Simulation & Gaming
43(3) 391–412

DOI: 10.1177/1046878111434255
http://sag.sagepub.com
SAGE

Design for Research Results: Experimental Prototyping and Play Testing

Mirjam P. Eladhari[1] and Elina M. I. Ollila[2]

Abstract

In this article, the authors examine iterative design methods for experimental game prototype development. They recognize the area of game design as a wicked problem space, that is, an area where attempts at producing solutions change the understanding of the problems. They argue that it is vital in game-design research to build and test designs in order to explore how certain game mechanics can result in different play dynamics and play experiences. Depending on the scope of research questions and available resources, it is important to carefully plan the design process of prototypes, their development, and the testing of them. It is also important to consider what types of data to obtain, and how to treat the data, in order to acquire materials for analysis that can support the exploration of the research questions of a study. The purpose of this article is to provide a navigation aid through this process. Various methods of prototyping and types of prototypes are described, along with considerations regarding the type of game that is developed. Then, various types of play tests are presented along with recommendations, depending on timing within the production cycle and availability of test-players. Also, an overview of potential methods of obtaining data from play tests is provided.

Keywords

data types, design navigation, design process, evaluation, game design, game research, game type, iterative design, playability, play-based design, play test, production cycle, prototype development, prototyping, research method, sketching, wicked problem

[1]Gotland University, Visby, Sweden
[2]Nokia Research Center, Helsinki, Finland

Corresponding Author:
Mirjam P. Eladhari, Department of Game Design, Technology, and Learning, Gotland University, Cramérgatan 3, Visby, SE 621 57, Sweden
Email: mirjam.eladhari@hgo.se

This article is about the development of experimental research prototypes, sometimes also called *demonstrators* or *sketches*. A commonly used and recommended (see, for example, Wenzler, 2009) development method is iterative design, which, as described by Salen and Zimmerman (2001), is a play-based design process. A prototype is

> played, evaluated, adjusted and played again, allowing the designer or design team to base decisions on the successive iterations or versions of the game. Iterative design is a cyclic process that alternates between prototyping, play testing, evaluation, and refinement. (Salen & Zimmerman, 2001, p. 11)

In game research, methods from social sciences, humanities, and technological fields are used commonly. In social sciences, the effects of games on people are studied, using the methodological body of the field. Common questions concern learning processes in games. Humanities-oriented researchers commonly study the meaning and context of games, studying the games as artifacts, using the plethora of methods, such as critical analysis of the field of humanities. In the fields of engineering and industry, approaches often concern the understanding of design and developing of games. The focus is typically on how to make better games; at other times games are used as drivers of technological innovations.

It is vital for the future of the field of game research that methods from other fields of study are adapted to the nature and essence of games, and that new methods are developed for this research field. It is important to discuss the methodological challenges in the field of research, where it is not uncommon that success criteria are imported from other fields, sometimes not considering the profound aesthetic nature of games.

Because of the cross-disciplinary nature of game research, it is uncommon to use methods from only one approach. However, in research projects, methods from different disciplines are often used at certain stages, and for different ends. For example, in Eladhari's (2009) doctoral thesis, the first chapters lay the ground for the *why* of the design work conducted from a humanities perspective, reasoning about the genre (virtual game worlds) approached. The middle chapters describe the *how*—the prototypes built, using an engineering perspective in describing the technological implementation, and a cultural productions perspective in describing the game design. The final chapters concern the *results* of the research, describing the evaluations of the prototypes carried out using methods commonly used in the social sciences.

Our focus in this article is on how in general, through the work process, to plan for obtaining data from play tests that can yield answers to specific research questions. We provide a number of considerations that can be useful for researchers and prototype developers to take into account. The considerations and methods are based on our experiences with research projects, with a few exceptions where we have reviewed methods used by other researchers and added those into our "toolbox." We also point out interesting directions for future research in this field.

In the beginning of the process of developing research prototypes, researchers take the initial stance in the research question they set out to explore. This governs the type of game and game prototype that can be useful to create, how it can be tested, and how significant data can be obtained in the process.

Early in the process, the researcher's questions often include the following questions:

- *Research Question 1:* What types of game play dynamics and game play experiences can a certain, mechanic, feature approach or method result in?[1]
- *Research Question 2:* What are the qualities of the play experience, and how can these qualities be documented, analyzed, interpreted, and explained?
- *Research Question 3:* How can it be decided whether a certain game play feature results in something valuable, such as a new type of experience, a meaningful experience, or a "better experience' in some other way, and if so, compared with what? Researchers may find themselves in situations where innovative results do not necessarily lend themselves to comparisons.
- *Research Question 4:* What type of data can be useful to obtain in order to explore the research question?

We do not, in this article, focus on the design process of what makes a good game, a process described among others by Fullerton, Swain, and Hoffman (2004) and Brathwaite and Schreiber (2008). In the design process, researchers need to take into account what questions they aim to explore and stay focused on these throughout the design and implementation work. It is, in our own experience, easy to fall into a frame of mind where one aims to produce a good game, losing sight of the objective of obtaining research material. Ollila (2009) showed that excessive attempts on innovation[2] within a single game concept, while investigating specific research questions, can spoil the results. The very reason for the development of research prototypes is to find methods, features, or approaches that can be used for other games, games that are specifically made to be fun[3] and challenging, and perhaps carry a message. The researcher must approach the design of the prototype *both* as a researcher and as a game designer.

Game design has been identified as a wicked problem space (Mateas & Stern, 2005). The phrase *wicked problem* is used in social planning to describe problems where every attempt at producing a solution changes the understanding of the problems (Rittel & Webber, 1973). Mateas and Stern (2005) argued that although studying existing games can lead to deeper understanding, it is essential to also build them:

> If game studies is limited to analysing existing games and design spaces, it can be problematic to imagine or theorize about potential game features outside of these design spaces. Models about the nature of games and their features run the risk of being incomplete or wrong, simply because certain design spaces have not yet been explored.

For researchers and designers alike prototypes constitute important thinking tools. The motivation for writing this article arose when the authors recognized a common need for a navigational aid in the wicked problem space of game design, to use when planning for what type of prototype to build and how to test it, given the research question(s) and the resources (time, staff, and technology) at hand. Much of the available literature focuses on how to produce good playable games, but for the game researcher working with experimental game prototyping, the success criteria are different. It is vital to use the often scarce resources in the most efficient way to produce reliable communicable results that can be useful to others in related fields. As the scope of the myriad of different research questions can be varied, along with ways to validate data, we here focus on the middle stages; prototyping, play testing, and ways to obtain interpretable data from the tests. Our goal is that this contribution can be of use to others working with experimental game research prototypes.

