide the game through, called a puppet master [McGonigal, 2007]. It is easy to give more instructions to the player through a mobile phone and also for the player to contact the puppet master. Still there are many pervasive games that do not need a specific game master. Both of the case studies, Insectopia and Coup are such in nature that they do not need a game master.

Pervasive games can be single-player games or multiplayer games or something in between. Table 2 refers to these as single-player (SP), alone together (AT), or multiplayer (MP) games. Single-player means that a player plays the game solo without collaborating with other players. Still these games can include high-score lists. Many Bluetooth ID based games are like this (e.g. Hot Potato and Spy Blob). Insectopia is a bit different: even though it basically is a single-player game, it also allows collaborative searching with one friend at a close range. Alone together types of games are multiplayer games: the gameplay is based on other players’ actions but not necessarily on collaboration between the players. A good example of this type of game is Coup where actions are done mainly solo but against other real players. The third type is multiplayer game where players not only play against each other but together collaborating with each other. Savannah is a good example of this type of game. Groups of players play against each other but inside the group, players have to collaborate in order to proceed in the game.

2.4. Pervasive mobile games
Pervasive games often use mobile devices in their gameplay. There are also pervasive games that do not use any mobile device but they are a minority. The mobile device most used in pervasive games is a mobile phone. This is because mobile phones are already pervasive by nature [Niewdorp, 2005]. Out of the fifteen example games eleven use mobile phones in their gameplay. The four other games that use PDAs are Epidemic Menace, Feeding Yoshi, Savannah, and Uncle Roy All around You.

There are many reasons why especially mobile phones offer a good platform for pervasive games. People have mobile phones and they know how to use them so phones offer a cheap and easily affordable platform for pervasive games [Peitz et al., 2007]. This also means that there are no problems
that come from presenting new technology to players (technology acceptance, see e.g. [Jones and Marsden, 2006, p. 63]). Of course there are some limitations when using mobile phones as a platform for pervasive games (e.g. small displays, requirement for a new phone model, need to download the game material into your phone) but still they are much more feasible than building a new platform from scratch. Also these limitations are not that crucial for pervasive games because they usually use mobile phones only to pass information to a player. The gameworld itself can be part of the physical environment so the demands are not that high for mobile phones and the graphics that mobile phones can present.

Mobile phones offer support for every expansion style, spatial, temporal, and social [Holopainen, 2006]. They are usable anywhere and new phones have built-in ability for location tracking. These functionalities support spatial expansion. Temporal expansion means that a player does not have to play a game at some specific time but rather when she wants to. Mobile phones are personal devices so a player has her phone with her almost all the time. It is also possible to get online with a mobile phone any time and almost any place.

Probably the most important expansion style, social, can be well supported by mobile phones. People trust their mobile phones and feel that it is safe to use them when interacting with others. They also include existing social networks like address-books that can be used as part of the gameplay. Mobile phones also support both synchronous and asynchronous playing styles (e.g. phone calls vs. text messages).

2.5. Case studies

I introduced all the fifteen example games briefly, but now I will discuss the two case study games, Insectopia and Coup, in more detail. They were evaluated in the fall 2006 as part of the IPerG project. I was evaluating them myself so out of those evaluations I have first-hand knowledge. The data gathering methods and background information about these two game evaluations are described in this section. In Chapter 5 the results and experiences that were gained from those two evaluations are used for comparing different methods and their pros and cons.

2.5.1. Insectopia

The evaluation for Insectopia was done in Tampere, Finland during November 2006. The game actually had been available for public for some time but no evaluations had been done. The goal of this evaluation was to see how well the
pervasive game fits into players’ daily lives. The evaluation was summative in nature so even though it gave ideas for further development, the improvement ideas were not carried out.

Data gathering methods

_Insectopia_ requires players to walk around a lot, since there was no possibility to study the gameplay “live”. This is the case also in some other pervasive games; there are only few opportunities to study the gameplay “live” [Jegers and Wiberg, 2006]. Laboratory testing is not possible because the game is expanded to the environment. This was the case with _Insectopia_ and that affected the data gathering methods that were chosen.

Before field testing started, we (three project members) did a small heuristic evaluation of this game. In this heuristic evaluation we used the mobile game heuristics by Koivisto and Korhonen [2006]. After the heuristic evaluation we recruited participants to play the game. We had to use confidants who were specifically asked to test the game because we had to loan the phones to them. The lack of proper phones also caused that the playing time for each participant was really short, only a couple of days. Also all the phones were not the same model. Some of the phones were better suited for the game than others. The main problem with the phones was the difference in their screen resolution, which made the game appear differently in their screens.

After the playing period, the participants filled in a web questionnaire that was divided into different themes, and which also included many background questions. In semi-structured interviews the participants were asked more specific questions about the game and about the participant’s feelings. All but one of the interviews were recorded.

Participants

According to the web page of _Insectopia_, it has had 109 players (checked 9.1.2007) altogether [Insectopia, 2006]. For this evaluation eleven people filled in the questionnaire. The participants were between ages 14 to 31, and seven of them were male and four female. Out of those eleven seven also took part in the interview (two women and five men). Six of the participants were students.

Most of the participants said that they play some kind of games, only two participants said that they don’t usually play at all. The most played game genres were action games, platform games, and puzzle games (six answers each). The only game genre that no-one said that they play, was sport games. We asked from the participants also with whom they like to play the games. Almost everyone said that they enjoy playing with their friends (nine answers),
friends in the Internet (seven answers) and strangers in the Internet (six answers).

2.5.2. Coup

Coup was played in sessions that lasted from a couple of days to a week depending on how the game pace was set. Overall we had four game sessions (the lengths of each session are shown in Table 3) during November and December 2006 in Tampere, Finland. The number of players shown in Table 3 is the overall number of players gotten from the log data, not the number of the participants.

<table>
<thead>
<tr>
<th>Game</th>
<th>Game length</th>
<th>Average player life length</th>
<th>Number of players</th>
</tr>
</thead>
<tbody>
<tr>
<td>Game 1</td>
<td>6 d 21 h 17 min 54 sec</td>
<td>3 d 08 h 29 min 05 sec</td>
<td>32</td>
</tr>
<tr>
<td>Game 2</td>
<td>3 d 16 h 19 min 04 sec</td>
<td>2 d 04 h 18 min 28 sec</td>
<td>16</td>
</tr>
<tr>
<td>Game 3</td>
<td>4 d 21 h 10 min 38 sec</td>
<td>2 d 19 h 02 min 11 sec</td>
<td>12</td>
</tr>
<tr>
<td>Game 4</td>
<td>7 d 06 h 35 min 34 sec</td>
<td>3 d 08 h 45 min 01 sec</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 3 Game lengths and players in each test game.

Data gathering methods

Several methods were used in this evaluation: playability heuristics [Koivisto and Korhonen, 2006] used in very early phases, questionnaire, and semi-structured interview as in Insectopia. In addition to these three we also analysed the log data from each test game in order to see how the players’ view corresponded with the actual flow of the game. We asked the participants also to write a diary about their playing. None of them was willing to keep it as a full scale diary, but some participants did write short comments about their thoughts and feelings towards the game. We wanted to use multiple methods so that we could get as much data from different sources as possible. Also these methods were familiar to us beforehand and they give different types of data (both qualitative and quantitative) so we could compare the results.

We used a web questionnaire that was divided actually in two parts: questions about the playing experience and background questions. In developing the questionnaire (see Appendix A) previous evaluation work of IPerG was utilised. Participants filled in the questionnaire after their playing period which lasted for about a week depending on the season length.

Through the interview we wanted to get more precise knowledge about Coup; what things did the participants like about the game and what things needed to be improved. It would have been ideal to get more interviews and especially from participants with different backgrounds (in order to get
different kinds of playing experiences), but it was not possible due to the small number of participants.

Participants
We recruited people to play Coup through different game oriented web pages and a number of student-related mailing lists. We assume that more than a thousand people saw this invitation, yet only a few of those people actually tried the game. One reason for this lack of interest could have been the mobile phone requirements: people did not seem to have S60 phones as much as we thought. But there were also some other reasons like the fact that the players had to pay for the costs of the data connection by themselves. This cost is not particularly high with a suitable data transfer package in one’s subscription but without one, costs can get quite high as can be seen from this quote from one of the participants.

I designated somewhere around 4-6 hours to this game in three days. …
With my data fees playing like this took about 7-10 Mb (including downloading all the game and program files) so the start up cost is nearly 5 € to 10 € (a bit expensive…). (male, age unknown)

We also noticed that it was not that easy to install the MUPE-client into the phone. For these reasons we only had around 10-30 players playing Coup at the same time (Table 3). This was much less than we anticipated.

Overall nine participants filled in the questionnaire that was available in Coup’s web page. Out of these participants eight were male, one female. The age varied between 24 and 36. All the participants said that they play some kinds of games in their free time and all of them also had some kind of a device that enables playing. The background of most of them was university (five participants). All of the participants had quite a long playing history; they said that they had been playing games from 7 to 25 years. Five of the participants said that they had downloaded 1-5 items into their mobile phone, one hadn’t downloaded anything, and two had downloaded 15-50 items.

Three participants also took part in the interview (all males). All these three participants agreed on recording the interviews. The interviews were carried out after the play session and the participants were free to answer as briefly or in depth as they wanted to. Also if they raised some interesting details about the game they were asked for more information about them.
2.6. Research questions
The starting points for this study were the evaluations of *Insectopia* and *Coup*. They got me interested in this field and made me realize the lack of documented evaluation reports on pervasive games. We need a deeper understanding of pervasive games: how do people feel about them, and how well do they fit to people’s lives? To solve this, more pervasive games need to be developed and evaluated, as done in the IPerG project. My contribution to this game research area is to clarify how the special nature of pervasive games affects data collection for their evaluation. In short the research questions of this thesis are:

- what makes pervasive games different from traditional games,
- how this affects evaluation, and
- what are the data gathering methods that are suitable for evaluating pervasive games.
3. **Game evaluation**

Game evaluation should continue through the whole game development process, since the issues that are under the scope in different game development phases differ. Early on in the development process the focus is on the functionality of the game whereas later on the focus is on gameplay. Gameplay is hard to define because it is not a single entity but a combination of many different elements. Rollings and Adams [2003, p. 201] define gameplay as “one or more causally linked series of challenges in a simulated environment”. In gameplay the interaction between a game and a player is significant.

Another important term in game evaluation is playtesting, which “is something the designer performs throughout the entire design process to gain an insight into how players will experience the game” [Fullerton et al., 2004, p. 196]. So playtesting means that one should experience the game and take the position of a player. Usually this happens simply by playing the game prototype in every game development phase.

3.1. **Why to evaluate games?**

There are many reasons for doing game evaluations. Some of these reasons are the same as with any other software but some are game specific. The business point of view is the tough competition between games so the game must be good in order to be successful. Not that many games are profitable and usually there are a lot of similar titles available. Competition has forced companies to compete against each other harder than before by offering customers new services and extra features. By doing evaluations in different development phases, software can be improved, and it creates a competitive advantage for the company.

Games are meant to entertain people. Using productivity software can be a necessity, but playing games is not. Games have to compete against other mediums for people’s free time. This means that if a game is not good enough, people will not play it, but use their free time in some other way. This puts a lot of pressure on game developers. Evaluating games throughout the whole
development process ensures at least that there is a chance that the game will be successful.

One reason for evaluation is also the fact that players have different skills and game experiences than game developers (Figure 15). Especially if the target group of a game is very different from game developers of that game, the evaluation is even more important. Game developers have usually much more game experience than players so they cannot see the problems that a casual gamer encounters. Still this does not mean that experts cannot or should not evaluate games. Especially in the early game development phases, experts can focus on right aspects of the game whereas gamers usually would concentrate on the issues that are not yet fully functional.

![Figure 15 The experience difference between gamers and game developers](Fullerton et al., 2004).

Game evaluation is very important in every game development process. Especially in pervasive game development it defends its place because no-one really knows yet that much about pervasive games and all the aspects that are part of the player experience in those types of games.

### 3.2. Components of playability

One very much used psychological theory in game studies, and the basis of defining playability is flow-theory about experiencing enjoyment. This theory was first introduced by Csikszentmihalyi [1991] and the most important finding was that people experience enjoyment in a surprisingly similar way. According to him, flow experience describes “situations in which attention can be freely invested to achieve a person’s goals” [Csikszentmihalyi, 1991, p. 40]. Flow is a mental state, where a person is fully immersed in what she is doing. Csikszentmihalyi has identified eight elements that are preconditions for this flow experience to happen.
1. The task has clear goals.
2. The person has the ability to concentrate on the task at hand.
3. The person experiences the loss of self-consciousness during the task but has a stronger sense of self afterwards.
4. The person’s sense of duration of time is altered (the transformation of time).
5. The task provides immediate and clear feedback to the person.
6. The task offers a challenging activity that requires skills but that can be completed.
7. The person thinks that she controls the actions, even though they are pre-set (the paradox of control).
8. The person experiences merging of actions and awareness.

The flow idea has been used a lot also when talking about enjoyment of games [Sweetser and Wyeth, 2005; Jegers, 2006; Järvinen et al., 2002]. This is because it especially addresses player enjoyment in computer gaming [Jegers, 2006]. The key issue is how to be able to provide a player the necessary preconditions so that this flow experience will happen [Järvinen et al., 2002, p. 21]. The biggest problem with flow is that it does not explain why games are engaging and enjoyable even at times when they produce frustration to the player [Järvinen et al., 2002, p. 30].

Järvinen et al. [2002] have used flow theory as a starting point for their definition of playability. They have defined it to be “a qualitative term for the uses of both design and evaluation”. The concept of playability is formed from four smaller components: 1) functional, 2) structural, 3) audiovisual, and 4) social playability (Figure 16). Functional playability includes all the controls that are a requirement for successful gameplay. Usually controls should be as simple and as intuitive as possible. Structural playability includes both micro- and macro-level structures like rules and interaction patterns. Audiovisual playability is about the audiovisual style and appearance. Social playability is about what kind of social practices the game is suitable for. For example, social playability includes aspects that unfold in multiplayer games when players interact with each other, but it also includes player-observer situations.
Koivisto and Korhonen [2006] have proposed a set of heuristics for evaluating playability of mobile games. These heuristics include issues that both Csikzentmihalyi [1991] in his flow theory and later Järvinen et al. [2002] in their following work found important for enjoyment and further on, playability. Koivisto and Korhonen [2006] have decided to split the concept of playability in smaller modules of playability, usability, and mobility for evaluation purposes. Järvinen et al.’s proposition is also based on components but whereas their components are somewhat fixed, Koivisto and Korhonen’s proposal is more flexible because the modules can be selected according to the game at hand. Whereas usability and gameplay modules are essential for all games, mobility is game type specific and focused only on the evaluation of mobile games.