Both testing methods and ways to analyze obtained data are in the field of game research prototypes heavily influenced by methods in the human-computer interaction (HCI) field. An open question is whether the prototypes and the players' experience of them can be analyzed using HCI methods "off the shelf" or if something else entirely is required? For instance, Ollila (2009) showed that HCI methods can be applied; however, they need to be adapted to the field of game research.

Usability methods have traditionally been used to assess the usability factors, such as learnability and memorability, of user interfaces (UIs). More recently, the focus has shifted to understanding User eXperience (UX) more broadly, including the emotional and social dimension, and the context of the user (see, for example, Roto, 2006). User experience can be defined as "A person's perceptions and responses that result from the use and/or anticipated use of a product, system or service" (Standard, 2009). UX is a subjective experience of using or thinking of using a product or service in a certain (kind of) context, and it can be examined in various granularity levels. User experience consists of both functional and emotional sides, with greater focus on the emotional side, and it changes dynamically over the time (Roto, 2006).

Game evaluation can be said to fall into three main areas:

- quality assurance (QA), which typically focuses on the quality of the game software and seeing that all the functionality is in place
- game usability testing and
- playability testing, which focuses on the game play

Quite often, these activities focus more on the game itself than on players' context or culture, and the tests are conducted in laboratory settings (Ollila, 2009; Pagulayan, Keeker, Fuller, Wixton, & Romero, 2008). Players' emotions are often assessed in interviews or analyzing players' voices, thinking-aloud-while-playing, body language, or facial expressions (Lazarro, 2008). Some researchers have studied player's reactions to games by recording their physiological signals during the play (Rajava, Saari, Laarni, Kallinen, & Salminen, 2005). In field (or beta) testing, players play the game in their natural settings; however, the data are usually gathered in interviews or via play diaries (Ollila, 2009), with the focus being more on the game itself than on practices or culture.

Many recent games could be better described as services than products. Many of these games are released at social platforms such as Facebook or distributed as applications for mobile devices such as smart phones. When these kinds of games are released, they do not necessarily have a full feature set ready, and will evolve by time. In this case, prototyping is important as well—to create a game that is good enough for the players to get started; after releasing the game, new features can be added and prototyped. This type of production allows developers to test features on their actual target group in production at an earlier stage than what has been customary. In this article, we look at games in general, not particularly games as services, but acknowledge that this will be an important area of research for the future.

This article is organized in the following way. First, different methods of prototyping and types of prototypes are described, along with considerations regarding the type of game that is developed. Then, various types of play tests are described along with recommendations for which types to use depending on timing within the production cycle and availability of test-players. Also, an overview of potential methods of obtaining data from play tests is provided.

Prototypes

The term *prototypes* can mean many things. For us, a prototype is anything that can be interacted with and demonstrate how a system works. Also, prototypes are usually disposable.

Some researchers in the area of HCI (Buxton, 2007) and in game research (Agustin et al., 2007) have pointed out a need for separating very early prototypes from more fine-tuned ones. They claim that just talking about lo-fi prototyping is not enough; the word *sketching* would better describe very early prototypes. In addition to the time aspect, an important difference between them is that a sketch aims to broaden the design space, whereas a prototype is more focused on evaluating design ideas.

The prototypes vary also in their maturity level. Prototypes can be early draft creations or polished late prototypes that the players can already independently interact with. They can implement a wide variety of features in a shallow level or a focused feature in completeness (for instance, demonstrating some core game play aspect). The sketches and early prototypes may not be easily understood by a player recruited for a quick laboratory test, but rather benefit from frequent quick tests with other experts. However, we acknowledge the need to involve the players who belong to the target group of the game early in the process. In sketches or very early prototypes, it is often useful to leave many holes in them to allow inventing new design directions (Buxton, 2007). The more finished ones, on the other hand, typically aim to be rather complete and the evaluations focus more on confirming whether a design works or not than on creating new design. Our definition of a prototype requires them to be something with which players can interact. Buxton's (2007) sketches can be also noninteractive, for instance, comic scenarios.

Considerations Regarding Prototypes

In research prototype design, it is vital to plan for what types of data need to be obtained with it in the play testing. A number of considerations need to be made in order to decide on type and scope of the prototypes to be developed. First, the purpose of the prototype can be very different: Is it developed to answer specific research questions, sell a concept and demonstrate its Key Selling Points (KSPs), communicate the design idea, or test and iterate the concept? Is the prototype used to expand the design space or to validate design decisions? Second, what type of game is developed? Pervasive, context-aware, single-player or social media games have quite different requirements for testing. Pervasive and context-aware games need to take the environment into account in ways appropriate to the purpose of the prototype (Stenros, Waern, & Montola, 2011 [this issue]). These types of games are challenging to create prototypes because the environment is always unpredictable, and if the test environment assumptions are set up by the test organizers, the test is no better than their imagination is (as they set the events that can happen in the environment). However, even an "ad hoc" test would be better than no test at all. Third, practical circumstances such as skills of the team and availability of test-players may restrict the nature of the intended prototype. The level of novelty of the technology can affect for example the scale of the test, and thus the prototype—large-scale field test is not possible if only a limited number of technological devices are available. Fourth, the timing within the production cycle is an important factor affecting the type of prototype that is most effective to develop—early in the process, sketches and physical prototypes are often to prefer. However, software prototypes or prototypes constructed with ready-made software and some nonsoftware components[4] (Ollila, Suomela, & Holopainen, 2008) can also be useful very early in the project, right after creating the game idea.

Types of Prototypes

We describe a number of ways to sketch and prototype games. The prototyping methods can be used both in research and commercial projects; however, there are often differences in using prototypes in these two areas. In research projects, physical prototypes are built more often. In commercial projects, the game that is being built is often similar to some earlier products (e.g., so-called sequels), and it can be easy to build a software prototype using previous engines or frameworks. Sometimes, the prototypes are also used to demonstrate the development team's skills to create the final product, and software prototypes can seem more convincing to present for the customers. In research projects, there is typically time reserved for experimenting. In commercial projects, less time is often allowed for experimenting with various kinds of features and testing them thoroughly. However, sometimes commercial projects benefit of more experimental prototyping methods as well, when completely new kinds of games are developed.

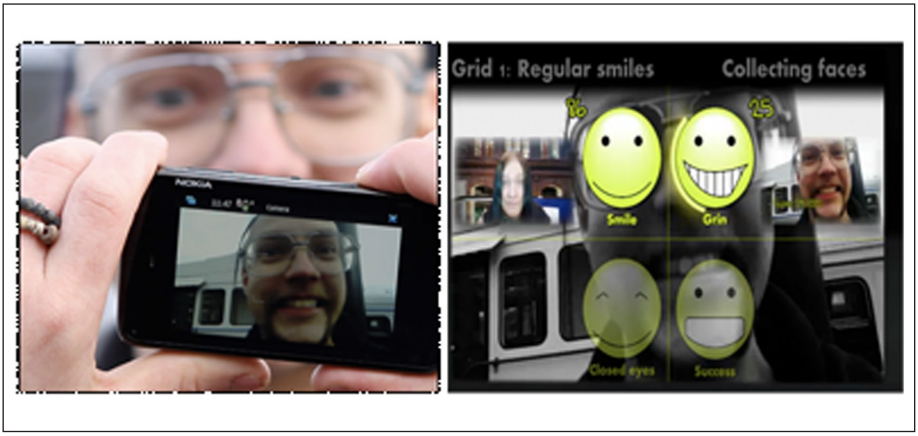

Figure 1. (a) The player is taking a picture of a person smiling (b) The smile is mapped to a smiley in a grid that the player needs to complete (i.e., take pictures with people having similar expressions on their faces as the smileys do)

Prototyping by acting and showing. One effective way of prototyping is to show how a system would actually work. These prototypes can be more or less interactive, or sometimes better described as scenarios or journeys. The tools in this category range from bodystorming (Burns, Dishman, Verplank, & Lassiter, 1994) in the very early phases of the project to polished video scenarios, the purpose of which is to show how a concept would work in its natural settings.

Video scenarios can be effective in communicating design ideas. They are not interactive, but can demonstrate interaction and are used to communicate design ideas to other team members, customers, or test subjects who can participate in evaluating the ideas or codesign. The video scenarios can also concentrate on the holistic player experience. Even rough video scenarios can be useful in conveying the ideas of concepts and have an advantage of being able to focus on the holistic experience instead of the product or service itself (Moggridge, 2007). Some evidence shows that it could be useful to separate the cognitive tasks of learning to interact and understanding the concept by giving the test subjects a possibility to only observe others to interact with the prototype instead of interacting with the prototype itself (Cui, Oulasvirta, & Ma, 2011; Goren-Bar, Graziola, Pianesi, & Zancanaro, 2006).

In one of our earlier studies, we created a video prototype of a playful concept where the players take pictures of people to match faces with smileys, complete collections, and quests (Holopainen & Ollila, 2010) as illustrated in Figure 1. The video was shown to various stakeholders, with good success.[5] We could see that using the video format enhanced the understandability of our ideas and made it easier to "sell" the idea to a third party. Our video was created with semiprofessionals; however, videos created by the designers can also be useful, even if the acting would not be professional quality. Also sometimes, a video is not needed, but the designers can act the concepts live, preferably in realistic settings and interactively. Sometimes, acting out is used already when ideating.[6] One useful method in this area is bodystorming (Burns et al., 1994). When bodystorming games, the participants imagine the game and act as though it would exist, in either a real or imaginary place where it could be played. Oulasvirta, Kurvinen, and Kankainen (2003) have experimented with the bodystorming method and argue that it can provide a more accurate understanding of the contextual factors when creating ideas. They recommend using the method for activities that are accessible (for instance, no ethical or physical restrictions) and unfamiliar to the researchers. Bodystorming has not widely been used in game design, with some exceptions (e.g., Bidwell & Holdsworth, 2006). Bodystorming could be a good alternative in very early game prototyping.

Paper or physical prototypes. Paper, or physical, prototypes are commonly associated with testing of UIs. In game design, they are often used to test playability (Ballagas & Walz, 2007; Fullerton et al., 2004). A physical prototype can at an early stage give pointers to whether a designed game mechanic results in the intended game dynamics when played. A physical prototype is often made of paper mock-ups of the intended game, but can also include physical "bits" such as figurines and tokens.

In addition, physical prototypes can also be game-mastered to make the prototype more holistic—not only focusing on the core game mechanics, but also on the events in the environment and even social acceptability of certain concepts, at least in the imaginary level. Iacucci, Kuutti, and Ranta (2000) have used game-style paper prototypes to demonstrate nongame services. We prototyped a pervasive game that included also socially challenging features, such as the game contacting the players when they were not actively playing (Koivisto & Eladhari, 2006). In our physical prototype, we used a game-mastering method, where the player was brought to various game situations with the help of a game master, like in pen-and-paper role-playing games (RPGs). The method was successful in pointing out aspects in the UI design, core game play and also pervasive and social aspects. However, we acknowledge that test setup was no better than the imagination of the test organizers and the test-players. Testing the game with a fully functional prototype could have lead to emergent situations that no one would have even thought of.

Computationally aided physical prototypes. In cases where the game mechanics demand higher degrees of computation of significant values, it can be useful to add aids for calculation to a paper prototype. A common and simple method is to use excel sheets as a tool as illustrated in Figure 2. This particular prototype mimicked a pervasive game that was played both in a virtual world with a stationary computer and in the physical world with a mobile phone. In this test, the values of the characters' projected movements and actions in the two world-representations were calculated using excel sheets in between each "round."