Modules of playability and some issues that are part of the player experience are shown in Figure 17. This figure is an extension of playability modules of Koivisto and Korhonen [2006]. In this figure there are also three main modules like in their work, but whereas the third module in their work is mobility, I have included all the game type specific issues into the third module. Among these game type specific issues are multiplayer aspects that were also acknowledged in Korhonen and Koivisto [2007].

One should select the right sub-modules based on the game type that is the target of the evaluation. It should be noted that in Figure 17 game specific issues mean features of the game, not the player. Casual in this figure means features of the game that support casual playing style. Modularity makes this approach suitable for evaluation purposes because it is possible to choose the appropriate modules for each case.
There are similarities in Järvinen et al.’s [2002] and Koivisto and Korhonen’s [2006] approaches to playability, for example structural playability [Järvinen et al., 2002] includes the same issues that Koivisto and Korhonen have in their gameplay module. In a similar way functional and audio-visual playability issues can be found from the usability module.

Table 4 compares gameplay issues of Nokia’s playability heuristics [Koivisto and Korhonen, 2006], definition of playability [Järvinen et al., 2002], usability heuristics [Nielsen, 1994] (full list as an Appendix D), and the flow-model [Csikzentmihalyi, 1991]. This table shows the similarities that these models have with each other.
Table 4 Comparison about gameplay between Koivisto and Korhonen’s playability heuristics [2006], Nielsen’s heuristics [1993], Järvinen et al.’s definition of playability [2002], and Csikzentmihalyi’s flow-model [1991].

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>GAMEPLAY</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The game provides clear goals or supports player-created goals</td>
<td>structural playability</td>
<td>The task has clear goals</td>
<td></td>
</tr>
<tr>
<td>The player sees the progress in the game and can compare the results</td>
<td>1. Visibility of system status</td>
<td>Task provided immediate and clear feedback</td>
<td></td>
</tr>
<tr>
<td>The players are rewarded and rewards are meaningful</td>
<td>structural playability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The player is in control</td>
<td>3. User control and freedom</td>
<td>The paradox of control</td>
<td></td>
</tr>
<tr>
<td>Challenge, strategy, and pace are in balance</td>
<td>structural playability</td>
<td>A challenging activity that requires skills but that can be completed.</td>
<td></td>
</tr>
<tr>
<td>The first-time experience is encouraging</td>
<td>structural playability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The game story supports the gameplay and is meaningful</td>
<td>structural playability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>There are no repetitive or boring tasks</td>
<td>structural playability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The players can express themselves</td>
<td>3. User control and freedom</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The game supports different playing styles</td>
<td>3. User control and freedom</td>
<td>structural playability</td>
<td></td>
</tr>
<tr>
<td>The game does not stagnate</td>
<td>structural playability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The game is consistent</td>
<td>4. Consistency and standards</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The game uses orthogonal unit differentiation</td>
<td>structural playability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The player does not lose any hard-won possessions</td>
<td>structural playability</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

There are a lot of similarities between all of these models as can be seen in Table 4. Still the other models are not exact matches for the gameplay module of playability heuristics. Five gameplay heuristics have a corresponding item in Nielsen’s heuristics. Nielsen’s third heuristic “user control and freedom” fits three heuristics in Koivisto and Korhonen’s module. It is not a surprise that there are not more similarities between these two heuristic sets. This is because the table presents only the gameplay module by Koivisto and Korhonen, and it does not include usability issues that are the focus of Nielsen’s heuristics. Still it is interesting to see that also these two have some similarities despite the fact that they are concentrating on different issues.
Nine of the gameplay heuristics by Koivisto and Korhonen can be placed into the structural playability module of Järvinen et al. which includes rules and interaction patterns. Other heuristics cannot be placed properly into any of the four modules that Järvinen et al. presented. This is because gameplay is supposed to include mainly the rules and other issues that are related to game-player interaction. Issues that could be included in functional or audio-visual playability are more or less usability issues that will be discussed later on.

Four flow-model elements are covered in gameplay heuristics. These four elements are concentrating on the task at hand. The other four are not covered in heuristics, because their focus is on the player, not the task. They are impossible to assess by the expert evaluator who is supposed to use the heuristics. For example, one of the flow elements is that a player experiences the loss of self-consciousness, but this cannot be assessed by the expert.

3.3. Playability vs. usability
When talking about productivity software, usability and user experience are in focus. In games it is playability and player experience (like was shown in Figure 17). Usability is according to ISO standard 9241-11:1998 “The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” [ISO, 1998]. User experience is an umbrella concept for usability. Whereas usability is concerned about usefulness and productivity of a system, user experience is about how the user experiences the interaction with the product [Preece et al., 2002, p. 9]. So usability “refers to how easy it is for users to learn a system, how efficiently they can use it once they have learned it, and how pleasant it is to use” [Nielsen and Mack, 1994 p. 3] but it can cover user experience only partially.

Preece et al. [2002] have defined usability and user experience goals in Figure 18. Usability goals are presented in the middle. They are usually quite easy to determine and experts can usually evaluate whether a system meet these goals or not. User experience goals in the outer circle include issues like is the system fun to use, or is it aesthetically pleasing. These goals are much less defined than usability goals, and they are based on subjective experiences. Since they unfold when a user interacts with the system, they are quite hard to evaluate without actual users.
Playability is in certain sense similar to usability. Also playability goals can be evaluated without actual users although when talking about entertainment media such as games, the experience of users is essential.

Usability is part of playability evaluation. Good usability is the core of good playability and player experience just like usability is the core of user experience. Heuristics that have quite a long history in usability evaluation have been adapted to games. Especially playability heuristics that are targeted for detecting usability problems can be directly taken from Nielsen’s usability heuristics. Table 5 is a comparison between the usability module of the playability heuristics [Koivisto and Korhonen, 2006], Nielsen’s heuristics [1993], Csikzentmihalyi’s flow-model [1991], and Järvinen et al.’s playability definition [2002]. As can be seen from the table there are a lot of similarities in the usability module of the playability heuristics and the usability heuristics. Only one of Koivisto and Korhonen’s usability heuristics does not have a corresponding heuristic in Nielsen’s heuristics: “audio-visual representation supports the game”. Since it is especially for games, there is no exact match found in Nielsen’s heuristics. Other than that, the heuristics are similar, only the wording differs.
### Table 5 Comparison about usability between Koivisto and Korhonen’s playability heuristics [2006], Nielsen’s heuristics [1993], Csikzentmihalyi’s flow-model [1991], and Järvinen et al.’s definition of playability [2002].

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>USABILITY</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audio-visual representation supports the game</td>
<td>8. Aesthetic and minimalist design</td>
<td>audio-visual playability</td>
<td></td>
</tr>
<tr>
<td>Screen layout is efficient and visually pleasing</td>
<td>4. Consistency and standards</td>
<td>functional playability</td>
<td></td>
</tr>
<tr>
<td>Device UI and game UI are used for their own purposes</td>
<td>1. Visibility of system status</td>
<td>audio-visual playability</td>
<td></td>
</tr>
<tr>
<td>Indicators are visible</td>
<td>2. Match between system and the real world</td>
<td>functional playability</td>
<td></td>
</tr>
<tr>
<td>The player understands the terminology</td>
<td>4. Consistency and standards</td>
<td>functional playability</td>
<td></td>
</tr>
<tr>
<td>Navigation is consistent, logical, and minimalist</td>
<td>2. Match between system and the real world, 4. Consistency and standards</td>
<td>functional playability</td>
<td></td>
</tr>
<tr>
<td>Control keys are consistent and follow standard conventions</td>
<td>2. Match between system and the real world, 4. Consistency and standards</td>
<td>functional playability</td>
<td></td>
</tr>
<tr>
<td>Game controls are convenient and flexible</td>
<td>7. Flexibility and efficiency of use</td>
<td>functional playability</td>
<td></td>
</tr>
<tr>
<td>The game gives feedback on the player’s actions</td>
<td>9. Help users recognize and recover from errors</td>
<td>structural playability</td>
<td>Task provides immediate and clear feedback</td>
</tr>
<tr>
<td>The player does not have to memorize things unnecessarily</td>
<td>6. Recognition rather than recall</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The game contains help</td>
<td>10. Help and documentation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The player cannot make irreversible errors</td>
<td>5. Error prevention</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Järvinen et al. [2002] defined playability in four modules: functional, structural, audio-visual, and social playability. All of the usability issues in games can be placed into these categories. Most of the usability issues are either functional (for example game controls) or audio-visual (the way the game is represented to a player) playability issues.

The flow-model presents the elements that are part of the ultimate enjoyment. Just for the comparison it has been added to Table 5 as well. As can be seen, there are hardly any similarities between the flow-model and usability heuristics. Mostly this is because the flow-model does not concentrate on any targeted issues like heuristics but rather on the overall situation where the
experience happens. That is why there is only one flow element that matches the others: “task provides immediate and clear feedback”.

3.4. The game evaluation process
The starting point when evaluating games is very different from productivity software. Playability evaluation concentrates on enjoyment and fun [Koivisto and Korhonen, 2006] whereas evaluation of productivity software concentrates on usability and effectiveness. When evaluating games one has to take a new perspective. It is not enough that the game is easy and effective to use. Quite the contrary; if the game is too effective and offers no surprises (the enjoyment of discovering new things), it will be predictable and boring. What makes games fun, is that the players do not know what to expect [Koivisto and Korhonen, 2006]. Good usability is important in games as well, but the absence of usability problems does not mean that the game is fun and engaging [Monk et al., 2002].

Many times improving usability of productivity software means making the functionality as clear for the user as possible. In games, the playability often relies to some extent on withholding certain functionalities or options from the player. If all the options would be clear to the player, the game would soon be too easy and boring. Withholding and revealing information in the game is one important aspect that determines the difficulty level of the game.

Traditional evaluation methods can be used when evaluating playability. Game industry has adopted evaluation methods and experiences from other disciplines since it is in itself so new [Benford et al., 2005a]. The discipline that has affected the most is Human-Computer Interaction (HCI) which is based on psychology. Two other disciplines where game industry has adopted ideas and experience especially recently are sociology, and art and design.

There are many similarities between productivity applications and games; for example, they both typically have screens and menus [Pagulayan et al., 2003, p. 137]. Still there are also some very important differences that need to be taken into consideration when evaluating playability of games. When designing a game, it also has to provide an enjoyable experience. This is something that designers of other applications do not need to take into account that much. When designing a game, the goal cannot be the same what it normally is; that the user can achieve task completion as quickly and efficiently as possible. It would ruin the whole game experience. A game has to provide a right level of challenge and engagement, otherwise people would not play it anymore [Pagulayan et al., 2003, p. 138].

Evaluation can be done in early phases of game development (formative evaluation) or in the last phase of the development (summative evaluation) (e.g.
[Montola et al., 2005]). An outline of the game development process is presented in Figure 19. If evaluation is done in early phases of the development, its main task is to discover design flaws and things that prevent having good playability. Then they can be corrected before the game is launched by giving some feedback for the developers. If the evaluation is done in the last phase of development, the goal is to see how well the requirements that were set to the game in the beginning were reached.

It is preferable to use both formative and summative evaluations. It is recommended that the game is tested through the whole development process so that most of the obvious playability problems can be fixed before the game is finalised. Still it is also important to test the game with the real users after the game is finalised to see how they react to the game. This is especially important in pervasive games because there is not that much research done about the things that affect their playability. It is not fully known how people will react to the pervasive features of the game. Evaluating the game is important so that it would be easier to develop better games in the future.

3.5. Selecting and recruiting participants

Usually the game development team runs the evaluation in the early phases of the game development process but later on it is good to test the game with outsiders. Table 6 shows who one should recruit to playtest in which stage. For example, bug testing (seeing if a game works in the way it is supposed to) should be done thoroughly before the game is released or even playtested with the real target audience because bugs take time and also attention from the real playability issues. It will be hard for the users to concentrate on things that are really important if the game does not work properly.
Table 6 Types of playtesters appropriate in game development [Fullerton et al., 2004, p. 200]. The column on the right is added.

<table>
<thead>
<tr>
<th>Prototyping stage</th>
<th>Playtest on your own</th>
<th>Playtest with people you know</th>
<th>Playtest with target audience</th>
<th>The phase of the game development process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foundations</td>
<td>X</td>
<td></td>
<td></td>
<td>From implementation to release</td>
</tr>
<tr>
<td>Structure</td>
<td>X</td>
<td>X</td>
<td></td>
<td>Design</td>
</tr>
<tr>
<td>Formal details</td>
<td></td>
<td></td>
<td>X</td>
<td>Features complete</td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
<td></td>
<td>X</td>
<td>Content complete</td>
</tr>
</tbody>
</table>

Structure is an issue that can be tested by a game development team or by confidants. In this stage the prototype is somewhat playable even though a game developer may need to explain it to participants because all the functionalities do not yet work. The idea of this stage is to get knowledge about how the gameplay proceeds. Both foundations and structure are straightforward and relatively easy to evaluate. Formal details and refinement should be tested with target audience. In formal detail stage the implementation is almost complete. All the functionalities and features have been implemented. Refinement is the last stage, which basically includes only some minor changes to the content or some other small improvements. At this stage, the game is almost ready to launch.

Recruiting participants is one of the most important but often the most challenging part in an evaluation. Without participants it is impossible to throughoutly test the playability of games. In the example game evaluations many recruiting means were used. All these means have their pros and cons, and they fit in different situations. Table 7 presents more specific information about recruiting; what are the advantages and disadvantages of using different groups in evaluation. Table 7 also contains information about what means were used in the example game evaluations.
Table 7 How to select participants according to Jones and Marsden [2006, p. 128]. The right column has been added.

<table>
<thead>
<tr>
<th>Participants</th>
<th>Pros</th>
<th>Cons</th>
<th>Use</th>
<th>Example games that used this method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selecting people &quot;off the street&quot;</td>
<td>Can pick people actually in situation under investigation.</td>
<td>Time intensive - many rebuffs likely.</td>
<td>Short, focused surveys. Getting impressions of population profile to help recruitment for more extensive study.</td>
<td>Yum Yum Sheep, Your Story, Hot Potato, Spy Blob, Pirates! Coup, Songs of North, Epidemic Menace, Uncle Roy All around You</td>
</tr>
<tr>
<td></td>
<td>Can observe some of potential participants' behaviour before approach is made. Fast exposure to many potentials. Quick responses.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advertising</td>
<td>Can reach broad numbers of potential respondents.</td>
<td>Can be expensive. Self-selection is a problem.</td>
<td>Focus-group style discussion. Profile gathering to assist in selection of people to study in more depth.</td>
<td>Yum Yum Sheep, Your Story, Hot Potato, Spy Blob, Pirates! Coup, Songs of North, Epidemic Menace, Uncle Roy All around You</td>
</tr>
<tr>
<td></td>
<td>Can target user groups through advert placement (relevant types of magazine, website, etc.).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Existing customers</td>
<td>Large amount of profile data already exists.</td>
<td>Can overlook needs of non-customers who might adopt important 'disruptive technologies'.</td>
<td>Widely useful, from large survey-type data gathering to in-depth interviews with individuals.</td>
<td>The Garden of Earthly Delights, Botfighters (participants had played the game or similar type of the game before)</td>
</tr>
<tr>
<td></td>
<td>Easy incentives to participants (e.g. reduces call costs).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employees of own organization</td>
<td>Compliant and accessible subjects.</td>
<td>Subjects too subjective or reticent to criticize.</td>
<td>Widely useful, from large survey-type data gathering to in-depth interviews with individuals.</td>
<td>Insectopia, Day of the Figurines, Garden of Earthly Delights, (Coup)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Referral</td>
<td>Much known about subjects. Source of high-authority individuals.</td>
<td>Availability of subjects may be low (or at high cost).</td>
<td>Gathering of emerging trends: broad, 'big-picture' insights.</td>
<td>Insectopia, Savannah, Feeding Yoshi</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Short and focused surveys that are done for selected people “off the streets” were not used in the example game evaluations. This method is not specifically good for playability evaluation but it can be used for gaining information about the game concept.

The most popular method for recruiting participants was advertising. In some cases (like in Coup for example) there was a web questionnaire, and invitations to fill in the questionnaire were sent using different channels like web forums or e-mail lists. This was the case also in Spy Blob, Hot Potato, Your
Story, Yum Yum Sheep, and Songs of North. In the evaluation of Songs of North one main function of the questionnaire was to get participants for the scenario based player study [Ermi and Mäyrä, 2004]. It is possible that this same tactic was used also in the evaluations of Spy Blob, Hot Potato, Your Story, and Yum Yum Sheep but it did not come out in the paper [Niemi et al., 2005]. Therefore I cannot say how they recruited those participants that were interviewed.

In the case of Pirates! straight advertising was used. The game was presented at the Handheld and Ubiquitous Computing (HUC) conference in August 2000, and people who participated in the conference were asked to take part also in the game evaluation.

The method of recruiting existing customers or in game context, choosing people who have played that or similar game before was used only in two example games: Garden of Earthly Delights and Botfighters. In the evaluation of Botfighters, all the participants were former players of Botfighters (the game did not run anymore at the time of the evaluation) [Bjerver, 2006]. These players were reached through one of the founders of the game.

The participants of Garden of the Earthly Delights had not played that particular game before, but they were all role players so the game type was familiar to them. The reason why former players are used quite seldom is simply the lack of pervasive games. Very few pervasive games are running all the time. That is why not that many people know what pervasive games are, or have played that type of game.

There are many possibilities for choosing how to recruit the participants as Table 7 shows. The easiest and cheapest way is to do the evaluation by oneself or to use colleagues. Still in some situations this is not enough. For example, getting people’s opinions about a specific game idea is something that has to be tested in the field.

3.6. The challenges of game evaluation
In every evaluation process there are challenges and potential difficulties. It is good to prepare beforehand for these difficulties and by doing that ensure that in every situation it is possible to get some information out of the evaluation. Since pervasive games break the known magic circle in novel ways, they also have some problems that traditional games might not have.

3.6.1. Getting enough participants to play the game
The biggest challenge in many cases is to get enough participants to test the game. If there are not that many participants, the evaluation results may be
wrong because of the distortion. This means that the results are neither reliable nor valid.

In pervasive games this problem is even more severe than in some other types of games because pervasive games are usually heavily based on communication. The player experience is tied to other players and their activity as well as to the player’s own performance and the game features. Critical mass means the number of people that is required in the game at the same time in order to be enjoyable (see e.g. [Korhonen and Koivisto, 2007]). The critical mass problem is a vicious circle where too few players cause that the game is not enjoyable and some players leave the game. Then the game becomes even more unplayable and again players leave the game. This happened in the case study game *Coup*.

This challenge of recruiting participants should be taken into account as well as possible already when designing an evaluation. If it is possible one should make sure that there are enough players playing the game if the gameplay is based on player collaboration. It is recommendable for a game to have also single-player content [Korhonen and Koivisto, 2007] or content that requires only few players so that it is possible to enjoy the game even if there are not that many players at the beginning. Sometimes it is necessary to sustain a game for a while to allow a critical mass to be formed. This can be done for example recruiting people outside the evaluation to play the game.

Another suggestion solving the critical mass would be making the start of the game a major launch event. If launch is advertised and prepared well and all players participating within the first day or two are given significant benefits, players are likely to start the game fast enough to achieve the critical mass in a burst.

The design documentation of all pervasive, multiplayer games should include an estimate of the critical mass required for adequate and full gameplay, and also a brief plan on how to achieve and sustain that mass.

3.6.2. Ethical issues and privacy

Another problem is that pervasive games based on social expansion use public interaction as their game mechanic. This means that people who are not themselves in the game can be drawn in, for example, they can be used as a source of the game. This can cause ethical issues (see e.g. [Benford et al., 2005a; Niemi et al., 2005]).

Another issue caused by playing in a public space is that other people can observe playing or be influenced by it [Montola, 2005]. The game events can cause some harm or discomfort to people who are witnessing playing but
bystanders can also be a negative thing for a player. Other people can react in a way that distracts a player.

3.6.3. Technical problems

Pervasive games use technology and when using technology there can always be some problems with it. This came out very clearly also in our evaluations.

It is hard or even impossible to prevent all technical problems from happening. The internet connection or the server can fail just in the critical moment. That is why one should always have some back-up plan for those cases if possible. Also before playtesting one should always inform all the people so that it is possible to react fast to problems that may occur.

Another problem especially for mobile games is the fact that the game usually has to be downloaded from the Internet. Even though downloading content for a mobile phone has increased it still can be a challenge for some participants. This came through in the evaluation of Coup; some players said that they were not able to get the game work properly in their mobile phones. This should be taken also into consideration when designing instructions for playtesting. Getting the game should be as easy as possible for the participants if they have to get it for themselves. Otherwise technical problems may cause that there will not be that many participants joining the game.
4. Literature overview to data gathering methods

It is not an easy task to evaluate pervasive games. They are even harder to evaluate than other types of games because the player also interacts with the environment. This is why field testing is usually the only possible way to get real information from the users. Still involving users is costly; it takes much effort, time and expertise to conduct a field study with the real users [Jones and Marsden, 2006, p. 109]. This is why especially in the early development phases other methods are recommendable.

Games are an old phenomenon but still there are not that many evaluation methods that are specifically targeted for games. Game evaluators should focus more on what are those important aspects in game experience (like playability) that we should measure in order to be able to evaluate the game properly, and then select tools that meet those requirements. In this chapter data gathering methods that were used in example game evaluations are described.

4.1. Heuristics

Heuristic evaluation is one of the usability inspection methods that is conducted by usability experts. In heuristic evaluation an interface is evaluated against rules of thumb [Cockton et al., 2003, p. 1122]. There are three major advantages in using heuristics: they are cheap, fast, and easy to use [Nielsen, 1994, p. 25]. This is also why different researchers have developed heuristics for games (e.g. [Fedoroff, 2002; Desurvire et al., 2004; Sweetser and Wyeth, 2005; Koivisto and Korhonen, 2006; Korhonen and Koivisto, 2007]). They offer a valuable way to evaluate games during the whole development process and in a tight schedule. This method can be attended to as part of an iterative design process [Nielsen, 1993, p. 155].

Heuristics are checklists where one can look at different aspects of the game that are important for the playability of the game. This fixed list is helpful for the game evaluator because then different evaluators will take into account the same aspects of the game. Usually expert evaluation is done by three (or
of evaluators is crucial. The results from six different studies (see e.g. [Nielsen, 1993, p. 156]) have proved that one expert can only detect about 35% of all problems. Therefore heuristic evaluation is a method that an expert should not be doing alone. Depending on the user interface, the recommended number of experts is 3 to 7 [Nielsen, 1993, p. 156]. This is the number of experts that should be used in evaluation so that most of the problems will be detected, but the costs are still not overwhelming. After the experts have separately searched for problems from the game they discuss together the problems found and rate them according to their severity. Usually the result of heuristic evaluation is a report where problems found are described in detail and solutions to these problems are suggested [Nielsen, 1993, p. 156-157].

Sometimes heuristics can be violated intentionally in order to build some interesting new dramatic in the game [Koivisto and Korhonen, 2006]. Still this should not be done without good reasons. In the best case scenario, violating heuristics can be an effective and interesting way to make a game even more engaging and fun but in the worst case it can also make the gameplay experience a lot worse.

Heuristic evaluation is done by experts; it does not involve the target group of the game. This means that some errors can be unseen for that reason. Experts may see some things as problems that do not bother the real players at all. Experts and players have different ways of looking at the game and they concentrate on different things [Fullerton et al., 2004]. This is inevitable but not necessarily a bad thing. It just needs to be realized when conducting an evaluation.

There are some problems with heuristics. For example it may be hard to select which heuristic list to use. Koivisto and Korhonen [2006] tried to use many heuristic lists for evaluating mobile games, but they noticed that none of them were appropriate for mobile games. That forced them to develop a new set of heuristics that takes also into consideration the special nature of mobile games, like the fact that the mobile phone handsets have not been developed for playing.

Heuristics for especially mobile games have been divided by Koivisto and Korhonen [2006] into three modules (Figure 20): gameplay (presented in Table 4), usability (presented in Table 5), and mobility. Together these three modules cover the whole playability of mobile games.

Mobility heuristics include three rules that are in scope when dealing with mobile phone games.
1. The game and play sessions can be started quickly.
2. The game accommodates with the surroundings.
3. Interruptions are handled reasonably.

As said earlier, the gameplay and usability modules include rules that are applicable to all games. The other modules should be selected according to the game at hand. The flexibility of Koivisto and Korhonen’s [2006] heuristic list is the reason why I think this is the best heuristic set for games (at least so far). Another good thing about this heuristic set is that the rules are simple and easy to follow. In some previous works, the rules have been too vague, complicated or hard to apply to games (see e.g. Malone [1982]; Desurvire et al. [2004]). The modular approach to playability is very fruitful especially from the evaluation point of view. Modules are easy to use when needed. Still there is a need for heuristics and evaluation guidelines that take pervasive aspects of a game into account.

The flow theory (Section 3.2) can be used to describe the enjoyment of pervasive games. Sweetser and Wyeth [2005] have used the flow-model and made heuristics for games based on that. Jegers [2006] has taken this heuristic set even further, and added criteria that are especially targeted for pervasive games. Table 8 presents the pervasive criteria. The whole list is in Appendix C. The rightmost column tells what expansion styles these new criteria support. Some of these new criteria do not match any particular expansion style but they still are important features to some pervasive games. Most of these new criteria support the temporal expansion style, which is often the most obvious difference in pervasive games compared to more traditional types of games. The two criteria that do not support specially any of the expansion styles are about giving players the freedom to make their own goals, and advancing in a game in a pace they want to. These two criteria could be applied to all games.
The feedback element in Table 8 does not include new criteria for pervasive games. Even though pervasive games differ in many ways from traditional games, the feedback should be given to players in a similar way and efficiency as in traditional games.

### Table 8 Pervasive game flow and criteria that are targeted especially for pervasive games [Jegers, 2006]. The column on the right is added.

<table>
<thead>
<tr>
<th>Element</th>
<th>Criteria</th>
<th>Expansion style the criteria supports</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clear goals</strong></td>
<td>Games should provide the player with clear goals at appropriate times</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pervasive games should support the players in forming and communicating their intermediate goals</td>
<td></td>
</tr>
<tr>
<td><strong>Concentration</strong></td>
<td>Games should require concentration and the player should be able to concentrate on the game</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pervasive games should support the player in the process of switching concentration between in-game tasks and surrounding factors of importance</td>
<td>Temporal expansion</td>
</tr>
<tr>
<td><strong>Immersion</strong></td>
<td>Players should experience deep but effortless involvement in the game</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pervasive games should support a seamless transition between different everyday contexts, and not imply or require player actions that might result in a violation of social norms in everyday context</td>
<td>Temporal expansion</td>
</tr>
<tr>
<td></td>
<td>Pervasive games should enable the player to shift focus between the virtual and physical parts of the game world without losing too much of the feeling of immersion</td>
<td>Spatial expansion</td>
</tr>
<tr>
<td><strong>Feedback</strong></td>
<td>Players must receive appropriate feedback at appropriate times</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pervasive games should stimulate and support the players in their own creation of game scenarios and pacing</td>
<td></td>
</tr>
<tr>
<td><strong>Challenge</strong></td>
<td>Games should be sufficiently challenging and match the player’s skill level</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pervasive games should help the players in keeping a balance in the creation of paths and developments in the game world, but not put too much control or constraints on the pacing and challenge evolving</td>
<td>Temporal expansion</td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td>Players should feel a sense of control over their actions in the game</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pervasive games should enable the players to easily pick up game play in a constantly ongoing game and quickly get a picture of the current status in the game world (in order to assess how the state of the game has evolved since the player last visited the game world)</td>
<td>Temporal expansion</td>
</tr>
<tr>
<td><strong>Player skills</strong></td>
<td>Games must support player skill development and mastery</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pervasive games should be very flexible and enable the players’ skills to be developed in a pace set by the players</td>
<td></td>
</tr>
<tr>
<td><strong>Social interaction</strong></td>
<td>Games should support and create opportunities for social interaction</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pervasive games should incorporate triggers and structures (e.g. quests and events, factions, guilds or gangs) that motivate the player to communicate and interact socially</td>
<td>Social expansion</td>
</tr>
</tbody>
</table>

Some of the criteria in Table 8 are not that usable. One of the biggest problems is that they do not take into account the fact that pervasive games are very
heterogeneous. All of these features that are listed in Table 8 do not fit to all pervasive games. That is why it would be fruitful to develop new heuristic modules for different pervasive games subgenres.

4.2. Questionnaires

Questionnaires are one of the most familiar data gathering methods. This means that the positive and negative sides are quite well known. Questionnaire is an indirect method which means that it is based on the users’ opinions [Nielsen, 1993, p. 209]. There are two basic questionnaire types, open-ended and closed questionnaires [Hiltunen et al., 2002]. Open-ended questionnaires give no answer alternatives but all respondents answer in their own words. Closed questions have different answering alternatives. Usually a questionnaire has both of these question types.

Questionnaires are widely used because they are so easy to execute and they fit so well in various situations. Especially now when questionnaires can be put online they are a cheap way to reach a wide range of people that live in a wide geographical area [Preece et al., 2002, p. 211; Hiltunen et al., 2002; Nielsen, 1993, p. 211]. It is also a good way to get much quantitative information in a short period of time.

The negative sides of questionnaires are that they are rigid, questions need a lot refinement (it is easy for people who are filling the form to misunderstand the question), and it is quite risky to ask personal questions with questionnaires [Hiltunen et al., 2002]. Also the response rate may stay low [Nielsen, 1993, p. 210].