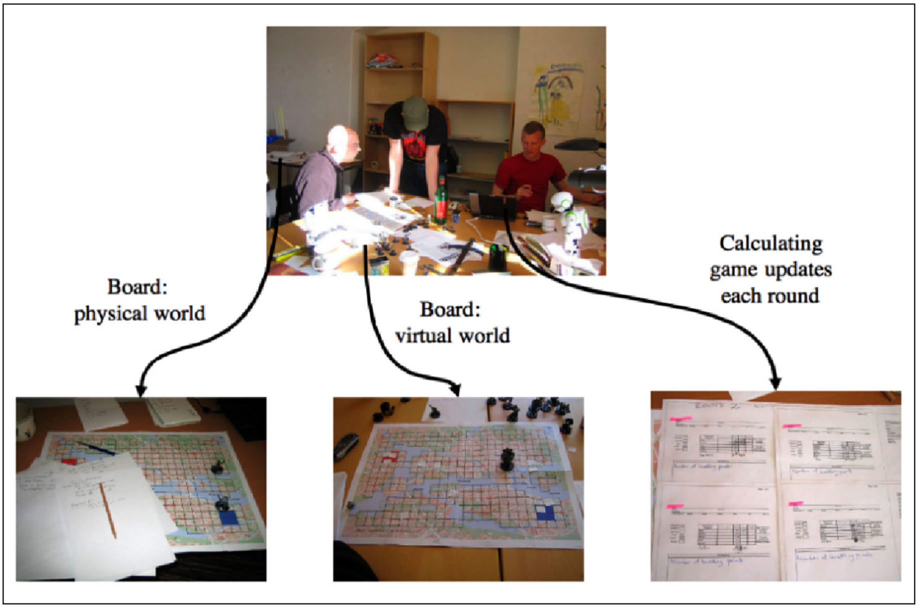

Figure 2. Use of stand-alone computation in a prototype

Augustin et al. (2007) developed a tool for sketching linear narrative games for single users. The tool provides a scripting language for interactive graphical scenes in first-person-shooter-game style where behaviors of nonplayer characters are controlled by the designers using the Wizard-of-Oz technique. The designer, or wizard, and the user connect from a PC each through a local area network to enact sketched scenarios, an activity which provides the designer with an idea of how the intended interaction may work.

Technical solutions are present that facilitate prototyping. Tools such as Microsoft's Surface Table allow for solutions where properties (such as location on a board or certain values) can be held in the "bits" of the prototypes. Physical objects placed on the screen are recognized. This gives designers the possibility to use more complex computational processes while still working with tactile, easy-to-change representations of game elements.

The tool Raptor (Smith & Graham, 2010), shown in Figure 3, is developed for tabletop interactions where several designers and users can interact simultaneously while refining game designs. In their comparative user study of Raptor, Smith and Graham found that the sketching activity using a tabletop was superior for collaborative sketching compared with the interface of traditional PCs. For more information on the raptor, see Smith & Graham, 2010.

Software prototypes. Software prototypes are quite often thought of as something created later in the project, when there is already game-design documentation available.

Figure 3. Raptor, a tool for sketching and prototyping games using a tabletop surface

However, plenty of fast prototyping tools are available, ranging from general purpose tools like Flash to more specialized frameworks.

Especially in the case of context-aware games, it can be difficult to take all the possible inputs from the environment into account when using a tool where test environment is completely set up by the test organizers.

One alternative is to use software components and physical prototyping together, such as web forums or text messaging on mobile phones. Web forums have been used together with a game master to enhance the design of a slow-update game (Ollila et al., 2008). The method worked well for understanding how the slow-update game and its core mechanics work; however, more attention should have been paid to the UI. The problem could have been avoided with using other fast prototyping methods that would have focused on the UI design. Paavilainen (2008) used text messaging and the Wizard-of-Oz[7] method to prototype a quiz game that is played with mobile phones over a long period of time. The prototyping method was used iteratively in such a way that the design was changed already during the evaluations based on the comments from the test. The method was considered successful in capturing the mobile use context and improving the design. Ballagas and Waltz (2007) combined using physical

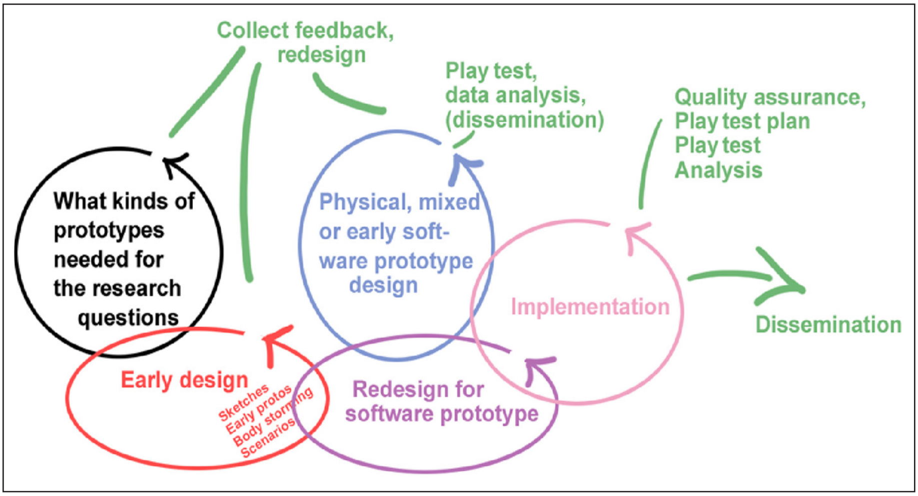

Figure 4. An example of a production cycle
Note: The process moves on from left to right.

hardware and prototyping with the Wizard-of-Oz method to explore social situations when playing a pervasive game. The method helped understanding social constraints in casting spells with gestures in a public space.

Example of a Production Cycle

The focus of our article is not the design process for games; however, we find it important to describe how the earlier mentioned methods can be used in the development. The traditional waterfall model (Royce, 1987) proposes a process where the design, implementation, and evaluation follow each other in a linear manner. The iterative design process for games (see, for example, Fullerton et al., 2004; Salen & Zimmerman, 2001) emphasizes on iterations where the game is designed, tested, and evaluated continuously during the process and developed further. Some criticize that even this kind of process would not be iterative enough (Fallman, 2003) and that design, implementation, and evaluation should be done tightly together. In our experience, the development of research game prototypes works the best when the designers, evaluators, and implementators are working closely together, doing new iterations of the game even during or between individual testing sessions. Similar approaches have been successfully used by others as well (Martin, 2000; Medlock, Wixon, Terrano, Romero, & Fulton, 2002). In a research setting, quite often the same people are doing the testing and development.

An example of production cycle where physical or mixed prototypes are followed by one or more software prototypes could include the steps shown in Figure 4.

In the design process, the design flows in parallel with the prototyping activities. Pugh (1996) talked about concept generation and controlled convergence. The design process alters between a generative process and a selection process. Ideally, these processes occur frequently in the design process (Buxton, 2007) when the new ideas are prototyped and generated and then again the best ones are selected and improved and some new features or ideas will be generated based on that. Eventually, though, the concept and feature set needs to converge into a smaller one. When the concepts have not yet been invested much in, there can be several concepts and features that are developed in parallel and either selected or thrown away in the design process.