Before one starts to develop a questionnaire, one should first think out some hypotheses that one wants to confirm [Hiltunen et al., 2002]. The questionnaire should be designed in a way that it is possible to see whether these hypotheses are true or not. The questions should be thought-out very carefully and they should be very clear and specific [Preece et al., 2002 p. 400]. It is possible that people do not understand questions in a similar way and that causes misinterpretations of results [Nielsen, 1993, p. 212]. This can be prevented at least to some extent by testing the questionnaire before the actual evaluation. There are also some other limitations that questionnaires have like that people have different answering styles (some exaggerate and others underestimate) and that people may not tell the truth but rather what they think is expected from them [Hiltunen et al., 2002].

While questionnaires are widely used, they are quite rarely used alone (e.g. [Preece et al., 2002 p. 211]), mostly because the data is usually quantitative
(unless open-ended questions are used). They fit well for questions about the player, the tasks that the player did in the game, physical and social aspects of the use environment, and the technology [Hiltunen et al., 2002, p. 69]. Nielsen [1993, p. 209] stated that questionnaires are good for revealing what features participants liked or disliked.

When using a questionnaire several issues should be taken into consideration. Preece et al. [2002, p. 400] have collected a checklist for designing questionnaires.

1. Make questions clear and specific.
2. When possible, ask closed questions and offer a range of answers.
3. Consider including a “no-opinion” option for questions that seek opinion.
4. Think about the ordering of questions.
5. Avoid complex multiple questions.
6. When scales are used, make sure the range is appropriate and does not overlap.
7. Make sure that the ordering of scales is intuitive and consistent, and be careful with using negatives.
8. Avoid jargon and consider whether you need different version of the questionnaire for different populations.
9. Provide clear instructions on how to complete the questionnaire.
10. A balance must be struck between using white space and the need to keep the questionnaire as compact as possible.

It is hard to design a good questionnaire, but by keeping these 10 guidelines in mind it is possible. Still it is always possible that the questionnaire will give false information. For example, even though questionnaires can be used to ask people’s predictions about some new features, the answers may not be in balance with the reality [Nielsen, 1993, p. 209].

4.3. Interviews

Many times interviewing is the most important data gathering method. Other methods are valuable but to really know what a user is thinking one has to ask him or her about it [Kuniavsky, 2003]. Interviewing, like questionnaires, is an indirect method [Nielsen, 1993, p. 209]. There is no way to make sure that users are telling the truth or that they remember exactly what happened in the use situation. People have the tendency to answer in a way that they think they are supposed to [Nielsen, 1993, p. 213]. This has to be taken into consideration when analysing the results. Asking questions and listening to answers helps us
to understand what the key issues in various use contexts are and how certain things are done.

There are two main styles of interviews: structured and unstructured. In a structured interview the questions are defined beforehand and so is the order [Hiltunen et al., 2002, p.68]. An unstructured interview is completely different from a structured interview. In this interviewing style the most important and determining factor for the interview is the interaction between the interviewer and the interviewed. Also the purpose is wider and one goal of the interview is to identify those issues that need more specified examination. This style fits well in situations when one does not know exactly what one is looking for [Nielsen, 1993, p. 211].

Interviewing is a useful method when users’ subjective satisfaction and possible anxieties are in the focus [Nielsen, 1993, p. 209]. This is many times the case with games. We need to get the information about the satisfaction of players in order to be able to analyse the playability of that game.

4.4. Gameplay diaries
In a diary study participants keep a diary about their actions that are involved with the researched issue. The diary method is the best when used for fully functional products [Kuniavsky, 2003, p. 369], to be filled in when participants play or think about the game that is being evaluated.

There are many reasons for selecting the diary study method. The biggest reason is that the evaluator cannot be present all the time and the presence of an evaluator also might affect the results [Kuniavsky, 2003, p. 369]. There might be very interesting and valuable things happening at times when the evaluation is not active [Jones and Marsden, 2006, p. 141]. So the diary gives participants the possibility to write down all the things that they want to and when they want to (preferably right away when that action has happened).

The duration of the study is one of the most important things that has to be determined. It should be long enough that any usage patterns can be seen but not that long that it wears down the participants [Kuniavsky, 2003, p. 370; Jones and Marsden, 2006, p. 141]. If it is expected that the game is played every day, it cannot be expected that participants have the energy to keep on writing the diary more than a week. The frequency of adding a new entry into the diary should also be instructed. If one does not give any instructions how often to write into the diary, it often leads to poor results [Jones and Marsden, 2006, p. 141]. Giving straight guidelines when to fill in the diary will teach the participants a routine that is easy to follow.
The diary method can be valuable because diaries can provide good knowledge at a low cost, and it is also the only geographically distributed qualitative data gathering method [Kuniavsky, 2003, p. 370]. Of course also this method has problems. The diaries are hard and time-consuming to go through and try to figure out what are real problems and how serious they are from the story that the participant is trying to tell the evaluator [Jones and Marsden, 2006, p. 141]. One has to remember that diaries are never objective. They include information that is not relevant, and people may exaggerate or underestimate. It is the evaluators’ task to see through these and remain objective.

4.5. Observation

Observation is a method where end-users are observed by the evaluator. Observation can be done before the actual product is ready. Then the target of observation is the context and the methods that people use instead of the product they are going to develop. This kind of observation is more common in software development processes than in games. In games this does not happen that often because games are not designed to perform any specific task. The game usually is an interactive prototype when observation is done and the setting is somehow controlled [Jones and Marsden, 2006 p. 199-204].

There are several ways to save the information that is gained from the observation. One is of course making notes at the same time when a participant is doing tasks that were given to her or is freely familiarizing with the game. This is usually not enough. It is hard to observe the participant so carefully and at the same time write all observations down. Also then the context of that event will be most probably lost (the situation where the player was in the game). Recording observation sessions with a video camera is very helpful [Jones and Marsden, 2006 p. 199-204]. The best style is to use two separate cameras so that one is recording game events and the other one the player. This way the player’s reactions can be targeted to specific game events. Still evaluators usually also need to know what players are thinking but without interrupting the flow experience they may be experiencing in the game. One popular method for this is the think-aloud technique (e.g. [Nielsen, 1993; Dumas, 2003]). This technique is used in playtesting [Fullerton et al., 2004, p. 201].

The challenges of observation include for example disturbing the playing situation and in the process unintentionally directing the player in some way. The observer should withdraw oneself as far from the subject as possible to
prevent this from happening [Jones and Marsden, 2006 p. 199-204]. It is also important to create a comfortable environment for the player, who should not feel like a lab rat.

Crabtree et al. [2006] have raised some possible challenges that the researcher has when evaluating pervasive applications “in the wild” with ethnography.

- Users are often mobile.
- Interaction often involves small displays.
- Users often interact with invisible sensing system.
- Interaction is distributed across different applications and devices.

One way to overcome these challenges is to use video recordings with system logs. Portable videocameras already have been used in some pervasive game evaluations [Crabtree et al., 2006].

### 4.6. Log data

Log data is the data that can be collected from the game server. Every action that a player does in a game can be recorded and used later to support evaluation. Usually log data is not used on its own but rather to support the information that has been collected using some other evaluation method. This is because log data is raw data [Kuniavsky, 2003, p. 403]. With the log data it is possible to get precise knowledge about who did what and when [Nielsen, 1993, p. 216]. Still there are also some problems with it, like sometimes only actions are recorded. In Coup’s case the data about those players who only browsed the game is left uncaught.

There are some problems with log data that should be identified. For example it may seem a great idea to collect every player’s individual actions from the game but that may be impossible because tracking the same person over different sessions is hard. This was the case with Coup.

Log data also brings ethical issues and privacy concerns into the scope [Nielsen, 1993, p. 218]. Is it acceptable to collect information about players’ individual actions? Many people do not like the idea that their actions will be followed [Kuniavsky, 2003, p. 408]. When doing an evaluation one must make sure to tell the participants how the collected information is going to be used.

### 4.7. Focus groups

Focus group is a convenient method that can be used both before the product has been designed and after it is already finished [Nielsen, 1993, p. 214;
Focus groups are basically group interviews that are targeted for a specific group of people [Kuniavsky, 2003, p. 201]. Usually the size of the group is from six to nine users [Nielsen, 1993, p. 214]. These discussion sessions are run by a moderator whose responsibility is to keep the discussion in focus and make sure that everyone gets their turn to speak. Focus groups are good for bringing up spontaneous reactions and ideas [Nielsen, 1993, p. 214]. They also fit well for situations when it is needed to get information about people’s attitudes and perceptions [Kuniavsky, 2003, p. 203]. This is why it fits well for the concept phase of the game development when people’s attitudes towards that specific idea are at scope. With this method in early phases of the game development it is also possible to get new ideas for the concept.

Focus groups are not good for revealing usability issues [Kuniavsky, 2003, p. 204]. Focus groups also do not replace surveys. They cannot be generalized to a larger population [Kuniavsky, 2003, p. 204]. There are also some problems when using focus groups. First, in order to discussion to flow fluently and effortlessly at least six users are needed in the discussion [Nielsen, 1993, p. 215]. It also may be quite hard to keep the discussion in focus and to make sure that no one takes too much control over the discussion.

When one prepares a focus group, a list of issues that should be covered in the discussion should be made. Also it is good to think beforehand what the goals for that discussion are, and what kind of information one wants to gather.
5. Analysis of the results – Different methods for evaluating playability

In this chapter I will go through different evaluation methods in the light of the case studies and the example games. The special focus is on what evaluation methods were used in the example game evaluations, what kind of data the evaluators managed to get with the methods, and the usefulness of the methods according to the evaluators. The games are quite different from one another but what is common to all of them is that they still are pervasive games that are played with a mobile device.

5.1. What playability problems can be found using different methods?
Different data gathering methods have their pros and cons. This is why they also can detect different playability problems. I will look at each method in the light of the example games in order to find special aspects that the methods can detect. Table 9 presents methods that were used in the example game evaluations. In the case of Songs of North the evaluation was done in two phases, first in game concept phase when the game did not exist yet, and later in the game prototype phase by letting participants play the actual game.
Table 9 Methods that were used in example pervasive game evaluations.

<table>
<thead>
<tr>
<th>Method</th>
<th>Heuristic evaluation</th>
<th>Questionnaire</th>
<th>Interview</th>
<th>Focus group (discussion)</th>
<th>Gameplay diary</th>
<th>Field observation</th>
<th>Log data</th>
<th>Did participants play the game for real?</th>
<th>How long did they play?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Botfighters</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
<td>it was different for every participant</td>
</tr>
<tr>
<td>Coup</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>Yes</td>
<td>about a month (four sessions)</td>
</tr>
<tr>
<td>Day of the Figurines</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
<td>24 days</td>
</tr>
<tr>
<td>Epidemic Menace</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
<td>two days (three sessions)</td>
</tr>
<tr>
<td>Feeding Yoshi</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>Yes</td>
<td>a week</td>
</tr>
<tr>
<td>Garden of Earthly Delight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>Hot Potato</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>Hot Potato (in prestudy phase)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>Insectopia</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
<td>about a couple of days</td>
</tr>
<tr>
<td>Pirates!</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>Yes</td>
<td>less than 4 hours</td>
</tr>
<tr>
<td>Savannah</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>half a day each group</td>
</tr>
<tr>
<td>Songs of North</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No / Yes</td>
<td>- / for two weeks</td>
</tr>
<tr>
<td>Spy Blob</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>Spy Blob (in prestudy phase)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>Uncle Roy All around You</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>Yes</td>
<td>several hours a day for one or two weeks</td>
</tr>
<tr>
<td>Your Story</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>Yum Yum Sheep</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>No</td>
<td>-</td>
</tr>
</tbody>
</table>

I have gathered in this table also the information on how long the participants played the game, because it affects how well the participants got acquainted with the game and also what data gathering methods were possible to use. In some cases the participants did not play a working version of the game at all (Garden of Earthly Delights, Hot Potato, Spy Blob, Your Story, and Yum Yum Sheep). In these cases the evaluation was done in the game scenario or paper prototype phase.
5.1.1. Heuristics

Heuristics were used only in the case studies *Coup* and *Insectopia*, or at least none of the other game studies reported on using this method. In both cases, the heuristic set that was used was the playability heuristics by Koivisto and Korhonen [2006]. In *Coup*’s evaluation we used heuristic evaluation for alpha version of the game. It already had all the functionalities but they were not tested properly yet. The heuristic evaluation was done by the project members.

The heuristic evaluation of *Coup* took quite a lot of effort and it was time-consuming. *Coup* is a multiplayer game and it is based on player interaction so project members had to agree on certain times when to play the game. This is somewhat different than in the traditional usability heuristic evaluation. When doing a heuristic evaluation for a game it is not enough just to look at the screen shots (e.g. [Nielsen, 1994, p. 30; Nielsen and Mack, 1994, p. 25-30]), one actually has to play the game. The playability heuristics include issues which would be impossible to assess simply by looking at images of the game.

In the evaluation of *Coup* it was not enough that only a couple of project members played the game because of the critical mass problem. All the project members were in different locations so everyone wrote down the problems they had discovered and then we discussed them via e-mail. Finally we combined the findings into one file which was sent to game developers. In *Coup*, heuristics were used as a formative evaluation method (e.g. [Montola et al., 2005]), in other words to give developers more insight about the problems that needed to be fixed before the game was launched.

In *Insectopia* doing a heuristic evaluation was somewhat easier than with *Coup*, because we did not have to pre-plan so much how to conduct the evaluation. This was because *Insectopia* is not a multiplayer game. Heuristic evaluation of *Insectopia* was done by three project members and it took only about two hours (partly because of the technical problems that occurred during this evaluation phase so that we were not able to go through the whole game).

Heuristics proved to be helpful for evaluating playability in very early phases of the game as shown by our experiences from *Coup*, but they also worked well for evaluating a final game product like *Insectopia*. In both cases we found many problems (both technical and playability problems). Of course in *Coup*’s case we found more problems because the game was not yet finished. The developers got valuable information on how to improve the game and what kind of technical problems occurred and when.

In the evaluation of *Insectopia* there was something wrong with the server during the evaluation so we could not test all the features of the game. This might have been avoided, if we had contacted the game developer before we
started to evaluate. The reason why we did not see this necessary was that *Insectopia* had been published already a couple of months before the evaluation, so we thought that it would work. Also it had worked without problems just a couple of days earlier. Still this shows that one can never be quite sure if the game is going to work or not. It is always possible that some technical problems occur just in the middle of the evaluation.

### 5.1.2. Questionnaires

Using questionnaires proved to be popular to collect data for evaluations of games. Ten example game evaluations out of fifteen used questionnaires as a source for getting information. The game evaluations that did not use questionnaires were *Feeding Yoshi, Savannah, Botfighters, Day of the Figurines,* and *Garden of Earthly Delight.*

The strength of questionnaires is that despite the low number of participants it is still possible to get a lot of data. *Coup's* and *Insectopia's* evaluations proved this. Counting correlations between different variables was quite fast. The only problem we had in the analysing phase was to find out the relevant information. We made a lot of different diagrams and computed correlations in order to be able to see the most relevant issues to be looked at for the project’s goals. We faced this problem probably because of the novelty of pervasive games. We did not know what to expect beforehand. That is why we needed to ask many questions which is not usually recommended in questionnaires (e.g. [Nielsen, 1993, p. 213].

In most example game evaluations a questionnaire was used to gather background information or some overall knowledge about the phenomenon. For example in *Yum Yum Sheep, Your Story, Hot Potato* and *Spy Blob* questionnaires were used to gather information about who are the users of Bluetooth and what people think about that technology [Niemi et al., 2005]. This pre-study included 150 people living in or around Stockholm, Sweden. This information was then used in the actual evaluation, and based on the results two different age groups were selected: adults aged 28 to 46 and teens aged 15 to 16. In *Songs of North* the questionnaire was also used in a similar manner [Ermi and Mäyrä, 2004]. The main function of it was to collect participants for the player study and also to collect quantitative data of the playing habits and playing contexts.

In the light of the example game evaluations questionnaires can be a valuable source of data when collecting information about participants’ background and their opinions. Still it was obvious that it does not fit to be the only source of information when evaluating pervasive games. This is because
all the relevant issues are not usually known beforehand due to the still experimental nature of pervasive games.

5.1.3. Interviews
Interview was the most popular method among the example game evaluations. Out of fifteen example game evaluations eleven used interviews. Only Savannah, Day of the Figurines, Uncle Roy All around You, and Garden of Earthly Delights did not use this method. In the evaluations of the case studies Coup and Insectopia interviews had a big role. In both cases the interviews were semi-structured which means that the main questions were thought out beforehand but the interview session was quite open and the participant’s answers had an effect on the questions. If necessary more clarifying questions were asked.

In the evaluation of Botfighters interview was the only method that was used [Bjerver, 2006]. The reason for selecting interviews in that evaluation was the lack of participants (Bjerver managed to get 7 participants altogether and 2 of them were staff members) but also the want to get qualitative information about the game. The interviews were conducted via telephone except for one which was done via chat. According to Bjerver [2006] conducting an interview this way worked out well even though he did not have face-to-face contact with the participants.

The experiences from the example games show that interviews are a good source of information. Even if some participants were not as talkative as others, evaluators were able to get much valuable information. It was proven that the role of the interviewer is very important especially if the participant gives very short answers.

5.1.4. Gameplay diaries
Two example games used gameplay diaries in their evaluation: Coup and Feeding Yoshi. In Coup we gave the opportunity for the participants to keep an online gameplay diary on the game webpage. First they had to register to the webpage and then they could write their own comments there whenever they wanted to. The participants were not allowed to see what other participants had written so their comments were strictly confidential between the participant and the game developers.

What we noticed about gameplay diaries was that people did not like to fill them in. No-one volunteered to fill in the diary regularly. Still some participants did write short comments. These comments were not concentrated that much on gameplay itself but rather to the technical problems.
In *Feeding Yoshi*, a daily diary was used to get more insight into play and the players’ other daily activities [Bell et al., 2006]. This information together with the information that was gained from the interviews made it possible for the researchers to detect gaming patterns of each team. So especially for evaluating pervasiveness of the game these two methods seem to be efficient to tell how well the game adapts to player’s daily life.

Bell et al. [2006] did not separate what results were gained using which method. Apparently the interviews and the gameplay diaries produced quite similar results. It seemed that in this evaluation using gameplay diaries worked out much better than in the evaluation of *Coup*. One reason for this may be the fact that in *Feeding Yoshi* the players were known beforehand [Bell et al., 2006] so their commitment to the evaluation was higher and therefore they perhaps kept the diary more regularly and carefully. Of course, if there are enough resources to put into evaluation, one can also try to increase the commitment by offering gift tokens to participants. In the evaluation of *Coup* we did offer participants free movie tickets but still that was not enough to make them to commit in writing a gameplay diary. The two movie tickets were not enough for this time consuming task. Trusting that participants would volunteer in writing the gameplay diary in exchange for movie tickets proved to be a failure tactic even though we do feel that it was valuable at least to give the players the opportunity to keep a diary or write about their feelings towards the game. Some problems were learned from the diaries that might have been missed out entirely without them.

Gameplay diaries can be a valuable source of data if one can make participants to commit in writing the diary regularly. In *Coup*’s case this did not work because we did not know the players beforehand, but in the evaluation of *Feeding Yoshi*, gameplay diaries were one of the most important data gathering methods.

5.1.5. Field observation

Field observation, or ethnography as some researchers (e.g. [Crabtree et al., 2006]) like to call it, was used in seven example game evaluations, *Savannah*, *Day of the Figurines*, *Feeding Yoshi*, *Epidemic Menace*, *Uncle Roy All around You*, *Garden of Earthly Delights*, and *Pirates!*. In *Feeding Yoshi* and *Pirates!*, however, the role of the observation was more or less just to support the data that was gathered using other methods. What is in common to half of the games is that the gameplay happens in a small, restricted area (*Savannah*, *Epidemic Menace*, *Pirates!*). In the other games the location of the players can be anything which is quite a challenge for field observation.
In Savannah the data was collected using video observation together with an extended version of the system log that recorded all the things that happened during the game. In video observation two cameras were used to follow one child. The children that were chosen to be followed also wore a radio microphone so that the researchers could record their conversations. This method showed to be rather good especially when the target group was children. It enabled getting valuable data without having to ask from the children directly. A similar method was used in the evaluation of Uncle Roy All around You. Twelve street players had a video camera that they carried around as they were playing the game [Crabtree et al., 2006]. Also in this study the method seemed to work well together with other methods to reveal important aspects of playability. Still the devices used in these two evaluations are quite expensive (many video recording devices are required). Furthermore, analysis is time-consuming so this kind of observation method is not possible in every evaluation situation.

In Day of the Figurines [Crabtree et al., 2007] the target of observation was not the players but the moderators of the game, which made the problem of players’ location to disappear. The game is based on sending and receiving SMS text messages on mobile phones. The game moderators control the game by reading the incoming text messages and replying to them appropriately. Observation made it possible to get information from the moderators’ point of view but lack the players’ perspective. From the content and number of text messages it is possible to make some assumptions about the playability of the game as well but they are only very rough assumptions.

The evaluation of Epidemic Menace was mostly based on field observations. Players were playing the game while four observers wrote down their observations about the player-environment, player-player, player-devices, and player-game master interactions [Ohlenburg et al., 2006]. The issues that were examined in this field study were determined beforehand, and they had a certain style to write down the observations. A player feedback discussion and a questionnaire followed after the field study so the observations were not the only data that was used to evaluate the playability of Epidemic Menace. In this case the field observation gave a too optimistic picture, because the participants learned quickly how to use the devices and how to play the game, so the game session seemed very fluent and successful. Still that was only a surface; the discussions and the questionnaire revealed issues that would not have been noticed in field observation.

In Garden of Earthly Delights the observation was done while the participants were playing the game with a paper prototype. Each test session
was videotaped, and a transcript of how the session progressed was written afterwards [Koivisto and Eladhari, 2006]. The evaluation showed that especially usability problems were quite easy to detect by observing the participants. Still they were also able to gather valuable data from pervasive aspects and gameplay as well. The fact that participants were able to play the game (even though only with a paper prototype) made the game concept more real to them and helped them to understand the pervasive aspects of the game.

5.1.6. Log data
Log data was used in four example games, Coup, Feeding Yoshi, Savannah, and Uncle Roy All around You. Experiences from all of these games showed that log data is very good for gaining basic knowledge, the facts about the gameplay. Log data was used to get automatically accurate information about when certain actions happened and how many players and when those players were attending the game.

In the case study game Coup, there were initially in total 23 event types that were recorded but one of them was removed after the first play session. Appendix E presents a list of the recorded events. We were able to get a lot of data from the log files but still we did not feel that we could make any assumptions based on them. For example, we had a problem that we could not tell whether a player is in the game or not if she did not perform any actions there but just looked at the current situation. Still for validating the data that was gained using other methods this worked well.

Log data was also used in the evaluations of Feeding Yoshi, Savannah, and Uncle Roy All around You. In all of these the role of the log data was more or less supportive to other data gathering methods. In the evaluations of Savannah and Uncle Roy All around You log data was used together with video observation and the findings were combined together later so they could be processed at the same time. In Feeding Yoshi the log data was used successfully to show patterns in the participants’ playing style.

The usability of log files can be really poor. For example in Coup’s evaluation we had to do some programming to help finding the relevant data from the file. Of course one could search for the relevant data manually as well but that would take a lot of time since all the actions are recorded. To help solving this problem a special log data recording system was used in the evaluation of Savannah [Benford et al., 2005b]. In this system the log data was presented in a graphical style with a graphical user interface (Figure 21), which made the analysing phase much easier.
The experiences from the example game evaluations showed that log data is a valuable source of information for playtesting especially for gaining more background information. Still it was clear that it should not be used alone without other methods. For example the data that can be collected from the logs does not give any reasons for certain actions. One can only make assumptions what the reasons behind the actions are. It can also give misleading information that could lead into misinterpretations if more specified information is not available.

5.1.7. Focus groups

Focus group discussions were used in six example game evaluations: Yum Yum Sheep, Your Story, Hot Potato, Spy Blob, Songs of North, and Garden of Earthly Delights. In all of these, focus groups were used in early phases of the game development process when only game scenarios existed. Focus groups were used in Yum Yum Sheep, Your Story, Hot Potato, and Spy Blob to decide which one of these game ideas would be the best to actually implement [Niemi et al., 2005]. The game scenarios were presented as text and there were two scenarios of all four games. The first one was an ambiguous scenario – a description of a non-player who observes gaming without knowing what was going on. The second scenario was an in-context scenario about the rules and goals of the game.

The game designers of Songs of North used focus groups (they refer to this as group discussion) when they evaluated gameplay scenarios for the game [Ermi and Mäyrä, 2004]. The speciality in this evaluation was that instead of explaining the game concept for the focus group they had prepared comic strips...
to do that for them. They thought that comic strips would be more illustrative and quite easily comprehended by the participants, thus better than plain explanation [Ermi and Mäyrä, 2004]. Focus group discussions were usable in this evaluation but it would have not worked alone without other methods. Also they did not get precise knowledge about all the issues they explored because there was some overlapping between concepts and not all of them were presented in the best possible way [Ermi and Mäyrä, 2004].

Focus groups can be a valuable source of information in the early game development phase when the goal is to get knowledge about the game concept. Still the method may not work as well in every game development process. Especially in pervasive game development it may be hard for participants to imagine the game that the developers have in their mind when they only hear a description of it. People do not have much experience about pervasive playing so it may feel strange and awkward at first. This can lead to more negative attitude towards the game idea than what would be the case if they actually could play the game. Perhaps the best way to use focus groups is to have multiple game concepts that are discussed in focus groups, as was done in the evaluations of *Yum Yum Sheep*, *Your Story*, *Hot Potato*, and *Spy Blob* [Niemi et al., 2005].

5.2. Methods in different evaluation phases

Evaluation can and should be done throughout the whole game development process. The methods that are sufficient and effective in different phases vary. Some methods fit better some phases than others. In the example games, evaluations were done in four phases of game development: game scenario, paper prototype, beta version, and final game product (Figure 22). The phases are added on Figure 19, p. 40. Some information about what method fits best to what phase already came out in the previous section. In this section evaluation requirements of different game development phases and methods that meet these requirements the best are presented.
Table 10 shows in detail when and where the evaluations of the example games were done and in which game development phase. As can be seen from the table, most of the example games were developed and evaluated in Finland, Sweden, or United Kingdom. An exception to this was *Epidemic Menace* which was evaluated in Germany.

The game development phase where the evaluation took place varied among the example games. Game scenario, beta version, and final game product phases were the most common phases to do a game evaluation. Only one game, *Garder of Earthly Delights*, was evaluated in the paper prototype phase. *Songs of North* was evaluated twice; first in the game scenario phase and then in the beta version phase.
Table 10 Detailed information about the evaluations of the example games

<table>
<thead>
<tr>
<th>Game</th>
<th>How many participants took part in the evaluation?</th>
<th>When was the evaluation done?</th>
<th>How long did the evaluation take?</th>
<th>Where was the evaluation done?</th>
<th>What game development phase was the evaluation done?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Botfighters</td>
<td>5 plus 2 staff members</td>
<td>November 2005</td>
<td>one week total, one interview 40 min to 1 hour</td>
<td>Sweden</td>
<td>finished</td>
</tr>
<tr>
<td>Coup</td>
<td>9 (questionnaire), 3 (interview)</td>
<td>October-November 2006</td>
<td>interview lasted max. 1 hour</td>
<td>Tampere, Finland</td>
<td>beta</td>
</tr>
<tr>
<td>Day of the Figurines</td>
<td>85 players, 6 operators</td>
<td>not said</td>
<td>24 days</td>
<td>UK</td>
<td>beta</td>
</tr>
<tr>
<td>Epidemic Menace</td>
<td>8</td>
<td>August 2005</td>
<td>not said</td>
<td>at the campus in, Germany</td>
<td>beta</td>
</tr>
<tr>
<td>Feeding Yoshi</td>
<td>16 players</td>
<td>not said</td>
<td>not said</td>
<td>Glasgow, Derby, and Nottingham, UK</td>
<td>finished</td>
</tr>
<tr>
<td>Garden of Earthly Delight</td>
<td>8</td>
<td>September 2005</td>
<td>2 days</td>
<td>Sweden</td>
<td>paper prototype</td>
</tr>
<tr>
<td>Hot Potato</td>
<td>10 youths and 10 adults</td>
<td>not said</td>
<td>not said</td>
<td>Stockholm, Sweden</td>
<td>game scenario</td>
</tr>
<tr>
<td>Insectopia</td>
<td>11 (questionnaire), 7 (interview)</td>
<td>October 2006</td>
<td>interview lasted max. 1 hour</td>
<td>Tampere, Finland</td>
<td>finished</td>
</tr>
<tr>
<td>Pirates!</td>
<td>13</td>
<td>August 2000</td>
<td>not said</td>
<td>conference hall, Bristol, UK</td>
<td>beta</td>
</tr>
<tr>
<td>Savannah</td>
<td>36 children</td>
<td>not said</td>
<td>three days</td>
<td>empty school playing field, UK</td>
<td>finished</td>
</tr>
<tr>
<td>Spy Blob</td>
<td>10 youths and 10 adults</td>
<td>not said</td>
<td>not said</td>
<td>Stockholm, Sweden</td>
<td>game scenario</td>
</tr>
<tr>
<td>Songs of North</td>
<td>1) 1917 web questionnaire, 6 interview</td>
<td>1) 2003</td>
<td>1) the scenario-based interviews lasted few hours</td>
<td>1) in Tampere, Finland</td>
<td>1) game scenario</td>
</tr>
<tr>
<td></td>
<td>2) 19 test-players, 17 answered to questionnaire</td>
<td>2) not said</td>
<td>2) not said</td>
<td>2) in Tampere, Finland</td>
<td>2) beta</td>
</tr>
<tr>
<td>Uncle Roy All around You</td>
<td>several thousands (at a time 10 players and 15 online players)</td>
<td>not said</td>
<td>not said</td>
<td>three cities in UK</td>
<td>finished</td>
</tr>
<tr>
<td>Your Story</td>
<td>10 youths and 10 adults</td>
<td>not said</td>
<td>not said</td>
<td>Stockholm, Sweden</td>
<td>game scenario</td>
</tr>
<tr>
<td>Yum Yum Sheep</td>
<td>10 youths and 10 adults</td>
<td>not said</td>
<td>not said</td>
<td>Stockholm, Sweden</td>
<td>game scenario</td>
</tr>
</tbody>
</table>
5.2.1. Game scenario

Game scenario means that only the game idea is somewhat fixed. Nothing concrete is necessarily ready at this phase and the idea has to be presented to the players either verbally as was done in the evaluations of Hot Potato, Yum Yum Sheep, Your Story, and Spy Blob, or graphically as in the evaluation of Songs of North. In all of these example games focus group discussions were used when evaluating the game scenarios. Focus group discussions can give deep information about the game concept, and evaluators have the possibility to lead the discussion in the desired direction. These reasons motivated using the focus group discussions in these example game evaluations. The participants for all of these example game evaluations were selected through web questionnaires which also provided background information about the subject and the participants for the evaluators.