Play Testing

The type of play test to conduct is tightly coupled to the nature of the prototype, its purpose, and practical considerations. Early in the process, it can be useful to make ad hoc tests, focus tests, expert evaluations, and participatory design workshops, while functional tests (QA) and tests with real players are more useful to conduct further along the process. In this section, different types of play tests are elaborated on along with recommendations dependent on potential practical circumstances.

Types of Play Tests

An ad hoc test is a quick informal test which requires minimal organization. It can be as simple as finding a colleague not working on the same project to quickly test some aspect of the prototype on them.

In a focus test, a group of potential players are probed about their perceptions, opinions, beliefs, and attitudes toward the prototype. Questions are asked in an interactive group setting.

An expert evaluation is the appraisal of a prototype by someone who has the professional training or experience to make an informed judgment on the design. Heuristics (Korhonen & Koivisto, 2006) can be used to support the evaluation.

The functional test, or QA, can be conducted by the production team or by QA-experts. QA is conducted in order to capture errors in the functionality of the prototype as well as for balancing game play according to the intended game play experience.

Participatory design workshops are conducted by the production team and a small number of invited guests, potentially experts in the field or potential players of the game. The workshops are often intended to aid in balancing the game and to eliminate dysfunctional elements or features. They can also be useful for gathering new ideas for further iterations or for enhancing the quality the prototype according to the goals by specific investigations.

In guided scenarios, a test-leader may use the Wizard-of-Oz method to simulate user interaction. The scenarios designed for the particular prototype are played

Table 1. Types of Play Tests, Test-Players, and the Earliest Timing in the Production Cycle

Type of play test	Who tests	Earliest timing
Ad hoc test	Team, colleagues, consultants	Very early
Participatory design workshops		Early
Quality assurance (function test)		Medium
Expert evaluations		Early-medium
Focus test	Players from target group(s)	Very early
Guided scenarios		Early
Game mastered play sessions		Medium
Free-form play sessions		Late

individually by participants. An approach to guided scenarios is described in Koivisto and Eladhari (2006).

In game-mastered play sessions, several participants may interact with a prototype, or scenarios designed for the particular prototype, guided by a game master. An example of this is provided by Tychsen, Mcilwain, Brolund, and Hitchens (2007).

In free-form play sessions, participants are interacting with the prototype unaided by guiding test-leaders or game-masters.

Guided scenarios, game-mastered play sessions and free-form play sessions can be conducted in laboratory settings where the interactions and the responses of the test-players are observed and documented. When games are tested in the field, researchers typically need to rely more on interviews, questionnaires, and play diaries.

Practical Considerations

The type of test to conduct is often dependent on timing within the production cycle and may require different types of participants as outlined in Table 1. It can be fruitful to conduct the same test on groups of participants with different backgrounds. For example, in a test where an elaborate player-character model was used, players who had extensive live action role-playing experience were more positive toward it than those lacking the same experience (Koivisto & Eladhari, 2006). As a more general note, the player participants of the test sessions should be carefully selected, otherwise it can be easily argued that the test results were not valid due to players “who would have not liked or understood that kind of game anyway.”[8] It is also important to find out early whether planned tests meet the ethical guidelines provided internally by the organization where the research is conducted. We recommend that necessary information to participants about how research data are used (such as videotaped material) and forms where players grant the use of this data are prepared well ahead of time.

Obtaining Data—Some Considerations

Regarding ways to obtain data from the play tests, further considerations need to be made, many of them dependent on the purpose of the prototype and ultimately dependent on the research questions of the investigation. For example, researchers may find it necessary to consider whether to use surveys, conduct interviews, or videotape players' interactions with prototypes. Sometimes, players' physiological responses can be recorded as well (Rajava et al., 2005). Factors to take into account may include at what point in the development process the tests are conducted, and how many players participate in the tests.

Transcribing the test sessions and conducting discourse analysis (Garfinkel, 1991; Sacks, 1995) on the text can be sometimes useful; however, it requires a lot of time. In quick prototype evaluations where speed is important, it is questionable how deep the analysis of the players' comments and reactions should be. However, again, if the purpose of the prototype is to, for instance, find out about player attitudes toward certain kinds of games, instead of creating the greatest game concept ever, time spent on discourse analysis can be well spent. Discourse analysis can be used to detect hidden attitudes and opinions of the players that cannot always be learnt from what they say, but from how they say it instead.

According to Buxton (2007), prototypes should be quick, timely, inexpensive, disposable, and plentiful. If very deep analysis methods are used for the prototypes, the prototypes cannot achieve these qualities. Especially in the very early phases of the design process, when the game concept itself and creating the game are in the focus, the analysis should be conducted rapidly. Moreover, the more the time invested in the early prototypes, the less disposable they become. Like mentioned earlier, using prototypes to find out about broader research questions than creating the game itself can benefit of deep analysis. Also, ethnographical (Blomberg, Burrel, & Guest, 2008; Blomberg, Giacomi, Mosher, & Swenton-Wall, 1993) studies that can be used to find answers to questions of what should be prototyped (instead of how) can benefit of deep analysis techniques.

As summarized in Table 2, there is a number of considerations researchers might want to take into account regarding how to obtain data from play tests.

The research methods range from surveys to interviews and play diaries. Sometimes teach-back tasks are used for a memory check (Puerta-Melguizo, Chisalita, & Veer, 2002; Veer, Wijk, & Felt, 1990). Teach-back tasks are often used in order to assess users' mental models of systems. For example, in a test of a prototype where the players' avatars had different "moods" given by a psychological model, players were asked to explain to a friend how the mood affected what they could do or not do in the game world prototype (Eladhari, 2009).

The videotaped materials allows for detailed study of body language, actions, and utterances. The material can also be used in interviews where the test-leader and the players together watch the videos as described by Jørgensen (2011 [this issue]).