In Hot Potato, Yum Yum Sheep, Your Story, and Spy Blob game scenarios were presented to the participants verbally on sheets of paper. There were two scenarios for each game, one that described a non-player situation who observes the game and another, where the rules and goals of the game were explained from the perspective of a player as well as from the perspective of a non-player [Niemi et al., 2005]. The goal in these evaluations was to get information on what is acceptable in games and what is not. The ethical considerations and revealing information to strangers came out in the discussions, and the evaluators got a lot of insight about where the thin line goes between what is acceptable and what is not. Based on the focus group discussions one of these games, Hot Potato, which was considered to be the least threatening game concept, was selected for further development [Niemi et al., 2005].

In Songs of North the game scenarios were presented as comic strips which worked out pretty well according to the researchers. It gave the participants a good overview of the game concept [Ermi and Mäyrä, 2004]. What seemed to be a problem was that the strips were designed for the purposes of the game designers, not for the study. This caused that the focus of some strips was not clear and the gameplay from a player’s perspective remained vague. Still the comic strips made the focus group discussions easier in a way that the game concept became more visible to the participants.

These examples show that pervasive games should be evaluated already in the game scenario phase because the game concepts can be very extraordinary and sometimes even too extreme to be accepted by people. It is good to get some pre-understanding about the concept at hand. It appears that
both written and graphical presentations of game scenarios are valuable as long as they are designed with evaluation and participants in mind.

5.2.2. Paper prototype

Paper prototyping is the phase that comes after game scenario. In this phase the game idea and the rules are already quite clear. The game is visualized and a paper version of the game that is more or less playable is built. Sometimes this phase is skipped, but it is good to have a paper prototype. If the design errors can be detected already in this phase, it saves costs and time.

Paper prototyping is not that usual in pervasive game development processes. Paper prototyping was used only in one example game, Garden of Earthly Delights [Koivisto and Eladhari, 2006]. Paper prototyping can be quite time consuming. It may be hard to make a prototype out of pervasive games that usually are very mobile and their gameplay is based on communication between players.

Garden of Earthly Delights is not a simple game but still the researchers managed to evaluate it using a paper prototype. The advantage in paper prototype phase evaluation is that participants are able to play the game for real even if the design is still much unfinished and nothing has been implemented yet. Koivisto and Eladhari [2006] also noticed that participants who played the game had more positive opinions about the gameplay than people who did not play. This suggests that pervasive game concepts are still so unfamiliar to people that they tend to have a bit negative reaction towards them at first but when they try to play these games their opinions change.

Paper prototypes are not necessarily that good for finding usability problems. In the evaluation of Garden of Earthly Delights only few usability problems were discovered. The reason for this might have been the decision to use simplistic drawings instead of more realistic graphic. This was a conscious decision which made participants not to concentrate too much on the graphical aspects but actually on the gameplay itself [Koivisto and Eladhari, 2006]. Even though paper prototyping did not appear to be a good method for detecting usability problems, it showed to be effective for detecting gameplay and pervasive problems which are crucial to find early on in the game development process. On top of this, by using the paper prototyping method for real participants it was possible to gather information about the player attitudes and behavior.

Koivisto and Suomela [2007] compared the paper prototyping evaluation of Garden of Earthly Delights to results of two other pervasive games evaluations which also used prototyping. It was noted in all the cases that prototyping gave
a lot of improvement ideas and revealed problems in the game design. Still there was a clear difference between the results based on what kind of prototyping was used.

What was important in this study was that it revealed that paper prototyping cannot detect all playability problems. Even if the evaluators present the scenarios with the paper prototype it still does not reveal problems that are not somehow expected or fall outside of scenarios.

Overall experiences from the evaluation of Garden of Earthly Delights show that paper prototyping is a method that requires a lot of time and effort especially if the gameplay is complex. It was estimated that 120 hours were spent doing this paper prototype evaluation in Garden of Earthly Delights [Koivisto and Suomela, 2007]. Still it is possible to use paper prototyping as a tool for improving a pervasive game. Participants can imagine pervasive features like for example moving in a real world very well even if there is only a paper prototype of the game available [Koivisto and Eladhari, 2006]. This suggests that paper prototypes should be used also in pervasive game development process.

5.2.3. Beta version

Beta versions are almost ready games. They have all the functions and features that are supposed to be in the final game product as well. Actually there is not that much that separates beta version games from final game products, except maybe the intention that their development is going to continue. Beta version testing showed to be popular, five example games were evaluated in their beta phase. These games were Coup, Day of the Figurines, Epidemic Menace, Pirates!, and Songs of North.

The evaluations in the beta phase were pretty similar. The participants were able to play the game for real so the evaluations were based on their actual experiences from the game. The evaluation phase did not have much effect on the results, and practically all the data gathering methods are available for use. The only method that was not used in any of these example game evaluations was focus group discussions. Instead, focus groups were the most popular method in the game concept phase. This shows that it is most suitable for the early development phases to uncover flaws in the game idea.

The heuristic evaluation method was used only in the evaluation of Coup. This does not mean that it would not be suitable for the playability evaluation in this game development phase but rather that the method is still quite new in a game context (e.g. [Fedoroff, 2002; Desurvire et al., 2004; Sweetser and Wyeth, Koivisto and Korhonen, 2006]. There is a need for heuristics that would take
pervasive aspects of the game better into account. Current heuristic sets (especially the one made by Koivisto and Korhonen [2006]) can be used, and it is possible to detect some playability problems using them, but at the moment heuristic evaluation is not the best method to be used for a pervasive game evaluation due to the lack of proper heuristics.

5.2.4. Final game product
There were five games that were evaluated when they were finished, *Insectopia, Feeding Yoshi, Savannah, Botfighters*, and *Uncle Roy All around You*. It is possible that some of these games are still improved but at least the papers indicated that these were somewhat final game productions. In these cases the goal of the evaluation was to measure how the pre-set goals were met. Some of these were developed during a research project and the evaluation was part of it.

Like in the beta version phase, also in this phase all the methods can be used. Still also in this phase none of the example game evaluations used focus group discussions. The reason for this might have been the same as in the beta version phase: in focus group discussions participants can affect each other’s opinions, which is not a desirable outcome when evaluating playability and individual player experiences.

Also in this phase the heuristic evaluation method was used only in one game evaluation, *Insectopia*. Other methods seemed to be adequate for evaluating playability in this phase.

5.3. The validity of information gained using different methods
It is important to make sure that the information gathered is as valid as possible. Sometimes it may be hard to figure out whether or not the results are accurate or not. The reason for invalid results can be for example that there was something wrong with the evaluation arrangements like that the questions might have been hard to understand and to answer.

In the evaluation of *Coup* we noticed that data gathered from other sources apart from log files yielded quite similar observations. For example, many participants complained in the interview about the unresponsiveness of other players and long waiting periods. That also became evident from the questionnaire. Figure 23 shows that seven out of nine *Coup* players felt that there is too much waiting in the game.
Figure 23 *Coup* players felt that there were too much waiting in the game (N=9).

The questionnaire and the interviews showed that almost all of the players felt that they were alone in the game as can be seen from these player quotes.

…So I really couldn’t play properly that because you always had to get some approval from the emperor, and it could take a day to get some approval… (male, 24)

You just waited there that those others would try to do something… (male, 25)

I had some difficulties to maintain the interest because there were so few players and then quite a many of them were not that active. So my own game stood still when I couldn’t really do anything else when those others didn’t come into the game to give an approval or something… (male, 24)

However, log data showed otherwise. According to log data, players were in the game almost at the same time. If the analysis had been based solely on the log data it would have appeared that the players did have the possibility to interact with each other and not having to wait so long.

Based on the evaluation of *Coup*, questionnaires and interviews give a lot of valuable and usually correct information about the player experience. This can be said because the information that was gained using these methods was so consistent. On the other hand, it also showed that it is not sufficient to use solely log data in the evaluation.
In the evaluation of *Coup*, questionnaires and interviews gave consistent information. This was not the case in the evaluation of *Insectopia*. *Insectopia* is a quite simple game so players can easily start playing the game even without reading the manual first. The fact that most of the participants did not read the manual at all became very clear in our evaluation. The players could not understand or had missed entirely the insect feeding functionality of the game. About half of the interviewed participants did not realise at all that they were supposed to feed the insects. This comes out very clearly in the following two quotes from interviewed participants.

*The feeding didn’t come out in any way in this game. I had no idea that I was suppose to feed them…* (female, 24)

*Hmm… how did the feeding actually work? I thought that it happened automatically when I did a search. Is there some specific place where you can feed them?* (male, 25)

This information differed from the information that we got from the questionnaire. In the questionnaire one question was about how easy the information given on the screen was to understand. Almost 70% of the participants replied that it was easy to understand (Figure 24). Still they had missed out totally one of the core functionalities of the game and we would not have noticed that without using also interviews.

![Figure 24 The participants felt that the information that was given on the screen was easy to understand (N=9).](image-url)
In the example game evaluations focus group discussions were used in the game concept phase which means that the actual games did not exist yet and the participants could not play the games for real. In the evaluation of the *Garden of Earthly Delight* [Koivisto and Eladhari, 2006] it was noticed that people who had playtested the game gave more positive feedback than those who participated only in the focus group discussions. This is probably because of the novel nature of pervasive games. Participants have difficulties in understanding this new way of playing so they look at it sceptically. So focus group discussions cannot predict what kind of reactions people will have towards the finished pervasive game.

The problem with focus groups is that one should be able to describe a concept without having anything concrete to show. This causes that the results are based on participants’ imagination and it can be far from the actual concept that was on designers’ mind. This happened also in one of the world’s most popular PC games *The Sims*. When Will Wright (the man behind *The Sims*) presented the idea to the focus group they hated the idea. It was also rejected by the product-selection committee [GIGnews, 2002]. As Will Wright said:

*Most people when they imagine that concept in their minds are not going to imagine something very compelling. So it’s the designer’s task to make that into a compelling activity with interactions, the way it’s visualized, etc.*

This shows that even though focus groups can give valuable information about the game concept it still does not necessarily predict what kind of a game it will be. Sometimes it is good to take the game concept further in the development process even if the focus group rejected the idea, but only if one believes in the game.

These experiences show that every method has its weaknesses and therefore none of them should be used alone. Interviews seem to be the most accurate method to use, so if it is necessary to choose only one method, then it should be an interview after a playtest period. Interviewing is the only method that gives an evaluator the possibility to ask clarifying questions and really find out what a participant thinks about the game. Also if the participant has difficulties understanding the question the evaluator can clarify it. The danger of misinterpretations is lower in interviews than for example in questionnaires.
6. Discussion

In this chapter I will go through the results and discuss what they mean for playability evaluation. I have also made an attempt to build a guideline for evaluating pervasive games; what are the special issues that should be taken into consideration when evaluating pervasive games.

6.1. How to evaluate playability of pervasive games?

Pervasive games are very different from traditional games and what makes evaluation even more demanding is that there is a lot of variation between pervasive games.

Figure 25 shows the modules of playability that were already shown in Figure 17 as part of the player experience. In this thesis the focus is on the playability evaluation of pervasive mobile games so the modules that are not relevant for pervasive mobile games have been excluded from discussion in the following sections.

![Figure 25 The playability modules of pervasive mobile games.](Image)

6.1.1. Gameplay

Gameplay is essential in games since it involves the core parts of games like rules. Gameplay problems are usually more difficult to detect since the focus
first goes to user interface which can be seen straight away when entering a
game. Gameplay problems on the contrary are not that obvious, and they are
noticed only after one has played a game for a while. When evaluating a still
unfinalized game, participants usually tend to concentrate more on usability
issues than gameplay issues.

Still it is not impossible to evaluate gameplay of pervasive games as
became obvious from the experiences of the example games. For example focus
group discussions where the game concept is described to participants through
a paper prototype or a game scenario can be valuable for getting information
especially about the gameplay issues.

When one wants more information about the gameplay issues it is best to
evaluate the game either in the beginning of the game development process or
after the game is almost finished. This is because then the focus can be really on
gameplay and not for example on the unfinished user interface. In the game
concept phase there is no working version of the game so it is easier for the
participants to concentrate on the idea of the game and give opinions about it.

6.1.2. Usability
Usability was the easiest to detect out of all playability modules. This may be
because usability evaluation has a long history, and evaluating usability issues
is quite straight-forward even though it demands expertise. Detecting usability
problems did not seem to be any more problematic with pervasive games than
with any other games.

6.1.3. Pervasiveness
Pervasive aspects depend on the game and how it breaks the magic circle. Still
what is characteristic for pervasive games is that they include context elements
into their gameplay. Therefore in the evaluation of pervasiveness the focus is on
these context elements and how a player has understood and experienced them.
Context elements can vary a lot. In Insectopia the context element is Bluetooth
Ids, and since it involves other people also ethical considerations become
important in the evaluation. There are also some other context elements that can
be used like weather, location, or time. The evaluation of pervasive games
should focus on these novel game elements; how they affect player experience,
do they increase the challenge level, do they make the game more interesting,
and so on.

Pervasiveness can be hard to evaluate since it is still so unfamiliar for both
evaluators and players. Evaluators may not know what are the most important
aspects, and especially how they affect a player. That is why interviewing is one
of the best methods to use when trying to gather information about pervasive elements.