Table 2. Obtaining Data

Data to obtain	Asking players via		Observing players		
	Surveys	Interviews	Interaction Choice of actions Patters of actions	Expressions Facial expressions Body language Voice (pitch and tone) Verbal expressions (think-aloud, conversation, etc.)	Physiological responses Heart rate variability GAZE (eye tracking) Electroencephalography (EEG) Temperature and galvanism in fingers
Capturing of data	(Documentation inherent in the actual survey forms)	Note taking Tape recording Video tapping	Video tapping Sound recording	Observation of measured values in real time and note taking Log-report from devices measuring physiological responses	
Processing of obtained data	Identification of relevant part of documentation Verbatim transcriptions of relevant parts and annotation of significant expressions Synchronization of used reports of documentation from different sources, such as software event-logs, verbatim transcripts, and log report of physiological responses. After this, data may be analyzed and interpreted with methods appropriate for the individual project				

Sometimes, physical data are gathered to understand better how the player interacts with the game. Potential physical data include eye-tracking, heart-rate variability, galvanism and temperature in fingers, electroencephalography (EEG) or functional magnetic resonance imaging (fMRI), and posture (posture sensors may be placed on participants' chairs; see Plass, Perlin, Nordlinger, & Isbister, 2010). For further reference, Kivikangas et al. (2010) provided an overview of psychophysiological methods in game research, presenting the most useful measures to date.

In the case that the test involves a software prototype, it can be useful to also develop or use a back-end system for capturing events in the form of player input to the system. Potential considerations may concern as to what extent is it useful to log the player's interactions with the system and which types of interactions to log. That is, when data mining is used, it is crucial to consider what data to mine. The team at Microsoft Labs who conducted the testing of Halo 3 (Thompson, 2007) used "heatmaps" showing recurring events mapped to the geography of the game. This was useful for the designers of the game when they iteratively redesigned the environment and the placement of items in it relevant to the game play. It can be also valuable to connect the log data to more qualitative methods, like the utterances players make while "thinking aloud"[9] as they interact with the prototype.

If multiple sources of data are collected simultaneously, researchers may want to consider methods for integrating these in order to make the process of analysis as practical as possible.

Finally, unless only surveys are used, the way to treat the material needs to be considered. For example,

- Is it useful to make verbatim transcriptions of the videotaped play tests? Should nonverbal expressions be annotated in the transcripts?

- Is it useful to make a rough coding scheme, for example, counting certain types of utterances, behaviors, and/or expressions or reactions?
- Is it useful to trace and describe the actions a player performs in relation to the paper prototype? For example, it can be useful to observe and describe in what order players do something, and whether there are patterns in how they approach different elements of the prototype.

Qualitative research methods are good particularly in the very early phases of the project when the sketches and early prototypes are evaluated to understand both what is the right design to take forward, and how the design can be improved and to generate new ideas. This is partly due to practical reasons; the very early prototypes often need a lot of explanation or assistance for the test subjects to use and understand them, and partly for getting the most out of the testing. The early prototypes should also be disposable, and thus investing too much on a single prototype is questionable.

Fran Samolins argues in Moggridge (2007) that “emphatic research methods,” when skillfully used, are good for yielding much inspiration from a small number of subjects, and statistically valuable data obtained in market research can reveal statistically viable truth, but is unlikely to yield inspiration. In the very early phases of game prototype development, interacting with the players in qualitative studies helps to understand better our own game design and develop it further.

Quantitative evaluations deal with large numbers and often the result is to prove which design alternative works the best or to find statistical dependencies in the data. An example where survey data collected from participants in a play test of a digital prototype and analyzed with statistical methods can be found in Tychsen et al. (2007). Two examples of analysis of data collected from a game-mastered play test of a pen-and-paper RPG where videotaped recordings of interactions were annotated are provided by Drachen, Hitchens, Jhala, and Yannakakis (2009) and Eladhari and Mateas (2008). The latter uses qualitative data analysis as advocated by Miles and Huberman (1994). The work conducted at the Games for Learning Institute provide examples of how to analyze and combine data from various sources, including event logs collected from digital prototypes, physiological data, and participants’ self-reports (Plass et al., 2010). At Microsoft Game Labs, maps are used for display of significant events which are logged during play testing (Hopson, 2010). The events are mapped to the geography of the game illustrated in Figure 5, where the dots signify deaths of player characters. Each dot is colored by the cause of death. Users can click the dots in order to display a recording of the last 10 seconds of play in the game-client. The events are also mapped to video recordings of participants’ faces. By using data mining, and triangulation of data from different sources, researchers and designers can get a picture of what patterns of play a certain iteration of a design elicits, as well as interpret players’ emotional responses by observing their facial expressions.

Qualitative and quantitative research methods should not be seen as competing approaches, instead, they can work really well together. For instance, a qualitative laboratory test with more focus on the first-time experience can complement a field

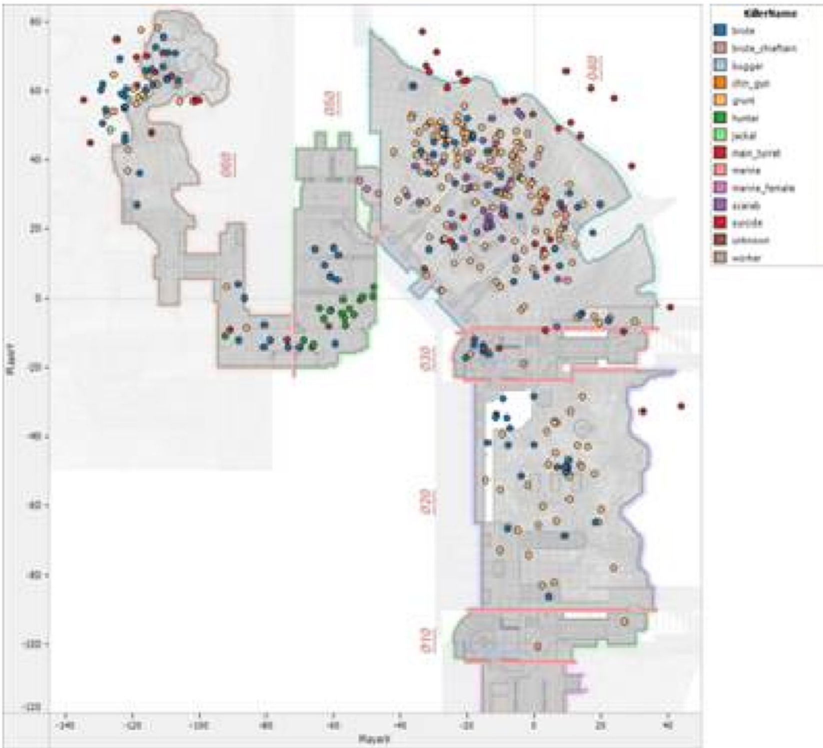

Figure 5. Map used at Microsoft Game Labs showing a top-down view of the fourth mission of Halo 3
Note: The dots show the locations of player deaths.

test with more focus on the long-term experience and using quantitative research methods. Furthermore, in the same test, qualitative and quantitative research methods can be combined. One way to do this is to make sure that the log data can be mapped with player interviews.