It is important to evaluate pervasive aspects also in the early game development phase. Still this might be quite difficult. Participants may not understand the game concept properly when they do not have any earlier experiences about pervasive games. To make it easier for the participants to understand the game idea one should visualize it. In the evaluation of Garden of the Earthly Delights [Koivisto and Eladhari, 2006] a paper prototype was made out of the game to show the participants the pervasive gameplay. Paper prototyping proved to be a good way to evaluate especially pervasive aspects. It was also noted that when participants had the possibility to play the game, their opinions were more positive than after they only heard a verbal game description.

In Songs of North game scenarios were presented as comic strips to make the game idea and pervasiveness clearer to the participants. Still in this evaluation the comic strips did not work as well as was intended. Instead, verbal game scenarios can work better, especially if the focus is on some special pervasive aspect, for instance ethical considerations (what people find acceptable), like in the evaluations of Hot Potato, Your Story, Yum Yum Sheep, and Spy Blob.

Heuristic evaluation was used only in the evaluations of the case study games. Heuristics are a valuable design and evaluation tool but still they are not good for detecting pervasive issues because there is no valid heuristic set developed yet for pervasive games.

These experiences show that pervasiveness can be hard to evaluate because the scope is not necessarily clear even for an evaluator. My own experiences with pervasive game evaluations have shown that unexpected issues usually come out in the evaluation. Still there are methods that can be used to evaluate also the pervasive aspects as well as possible. Experimental and novel ideas like using comic strips in evaluation are needed, even though they may not work at first. Still they are needed to make the evaluation of pervasive aspects of games easier.

6.1.4. Mobility
Mobility was one big part of the example games and therefore mobility is also part of their playability. Mobility problems are usually quite easy to detect even without actual players. Mobility problems were not in focus in any of the example game evaluations so there is not that much information on how efficient different methods were to detect them. Still it seems at least based on
the case study game evaluations that mobility problems can be found quite easily.

A game has to be playable when evaluating mobility issues because they are so closely connected to the device like how a game handles incoming calls. Without the device they do not exist. This is why mobility problems can be found only in the last game development phases when the game is already implemented.

Mobile devices seem to be a good platform for pervasive games because they support many pervasive features by nature. They support already both asynchronous and synchronous communication, they are almost always accessible, and they already have the technology that supports including context elements in a game. This is why mobility is more an advantage for a pervasive game and mobile issues do not usually cause that much problems even though mobile gaming itself may not be so familiar for everyone.

6.1.5. Multiplayer aspects
Some of the example games were multiplayer games. For them multiplayer aspects and especially collaboration and communication between players are essential. None of the example game evaluations concentrated especially on evaluating multiplayer aspects but still they were part in the evaluation because they affect the playability.

In the example game evaluations, multiplayer aspects were mostly covered through interviews and questionnaires. Playing multiplayer games is familiar to people so issues like collaboration and communication are easy to cover using almost any of the data gathering methods. Korhonen and Koivisto [2007] have designed a special heuristic module for multiplayer games but all of the evaluations were made before this existed so I do not have any experiences with how efficient their heuristics are for evaluating multiplayer problems.

Observation can also be very valuable for detecting multiplayer problems. In some of the example game evaluations the evaluation was based on ethnography or observation. This makes it possible to get first-hand information about how multiplaying works in a game. When using an interview or a questionnaire one has to rely on the memory of a player instead of witnessing directly the situation itself. Interviews and questionnaires also work but they are not as reliable as observation. On the other hand observation can also lead to wrong conclusions. It is possible that the objective view of the situation is completely different than subjective experience on that same situation. In the example game evaluations Garden of the Earthly Delights, Pirates!, Epidemic Menace, Feeding Yoshi, and Uncle Roy All around You used both observation and
questionnaire and/or interview. This combination seemed to work well for multiplayer games where communication and collaboration have a key role.

6.2. Choosing data gathering methods

There are at least three different factors that affect which methods to choose: 1) the nature of the project, 2) the skills of the team, and 3) the culture and style of the company [Jones and Marsden, 2006, p. 113]. The nature of the project determines what methods are applicable. This includes the phase of the game development and the goals of the evaluation. The second factor, the team’s skills also have an effect on what data gathering methods should be used. For example, conducting a field study or doing an interview requires at least some sort of expertise. Otherwise it is possible that wrong conclusions are drawn or some valuable information is left unnoticed. Usually the problem with the non-expert evaluators is that they are not accurate enough, i.e., they do not understand the importance of little details [Montola et al., 2005]. The third thing that affects is company culture and style, which basically means that some companies may have some traditions in what methods to use and how. Jones and Marsden [2006] do not mention resources as one of the factors, but they are also one very important factor that affects what methods can be used. Usually this factor is the most important when lack of time and money forces to take shortcuts in evaluation.

Based on the example game evaluations it is recommendable that at least two different methods are used (usually one that gives qualitative data and the other that gives quantitative data). One evaluation method is not enough to give reliable information so different methods should be used to back up each other and prevent from misinterpretations. Determining what those two (or more) methods should be is more difficult. It appears that despite the expansion style(s) of the game, the methods varied. One thing that of course affected a lot was that the games were in different development phases when the evaluation was carried out.

Combination of log data, questionnaires, and interviews gave a lot of valid information and proved to be efficient to evaluate playability and player experience. Therefore this combination is recommendable to use if possible. Still in some situations, this combination can be too complex and time consuming to be used, especially in the middle of the game development process. The situation of course is different in every gameplay evaluation so methods should be selected to meet the circumstances and demands of that particular situation.
Interviewing was the most often used data gathering method among example game evaluations; it was used almost in every evaluation. It seems that interviewing fits all cases. The same was with questionnaires. There were no noticeable differences in their use between the expansion styles. One reason why these methods were so popular despite the game might have been the fact that they are probably the most well-known and used data gathering methods. They also fit very different situations depending on how they are used. By using open-ended questions it is possible to get a lot of qualitative data and also notice issues that would not be seen otherwise. Based on the experiences of the example game evaluations interviews and questionnaires are data gathering methods that should be used in every evaluation, especially if the evaluation is done for the final game product.

Games that expand the magic circle socially used most often combinations of interviews and focus group discussions. The exceptions for this were *Insectopia* (only interview), *Botfighters* (only interview), and *Day of the Figurines* which did not use either of these methods. The focus group method was used in many example game evaluations in the scenario phase, when the game idea was tested. It is quite understandable why the focus group method is so popular especially when evaluating socially expanded games, because it is not always obvious what is socially acceptable and what is not. For example Niemi et al. [2005] used this method to solve which of the four games (*Yum Yum Sheep, Your Story, Spy Blob, Hot Potato*) would be the most acceptable for the target audience. This evaluation showed how effective focus groups can be in detecting issues that include for example ethical concerns that people may have. Therefore focus group discussions fit well for evaluating the acceptance of socially expanding pervasive games in their early development phases.

The field observation method was used the most in those example games that expand the magic circle spatially. Another joint characteristic for these games is that most of them are event games, which means that they run only for a short period of time and need moderators. It may sound weird that games expanding the magic circle spatially use field observation, because their playing field is not fixed in the traditional sense. This could make field observation impossible. However, some of the spatially expanded example games happen in some pre-set area so using field observation is possible, although not in the traditional way. In some of the example game evaluations where field observations were collected (like in *Savannah*) the observations were recorded using video cameras attached to a couple of players. This way the evaluators could get information on what happened from the player point of view. Based on the *Savannah* evaluation such recording worked well, even though it is
technically more challenging than for example an interview. Still using this method could be a good idea especially when the game is targeted for small children.

A closer look at the example games showed that it is not easy to say which methods would be the most suitable for which pervasive game type. It seems that other issues are more important when selecting data gathering methods than the expansion style. One of the most important affecting factors was the development phase of the game. Also there have been so few pervasive game evaluations so far that the methods that are most suitable for them are not yet clear. More research is needed to find out what different data gathering methods really would be the best for which pervasive game type.

6.3. Things to remember when planning an evaluation

There are a lot of things that one should take into consideration when planning an evaluation of a pervasive game. These guidelines have been collected from the case study game evaluations as well as the studied game evaluation literature.

1. If the game is location-based, remember also to evaluate how the location of the player affects playability.

In location-based games like *Feeding Yoshi* and *Insectopia* the urban area where the game is played has a significant role in player experience. For that reason, it would be ideal to test the game in different locations simultaneously. If the game is played in several locations, one gets a better idea of how player experience is connected to the location where it is played. This has a lot of meaning when it comes to selecting market areas for that game. Use interviews, questionnaires or daily gameplay diaries to do this.

2. Use always more than just one evaluation method, preferably at least one qualitative and one quantitative method.

The experiences from the example game evaluations show that it is not enough to use only one method; it may lead to wrong conclusions.

3. Start evaluation already in early game development phases by doing focus group interviews and heuristic evaluations.

People can evaluate also concepts or ideas that are not in any way implemented. Group discussions and questionnaires can help determining what
pervasive game concepts are accepted by the target group, and not found too offending. Also heuristics and paper prototypes are valuable ways for getting information in the early game development phases. Focus group discussions can give valuable information about the pervasive and gameplay aspects, and heuristics especially on usability aspects. Evaluating the game in the early development phases saves time and money, and helps to set the focus of the game developers into right direction.

4. **Give participants the chance to play the game for several days to see how the game fitted into their lives.**

Participants have to have the opportunity to play the game as part of their everyday lives. Especially if the game expands the magic circle temporally, it is impossible to evaluate the game properly without allowing the participants to play the game whenever they choose to.

5. **If the game is a multiplayer game, make sure that there are enough players playing the game when the game is launched.**

In an experimental multiplayer game, the problem usually is to get enough players to play the game. To make sure that the lack of players would not hurt the player experience of those who are playing the game for real, recruit a sufficient number of players to play the game at least when the game is launched. This is important especially if the game demands collaboration.

6. **Have a backup plan especially if the game requires new technology.**

Pervasive mobile games usually require a certain type of mobile device to be used. An evaluator must take into consideration that people may not have the equipment that is needed to play the game. The evaluation of *Coup* showed that not that many potential players had the proper phone that was needed to play the game. Also downloading the game into a phone caused problems to some players. Technology can cause also other problems, for example the server can go down in the middle of the test period. An evaluator should be prepared for these when planning an evaluation of a pervasive mobile game.

7. **If the game is done as part of a research project, its exploratory nature should be stressed when the game is launched.**

If players know that the game was developed as part of a research project, they will understand that it is not necessarily refined, and that their feedback is
really needed. Also this may increase the players’ commitment to the game by making them feel important.

6.4. Limitations of this study and further research
This study involves many types of pervasive games that use mobile devices somehow in their gameplay. A problem when doing this study was the lack of specific knowledge about the example game evaluations. The information was collected from published papers that did not necessarily concentrate on the evaluation part. In the future, it would be great to study evaluation methods of pervasive mobile games even in more detail. Except of the case studies, the information that was used was combined from the published papers which did not in many cases describe the evaluation process in detail. Getting detailed information and also applying the same methods to other pervasive games would give more accurate information about pros and cons of those methods regarding to evaluation of pervasive mobile games. In the future, this kind of comparative analysis would be fruitful to conduct.

We also need a more comprehensive set of heuristics for pervasive games. Jegers [2006] has made an attempt for pervasive game heuristics but they need to be used in real game evaluations and compared against other methods in order to see how valid and extensive they are. At the moment there are many heuristics lists for games but they do not take into consideration pervasive features.
7. Conclusion

In this thesis I have collected information about pervasive games and especially about their evaluation. This is a novel attempt since so far in the field there are no extensive disquisitions made about the evaluation of pervasive games. This thesis offers an extensive overview to existing pervasive mobile games and their evaluations. Also the efficiency of different data gathering methods are shown, not merely comparing two methods against each other as has been done so far, but taking a broader approach and comparing the methods in all the game development phases.

I have discussed the phenomenon of pervasive games and how the special nature of pervasive games affects their playability evaluation. Playability is part of the player experience that can be evaluated in different ways. I have made an attempt to clarify these concepts of playability and player experience, since there is a lack of definitions that can be used in evaluation.

I have also gathered some pervasive game evaluation guidelines based on my experiences and publications about pervasive games. These guidelines give some instructions for how to avoid some complications that may occur in pervasive game evaluations.

The results show that playtesting with real participants is in many game development phases the best way to gather accurate data, but also other data gathering methods can be used as part of the evaluation. Still none of the methods seemed to be efficient enough when used alone; the possibility to make wrong conclusions out of that data grows too high. If one can use only one data gathering method it should be interviewing after the playtest period where players have had the chance to play the game properly in their own time.

Some methods like focus group discussions may work well in some situations but in others they can be problematic. It is hard to explain a pervasive game concept to participants if they do not have any previous experiences with them. To make it easier for the participants to imagine the game concept, comic strips and paper prototypes can be used. The comic strips should be designed from the players’ perspective so that the participants can take the position of a
game player. The paper prototype works better in clarifying the game concept, but it also demands a further developed game concept. Pervasive features change the whole player experience and also the way of playing so dramatically that it is usually very hard for the participants to imagine what it would feel like to actually play the game. It was shown that after a play session participants had more positive opinions about the game than in focus group discussions. So a paper prototype makes it easier for evaluators to discuss about the pervasive game with the participants.

In the example game evaluations also some novel evaluation methods were tried out. Heuristic evaluation is one of the expert evaluation methods that have recently been used to evaluate playability of games. Especially the modular approach seems efficient and flexible to use in the game context. Still there is no valid and tested heuristic module made for pervasive games so far. The next research issue could be developing and testing a special module for pervasive games.

The main research questions were whether the existing data gathering methods are suitable for evaluating pervasive games, and how they should be used. As the results presented in Chapter 6 show, it is possible to detect playability issues of pervasive mobile games using the existing methods, but the evaluation process and methods vary a lot between the games.

To conclude, the most important finding of this thesis is that new improved practices for evaluating pervasive games are needed as well as new methods like heuristics. Currently the practices vary a lot and there is no coherence in discussions about evaluations.
References


References to Games


Links to Games


The questionnaire

This questionnaire was used in Coup. Insectopia had a similar questionnaire but with some modifications.