Conclusion

In this article, we presented various ways of sketching or prototyping game ideas in research projects. We discussed choosing the research questions and the methods to obtain data. An important point that we wanted to make was that in the area of game research, the question is not only how to develop "the greatest game ever," but also more how to create a game with which reliable data can be gathered on the questions the researchers are interested in. The methods and considerations are based on our experience on creating various game research prototypes in different kinds of teams and settings. A few of the methods that we described were from other researchers, added in our "toolbox" because we consider those as promising ways to study games.

Our take on how the process of creating research prototypes should be is somewhat similar to Fallman's argument about nonlinearity (Fallman, 2003). Finding the research problems or challenges, creating the solution, and evaluating the result should go hand in hand. Prototypes are extremely useful tools even right after the game idea. It usually makes sense to build a simple prototype right after the game idea, before writing long descriptions on how the game could actually work. The prototypes can be either quick software prototypes, physical prototypes, or even rough sketches. In our view, prototypes and sketches constitute important thinking tools for researchers and designers alike, aiding us in exploring the wicked problem space of game design.

Authors' Note

Authors are listed in alphabetical order.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Notes

1. As described by Brathwaite and Schreiber (2008), a core *mechanic* (such as flipping over tiles or selling items to another player) of a game results in a core *dynamic* when it is played.
2. Particularly innovation with creating a large set of completely new kind of, potentially engaging, game features that could make the game more fun to play, which may make it difficult to interpret the research data.
3. However, we do not suggest here that research prototypes should not be fun to play, which actually may be another pit fall.
4. Such as using web forums with a human game master.
5. With one party investing further in protecting the game idea.
6. Generating ideas for concepts.
7. A human actor operates as a computer "behind the screen."
8. A comment heard sometimes when the test results are not very positive.
9. When using the think-aloud protocol participants are asked to verbalize what they do, think, and feel when interacting with a prototype. This allows observers to see how users go about to complete tasks (Lewis, 1982).

References

Agustin, M., Chuang, G., Delgado, A., Ortega, A., Seaver, J., & Buchanan, J. W. (2007). Game sketching. In *Proceedings of DIMEA '07* (pp. 36-43). New York, NY: ACM.

Ballagas, R., & Walz, S. P. (2007). REXplorer. In C. Magerkurth & C. Röcker (Eds.), *Pervasive gaming applications vol. 2*, Aachen, Germany: Shaker.

Bidwell, N. J., & Holdsworth, J. (2006). Battleship by foot. In *Proceedings of IE '06* (pp. 67-74). Perth, Australia: Murdoch University.

Blomberg, J., Burrel, M., & Guest, G. (2008). An ethnographic approach to design. In A. Sears & J. A. Jacko (Eds.), *The human-computer interaction handbook* (2nd ed., pp. 964-986). Mahwah, NJ: Lawrence Erlbaum.

Blomberg, J., Giacomi, J., Mosher, A., & Swenton-Wall, P. S. (1993). Ethnographic field methods. In D. Schuler & A. Namioka (Eds.), *Participatory design: Principles and practices* (pp. 123-155). Mahwah, NJ: Lawrence Erlbaum.

Brathwaite, B., & Schreiber, I. (2008). *Challenges for game designers* (1st ed.). Rockland, MA: Charles River Media.

Burns, C., Dishman, E., Verplank, W., & Lassiter, B. (1994). Actors, hairdos & videotape—Informance design. In *Proceedings of CHI '94* (pp. 119-120). New York, NY: ACM.

Buxton, B. (2007). *Sketching user experiences*. San Francisco, CA: Morgan Kaufmann.

Cui, Y., Oulasvirta, A., & Ma, L. (2011). Event perception in mobile interaction. *International Journal of Human Computer Interaction, 27*, 413-435.

Drachen, A., Hitchens, M., Jhala, A., & Yannakakis, G. (2009). Towards data-driven drama management. In *Proceedings of DiGRA 2009.* London, England: DiGRA Publishers.

Eladhari, M. P. (2009). *Characterising action potential in virtual game worlds applied with the mind module* (Doctoral dissertation). Teesside University, UK.

Eladhari, M. P., & Mateas, M. (2008, December 3-5). Semi-autonomous avatars in world of minds. In *Proceedings of ACE 2008*, Yokohama, Japan: ACM.

Fallman, D. (2003). Design-oriented human-computer interaction. In *Proceedings of CHI '03* (pp. 225-232). New York, NY: ACM.

Fullerton, T., Swain, C., & Hoffman, S. (2004). *Game design workshop: Designing, prototyping, and playtesting games*. San Francisco, CA: CMP.

Garfinkel, H. (1991). *Studies in ethnomethodology*. Cambridge, UK: Polity. (First published by Prentice Hall, 1967)

Goren-Bar, D., Graziola, I., Pianesi, F., & Zancanaro, M. (2006, March 1). The influence of personality factors on visitor attitudes towards adaptivity dimensions for mobile museum guides. *User Modeling and User-Adapted Interaction, 16*, 31-62.

Holopainen, J., & Ollila, E. M. I. (2010, May). *Collecting smiles-augmented reality playful application for mobile phones*. In Pervasive 2010. Helsinki, Finland.

Hopson, J. (2010). Better games through better understanding, Keynote. In *Proceedings of FDG 2010.* Monterey, CA: ACM.

Iacucci, G., Kuutti, K., & Ranta, M. (2000). On the move with a magic thing. In *Proceedings of DIS '00* (pp. 193-202). New York, NY: ACM.

Jørgensen, K. (2011). Players as co-researchers: Understanding games through player interpretations. *Simulation & Gaming.* [doi:10.1177/1046878111422739]

Kivikangas, J. M., Ekman, I., Chanel G., Järvelä, S., Cowley, B., Salminen, M., . . . Ravaja, N. (2010). Review on psychophysiological methods in game research. In Proceedings of First Nordic DiGRA. Stockholm, Sweden: DiGRA. Retrieved from http://www.digra.org:8080/Plone/dl/display_html?chid=http://www.digra.org:8080/Plone/dl/db/10343.06308.pdf

Koivisto, E. M. I., & Eladhari, M. (2006). Paper prototyping a pervasive game. In *Proceedings of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology (ACE '06).* ACM, New York, NY. doi:10.1145/1178823.1178935. Retrieved from http://doi.acm.org/10.1145/1178823.1178935

Korhonen, H., & Koivisto, E. M. I. (2006). Gameplay heuristics for mobile games. In *Proceedings of mobile* HCI (pp. 28-35). Helsinki, Finland: ACM.