**PLAYER**

Your player name

**THE GAME EXPERIENCE**

(likert scale: I totally agree / I partly agree / I don’t agree nor disagree / I partly disagree / I totally disagree / I can’t say)

It was easy to understand the basic idea of the game
It was easy to download the game
The game offered enough different kinds of tasks for everyone
The game was too challenging for me
The requirement level of the game was suitable for me
When I started the game and was on low levels of game’s hierarchy, it was a challenge to me

It was easy to learn how to make basic actions in the game
When making a mistake, the game gave me useful information that helped me recover from that
It was easy to understand the information on screen when playing the game
The game is targeted to a different kinds of people than me
The visual appearance of the game was appealing
I liked the theme of the game
The game world was appealing
The game is made for people like me

The game offered clear and varying tasks
The game offered enough meaningful things to do
It was easy to monitor my successes and defeats in the game
I had clear goals while I was playing the game
I had a clear conception about what I was doing and how I am going to proceed

I had trouble finding time for playing
I was annoyed by the attention of bystanders
The battles were held at difficult hours
There were too much waiting periods in the game
There were not much to do during the game
There was nothing happening in the game at times when I wanted to play
The pace of the game was too slow

I liked that I was able to play the game wherever I wanted to
I liked that I was able to play the game whenever I wanted to
I liked that the game could demand my attention at any time
I played the game in school or at work
I usually played the game in the same time of the day
I always played the game in the same place
I played this game when I was spending time with my friends who were not involved in this game
I played this game only when I had nothing else to do
The game did not disturb my ordinary life
There were too many events in the game in a short while
It felt stressful to play the game
A game session lasted too long time

The other players did not take the game as seriously as I did
Some of the players took the game too seriously
The collaboration was difficult because the players were not present at the same time
It was easy to find things to do together with other players
The game supported group forming and community building
I helped a player(s) to perform an action in the game
It was hard to see, when other players were present
I felt I was alone in the game world
The game supports communication between players
There were too many messages coming from the game
The game provided meaningful things to chat with the other players

Invitations are a good way to get players to the game
Inviting players to the game makes the game fuzzy
I invited a/some player(s) into the game
When I was invited to the game, I was disappointed to find myself in such a difficult position
I would not have joined the game unless my friends made me do it

It was easy to be an Emperor
I took a certain role/attitude when I played the game
The hierarchy among players made the game interesting
I liked that new players were invited to the game
I liked that it was unclear who were part of the game and who were not
I liked that I could invite new players to the game

GAME VALUATION
(likert scale: Very satisfied / Satisfied / Not satisfied nor unsatisfied / Unsatisfaed / Very unsatisfied / I can’t say)

Overall, how satisfied were you with Coup?
Do you think that you would play Coup again?
Would you recommend Coup to others?
Were there technical problems that made playing difficult or even impossible at times?

I would play this game if it was a commercial product
I already want to play this game again
This game would be nice to play when waiting for something (eg. the bus)
This game would be nice to play when travelling
This game would be nice to play in school / at work
This game would be nice to play at home
I would take time out of work to play this game

If this game was a commercial product, it would feel natural to buy it at...

How often did you play Coup?
How much approximately did you play Coup?
**BACKGROUND INFORMATION**

I live in: ________________

I am a [ ] woman [ ] man
Age: ________________

Education: (In Finnish version this was adapted to mach local school system)
[ ] primary school
[ ] secondary school
[ ] university
[ ] other: ________________

Current job / currently studying: ________________

My mother tongue is: ________________

**Game habits and other leisure activities**

Most important leisure activity:
_____________________________________________________________________
_____________________________________________________________________

<table>
<thead>
<tr>
<th>How often do you do: Leisure Activity</th>
<th>Daily</th>
<th>Every week</th>
<th>Every month</th>
<th>Do this sometimes.</th>
<th>Don’t do this any longer.</th>
<th>Have never done this.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Board or parlour games (e.g. Monopoly)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Card games (e.g. Hearts, Solitaire)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Run-around Games (eg. Tag)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sports (e.g. Football)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hand-held Console Games (e.g. Gameboy)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mobile phone games</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Online Role-Playing Games</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Table-top Role-Playing Games</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Live role-playing games</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participate in theatre myself</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Watch theatre or opera</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Watch reality shows on TV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sing or play a music instrument</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>When I play games, I play</th>
<th>Daily</th>
<th>Every week</th>
<th>Every month</th>
<th>Do this sometimes.</th>
<th>Don’t do this any longer.</th>
<th>Have never done this.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-player games</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(solitaire, single-player computer or console games, etc.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multi-player games</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(competitive sports, board games, etc.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Multi-player online games (online games like World of Warcraft, Eve Online etc.)

If you play computer, console or mobile games at all, which games do you play? (Feel free to select several alternatives.)
- Action games or shooters
- Role-playing games
- Simulation games
- Sport games
- Platform games
- Adventure games
- Strategy games
- Rally games
- Puzzle games
- Other, what? _______________

Which of the following devices do you have at your disposal? (Feel free to select several alternatives.)
- Computer (PC, Mac)
- Game console (XBOX, PS2 etc.)
- Handheld console (Gameboy, PSP, etc.)
- N-Gage
- Mobile phone
- Other, what? ___________________

If you play games with others, who do you play with? (Feel free to select several alternatives.)
- Friends
- Siblings
- Parents
- My children
- My wife/husband
- Strangers on the Internet
- Friends in the Internet
- Someone else: _________

How many years have you been playing computer, console and mobile games altogether?

What is your favourite game? Why?

_____________________________________________________________________________

_____________________________________________________________________________

_____________________________________________________________________________
Mobile phone use

What kind of mobile phone do you own? Please state the manufacturer's name and the model type. ___________________________ I don’t own a mobile phone

On average, how often do you make / receive mobile phone calls?
[ ] more than 15 per day
[ ] more than 10 per day
[ ] more than 5 per day
[ ] at least once a day
[ ] every other day
[ ] few per week
[ ] hardly any
[ ] never

On average, how often do you send / receive text messages?
[ ] more than 15 per day
[ ] more than 10 per day
[ ] more than 5 per day
[ ] at least once a day
[ ] every other day
[ ] few per week
[ ] hardly any
[ ] never

What other kind of features does your mobile phone have?
[ ] photo camera
[ ] video camera
[ ] mp3 player
[ ] games
[ ] Internet access

Which of these features do you use?
[ ] photo camera
[ ] video camera
[ ] mp3 player
[ ] games
[ ] Internet access

How do you pay for use of your mobile phone?
[ ] contract
[ ] pre-paid
[ ] company
[ ] parents
[ ] other

When do you carry your mobile with you?
[ ] always
[ ] working-hours
[ ] spare time
[ ] on weekends
[ ] occasionally
For how much of the time you are awake is your mobile switched on?
_____ %

How much mobile content have you downloaded to your mobile phone?
[ ] Nothing
[ ] 1-5 items
[ ] 6-15 items
[ ] 15 – 50 items
[ ] More than 50 items, it would be more if the memory wasn’t full!

How many mobile games have bought?
[ ] None
[ ] 1-2
[ ] 3-5
[ ] More than five

From where have you downloaded you mobile content
[ ] The telecom operator
[ ] Online stores
[ ] The phone manufacturer
[ ] From an independent WAP-portal
[ ] Other places such as:______________________________
Appendix B  The interview outline
This interview-outline was used in Coup’s interviews. Still quite similar interview outline was used also in Insectopia’s evaluation.

- What was your player name?
- What kind of games you usually play?
- Have you ever played games like this? Why/why not?
- How much do you play games?
- In this game, what was different (better/worse) compared to those other games that you have played?

- What was it like to start the first game session?
- Did you feel it was ok to enter a game late?
- How did it feel to start at the bottom of the hierarchy?
- What was it like to learn to play the game?
- Do you think that you need a lot of practice before playing?
- In your opinion, what is required from the player in this game?
- Did the game remain interesting during the whole period of playing?
- Did you feel like you could suspend or leave the game at any time?

- How often did you play the game on a day? How long per session?
- Would you like to play more or less than you did? Why wasn’t it possible?
- What kind of players could play this kind of game?
- What was the most interesting/boring about the game?
- What kind of problems did you have to face in this game?
- Did you try to persuade people to join the game? Why/why not?
- Do you think that in some point you could play this game more?
  - If not, what changes would it take to make you continue playing?

- Where and in which situations did you usually play this game?
- Did you do anything else during your play session? What?
- Did you think of the game at times that you were not playing?
- Do you have a phone that is needed to play this game?
  - Could you think of getting one?

- Did you know anyone else in this game?
- How did you feel about playing with strangers?
- Did you talk about the game with your friends? What?

- What was your game character like? Was it important to you?
- Was it easy to identify with your game character? Why/why not?
- Did different players take different roles or positions in this game? Was it interesting?
- Do you think that it was hard to ascend (=go forward) in this game?
- Did you use any strategy in this game? Do you think it is possible to play this game with a different strategy?
- Did you recognize other players outside the game that you did not know before you joined in?

- If you think back the time you played – did it change your daily routines compared to the time before you joined the game?
Did your gameplay have some kinds of effects on other people around you?
Did you play the game when you were with your friends?
Did other things interrupt your gameplay? What kind of things?

What did you think about the visual style and theme of this game?
Were you aware of your success and did you compare it to other players successes?
Did you feel connected to other players during?
Did you sent messages to other player(s)? Why/why not?
Do you think that the player should try to form teams in this game? Why/why not?
Was it interesting / appealing to collaborate in this game with other players?
Did the game offer enough meaningful things to do? If not, what would you like to add in this game?

How do you think that people would react if you played this game actively in public places?
What is your opinion about playing in places that people can see you and are not usually used for playing?
Did this game change your daily routines? How?
## Appendix C  The pervasive game flow heuristics [Jegers, 2006]

<table>
<thead>
<tr>
<th>Element</th>
<th>Criteria</th>
</tr>
</thead>
</table>
| **Concentration**     | Games should provide a lot of stimuli from different sources  
Games must provide stimuli that are worth attending to  
Games should quickly grab the players' attention and maintain their focus throughout the game  
Players shouldn't be burdened with tasks that don't need important  
Games should have a high workload while still being appropriate for the players' perceptual, cognitive and memory limits  
Players should not be distracted from tasks that they want or need to concentrate on  
Pervasive games should support the player in the process or switching concentration between in-game tasks and surrounding factors of importance |
| **Challenge**         | Challenges in games must match the players' skill levels  
Games should provide different levels of challenge for different players  
The level of challenge should increase as the player progresses through the game and increases their skill level  
Games should provide new challenges at an appropriate pace  
Pervasive games should stimulate and support the players in their own creation of game scenarios and pacing  
Pervasive games should help the players in keeping a balance in the creation of paths and developments in the game world, but not put too much control or constraints on the pacing and challenge evolving |
| **Player skills**     | Players should be able to start playing the game without reading the manual  
Learning the game should not be boring but be part of the fun  
Games must include online help so players don’t need to exit the game  
Players should be taught to play the game through tutorials or initial levels that feel like playing the game  
Games should increase the players' skills at an appropriate pace as they progress through the game  
Players should be rewarded appropriately for their affords and skill development  
Game interfaces and mechanics should be easy to learn and use  
Pervasive games should be very flexible and enable the players' skills to be developed in a pace set by the players |
| **Control**           | Players should feel a sense of control over their characters or units and their movements and interactions in the game world  
Players should feel a sense of control over the game interface and input devices  
Players should feel a sense of control over the game shell (starting, stopping, saving, etc.)  
Players should not be able to make errors that are detrimental to the game and should be supported in recovering from errors  
Players should feel a sense of control and impact onto the game world (like their actions matter and they are shaping the game world)  
Players should feel a sense of control over the actions that they take and the strategies that they use and that they are free to play the game the way that they want (not simply discovering actions and strategies planned by the game developers)  
Pervasive games should enable the players to easily pick up game play in a constantly ongoing game and quickly get a picture of the current status in the game world (in order to assess how the state of the game has evolved since the player last visited the game world) |
| **Clear goals**       | Overriding goals should be clear and presented early  
Intermediate goals should be clear and presented at appropriate times  
Pervasive games should support the players in forming and communicating their intermediate goals |
| **Feedback**          | Players should receive feedback on progress toward their goals |
| Players must receive appropriate feedback at appropriate times | Players should receive immediate feedback on their actions  
Players should always know their status or score |
| --- | --- |
| Immersion | Players should become less aware of their surroundings  
Players should become less self-aware and less worried about everyday life or self  
Players should experience an altered sense of time  
Players should feel emotionally involved in the game  
Players should feel viscerally involved in the game  
Pervasive games should support a seamless transition between different everyday contexts, and not imply or require player actions that might result in a violation of social norms in everyday context  
Pervasive games should enable the player to shift focus between the virtual and physical parts of the game world without losing too much of the feeling of immersion |
| Social interaction | Games should support competition and cooperation between players  
Games should support social interaction between players (chat, etc.)  
Games should support social communities inside and outside the game  
Pervasive games should support and enable possibilities for game oriented, meaningful and purposeful social interaction within the gaming system  
Pervasive games should incorporate triggers and structures (e.g. quests and events, factions, guilds or gangs) that motivate the player to communicate and interact socially |
| Social interaction | Games should support competition and cooperation between players  
Games should support social interaction between players (chat, etc.)  
Games should support social communities inside and outside the game  
Pervasive games should support and enable possibilities for game oriented, meaningful and purposeful social interaction within the gaming system  
Pervasive games should incorporate triggers and structures (e.g. quests and events, factions, guilds or gangs) that motivate the player to communicate and interact socially |
Appendix D  The usability heuristics [Nielsen, 1994, p. 30]

1. Visibility of system status.
The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.

2. Match between system and the real world
The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.

3. User control and freedom.
Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.

4. Consistency and standards
Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.

5. Error prevention.
Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.

Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.

7. Flexibility and efficiency of use.
Accelerators -- unseen by the novice user -- may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.

8. Aesthetic and minimalistic design.
Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.

9. Help users recognize and recover from errors.
Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.

Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.
Appendix E  Events that were recorded in Coup

1. **Sent message**, which occurred every time a player sent a text message to another player.
2. **Read message**, which occurred when a player read a text message sent to her.
3. **Sent hosting request**, which occurred when a player sent her superior a request for permission to host a tournament.
4. **Responded hosting request**, which occurred when a player responded to a hosting request by another player. This could be either accepted or rejected.
5. **Sent gold request**, which occurred when a player sent a request for gold to another player.
6. **Rejected gold request**, which occurred when a player rejected a request for gold sent to her by another player.
7. **Sent gold offer**, which occurred when a player either spontaneously sent an offer of gold to another player or responded to a request for gold.
8. **Responded gold offer**, which occurred when a player responded to an offer of gold by another player. This could be accepted or rejected.
9. **Sent knight request**, which occurred when a player sent a request to join another player’s tournament.
10. **Responded knight request**, which occurred when a player responded to another player’s knight request. This could be either rejected or accepted.
11. **Army recruitment**, which occurred when a player recruited a new regiment to her army.
12. **Started overthrow**, which occurred when a player started an overthrow.
13. **Started liberation**, which occurred when a player started a liberation attempt. **Note that this action was removed from the game after the first test session.**
14. **Battle-info**, which occurred every season for each ongoing battle.
15. **Surrender**, which occurred when a player surrendered to another player in a battle.
16. **Ended overthrow**, which occurred when an overthrow-attempt ended with one player winning and other player losing.
17. **Ended liberation**, which occurred when a liberation-attempt ended.
18. **Demotion**, which occurred when a player demoted her underling.
19. **Promotion**, which occurred when a player promoted one of her underlings.
20. **Player dies**, which occurred when a player died.
21. **Player joins**, which occurred when a new player joined the game.
22. **Coup started**, which occurred when a player started her Coup-client and began a play session.
23. **Coup closed**, which occurred when a player closed her Coup-client and ended her play session.