Lazarro, N. (2008). Why we play: Affect and fun of games. In A. Sears & J. A. Jacko (Eds.), *The human-computer interaction handbook* (2nd ed., pp. 679-700). Mahwah, NJ: Lawrence Erlbaum.

Lewis, C. H. (1982). *Using the "thinking aloud" method in cognitive interface design* (Tech. Rep. No. RC-9265). Yorktown Heights, NY: IBM.

Martin, A. (2000). The design and evolution of a Simulation/Game for teaching information systems development. *Simulation & Gaming, 31*, 445-463.

Mateas, M., & Stern, A. (2005, June 16-20). Build it to understand it. In *Proceedings of DiGRA 2005 Conference* (pp. 299-310). Vancouver, Canada: University of Vancouver.

Medlock, M. C., Wixon, D., Terrano, M., Romero, R., & Fulton, B. (2002, July). Using the RITE method to improve products. In *Proceedings of Usability Professionals' Association Annual Conference*. Orlando, FL: Usability Professionals' Association.

Miles, M. B., & Huberman, M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: SAGE.

Moggridge, B. (2007). *Designing interactions* (1st ed.). Cambridge, MA: MIT Press.

Ollila, E. M. I. (2009). *Using prototyping and evaluation methods in iterative design of innovative mobile games* (Doctoral dissertation). Tampere University of Technology, Tampere, Finland.

Ollila, E. M. I., Suomela, R., & Holopainen, J. (2008). Using prototypes in early pervasive game development. *Computers in Entertainment, 6*(2), 1-7.

Oulasvirta, A., Kurvinen, E., & Kankainen, T. (2003). Understanding contexts by being there: Case studies in Bodystorming. *Personal and Ubiquitous Computing, 7*, 125-134.

Paavilainen, J. (2008, September 20). Mobile game prototyping with the Wizard of Oz. In *Proceedings of Dream 2008*. Odense, Denmark: Danish Research Centre on Education and Advanced Media Materials.

Pagulayan, R. J., Keeker, K., Fuller, T., Wixton, D., & Romero, R. L. (2008). User-centered design of games. In A. Sears & J. A. Jacko (Eds.), *The human-computer interaction handbook* (2nd ed., pp. 883-906). New York, NY: Lawrence Erlbaum.

Plass, J. L., Perlin, K., Nordlinger, J., & Isbister, K. (2010, March 9-13). Research on design patterns for effective educational games. In *Game Developers Conference*. San Francisco, CA. Retrieved from http://www.gdcvault.com/free/gdc-10

Puerta-Melguizo, M. C., Chisalita, C., & Van der Veer, G. C. (2002). Assessing users mental models in designing complex systems. In *Proceedings of IEEE 2002* (p. 6). Hammamet, Tunisia: IEEE. doi:10.1109/ICSMC.2002.1175734

Pugh, S. (1996). *Creating innovative products using total design: The living legacy of Stuart Pugh* (D. Clausing & R. Andrade, Eds.). Boston, MA: Addison-Wesley.

Rajava, N., Saari, T., Laarni, J., Kallinen, K., & Salminen, M. (2005, June). The psychophysiology of video gaming. In *Proceedings of DiGRA 2005.* Vancouver, Canada: University of Vancouver. Retrieved from http://www.digra.org:8080/Plone/dl/display_html?chid=http://www.digra.org:8080/Plone/dl/db/06278.36196.pdf

Rittel, H., & Webber, M. (1973). Dilemmas in a general theory of planning. *Policy Sciences, 4*, 155-169.

Roto, V. (2006). *Web browsing on mobile phones—Characteristics of user experience* (Unpublished doctoral dissertation). Helsinki University of Technology, Finland.

Royce, W. W. (1987). Managing the development of large software systems: Concepts and techniques. In *Proceedings of ICSE '87* (pp. 328-338). Los Alamitos, CA: IEEE Computer Society Press.

Sacks, H. (1995). *Lectures on conversation: Volumes I & II* (G. Jefferson [Ed.]). Oxford, UK: Blackwell.

Salen, K., & Zimmerman, E. (2001). *Rules of play: Game design fundamentals*. Cambridge, MA: MIT Press.

Smith, J. D., & Graham, T. C. N. (2010). Raptor. In *ACM FuturePlay 2010* (pp. 191-198). Vancouver, Canada: ACM.

Standard. (2009). *ISO FDIS 9241–210* (Tech. Rep.). Retrieved from http://www.iso.org/iso/catalogue_detail.htm?csnumber=52075

Stenros, J., Waern, A., & Montola, M. (2011). Studying the elusive experience in pervasive games. *Simulation & Gaming*. [doi:10.1177/1046878111422532]

Thompson, C. (2007, August 21). Halo 3: How Microsoft Labs invented a new science of play. *Wired Magazine, 15*(09). Retrieved from http://www.wired.com/gaming/virtualworlds/magazine/15-09/ff_halo?currentPage=all

Tychsen, A., Mcilwain, D., Brolund, T., & Hitchens, M. (2007). Player character dynamics in multi-player games. In *Proceedings of DiGRA 2007* (pp. 40-48). Tokyo, Japan: DiGRA.

Veer, G. C. Von der, Wijk, R., & Felt, M. A. M. (1990). Metaphors and metacommunication in the development of mental models. In *Cognitive ergonomics: Understanding, learning and designing human-computer interaction* (pp. 133-149). San Diego, CA: Academic Press Professional.

Wenzler, I. (2009). The ten commandments for translating simulation results into real-life performance. *Simulation & Gaming, 40*, 98-109.

Bios

Mirjam P. Eladhari (PhD) is associate professor at the GAME department of Gotland University. Her main area of research is Artificial Intelligence (AI)–driven game design. The research approach she has adopted includes exploration of the social multiplayer game-design

space through experimental implementations of prototypes where both novel and established AI techniques are used.

Contact: mirjam.eladhari@hgo.se

Elina M. I. Ollila (PhD) is a principal researcher and designer at Nokia Research Center in Finland. Her focus is game research, social media, virtual worlds, and interaction design. She has participated in projects creating several prototypes in this field in different roles and published articles on the work.

Contact: elina.ollila@nokia.com