BAUHAUS-UNIVERSITÄT WEIMAR

Frames and Lenses

Framing Gameplay Experience in Games with Eye Movement Based Adaptation

by

Katharina Spiel

A thesis submitted in partial fulfillment for the degree of
Master of Science

in the
Faculty of Media
Usability Research Group

September 2014

BAUHAUS-UNIVERSITÄT WEIMAR

Abstract

Faculty of Media
Usability Research Group

Master of Science

by Katharina Spiel

Gameplay Experience is affected by several attributes of games. Amongst them, two have been chosen here to look into further: Technological novelty and Framing. The work presented shows a development process for a digital game directly using eye movements for adaptation (next to conventional adaptation) and an extensive user study where the game has been presented in two different ways to players. Results of the study indicate that performance is more affected by eye movement based adaptation, that framing and adaptation have an influence on gameplay experience (at least in some settings) and that framing shapes expertise as measured by eye movements in conventionally adapted games. These are put in perspective and discussed thoroughly with an outlook on the future of research in eye movement based adaptation.

Contents

Abstract **i**

1 Aspects of Games, Technology and Adaptation **1**

2 On the Shoulders of Giants – Foundations of this Work **4**
2.1 Playing Games . . . 5
2.1.1 Expertise . . . 5
2.1.1.1 Acquiring a Skill . . . 6
2.1.1.2 TETRIS Experts . . . 7
2.1.2 User Experience in Games . . . 9
2.1.2.1 Presence, Involvement, Flow, Immersion . . . 9
2.1.2.2 Questionnaires . . . 10
2.2 Tracking Eyes . . . 13
2.3 Adapting Software . . . 15
2.4 Framing Experiences . . . 16
2.5 Framing the User Experience of Games with Eye Movement Based Adaptation . . . 17

3 Research Hypotheses **18**
3.1 Performance Based Hypotheses . . . 18
3.2 Gameplay Experience Based Hypotheses . . . 19
3.3 Expertise Based Hypotheses . . . 20
3.4 Intersectional Hypotheses . . . 20

4 Tetris **22**
4.1 General Description . . . 22
4.1.1 Mathematical Attributes . . . 23
4.1.2 Game State Analysis . . . 24
4.2 Choosing Blocks . . . 26
4.2.1 NICETRIS . . . 27
4.2.2 GRAB BAG . . . 27
4.2.3 TRUE RANDOM . . . 28
4.2.4 SKEWED RANDOM . . . 28
4.2.5 BUST HEAD . . . 28

4.3 Implementations . . . 29
4.3.1 Pytris . . . 29
4.3.1.1 Game Mechanics . . . 29
4.3.1.2 Data Recording . . . 30
4.3.1.3 Version for Skill Determination . . . 31
4.3.2 (N)EMTris . . . 31
4.3.2.1 Adaptivity of NEMTris . . . 31
4.3.2.2 Adaptivity of EMTris . . . 33
4.3.2.3 Data Recording . . . 36

5 Pre-Study, Pilot Study and Design of User Study 38
5.1 Pre-Study . . . 38
5.1.1 Leading Questions . . . 38
5.1.2 Procedure . . . 39
5.1.3 Results . . . 39
5.1.3.1 Performance Based Results . . . 40
5.1.3.2 Questionnaire Results . . . 41
5.1.3.3 Other Results . . . 41
5.1.4 Consequences for Further Development . . . 43
5.2 Study Setup . . . 44
5.2.1 Test Setup . . . 44
5.2.2 Study Procedure . . . 44
5.2.3 Determining Expertise . . . 45
5.3 Pilot Study . . . 46
5.3.1 Purpose and Leading Questions . . . 46
5.3.2 Results . . . 46
5.3.3 Consequences for Main Study . . . 48
5.4 Operational Hypotheses . . . 48
5.4.1 Performance Based Hypotheses . . . 48
5.4.2 Gameplay Experience Based Hypotheses . . . 49
5.4.3 Expertise Based Hypotheses . . . 49
5.4.4 Intersectional Hypotheses . . . 50

6 Results of User Study 51
6.1 Procedure . . . 51
6.2 Interpretation of Data . . . 52
6.2.1 Test on Normal Distribution . . . 52
6.2.2 Tests on Significance of Difference . . . 53
6.2.3 Determination of Effect Size . . . 53
6.3 Population . . . 54
6.4 Results According to Hypotheses . . . 55
6.4.1 Performance Based Hypotheses . . . 55
6.4.1.1 Framed vs. Non-Framed Players . . . 56
6.4.1.2 Type of Adaptation . . . 56
6.4.1.3 Effect of Framing on Conventionally Adapted Games . . . 57
6.4.1.4 Effect of Type of Game on Framed Players . . . 58
6.4.1.5 Framing and Player Improvement . . . 58

6.4.1.6 Summary . . . 59
6.4.2 Gameplay Experience Based Hypotheses . . . 59
6.4.2.1 Effect of Framing . . . 60
6.4.2.2 Effect of Type of Adaptation . . . 60
6.4.2.3 Effect of Framing on Conventionally Adapted Games . . . 60
6.4.2.4 Effect of Type of Game on Framed Players . . . 61
6.4.2.5 Framing and Difference in Reported Experience . . . 61
6.4.2.6 Summary . . . 62
6.4.3 Expertise Based Hypotheses . . . 62
6.4.3.1 Effect of Framing . . . 63
6.4.3.2 Effect of Type of Adaptation . . . 63
6.4.3.3 Effect of Framing in Conventionally Adapted Games . . . 63
6.4.3.4 Effect of Type of Adaptation on Framed Players . . . 64
6.4.3.5 Player Improvement . . . 64
6.4.3.6 Summary . . . 65
6.4.4 Intersectional Hypotheses . . . 65
6.4.4.1 Effect of Cleared Lines on Gameplay Experience . . . 65
6.4.4.2 Effect of Cleared Lines on Eye Movements and Epistemic Actions . . . 67
6.4.4.3 Correlation between Eye Movements/Epistemic Actions and GEQ Values . . . 68
6.4.4.4 Summary . . . 69

7 Discussion **70**
7.1 Interpretation According to Hypotheses . . . 70
7.1.1 Framing . . . 70
7.1.2 Use of Eye Movement Based Adaptation . . . 70
7.1.3 Framing in Conventionally Adapted Games . . . 71
7.1.4 Adapting Games for Framed Players . . . 72
7.1.5 Framing and Adaptation Differences . . . 72
7.1.6 Intersectional Influences . . . 73
7.2 Side Effects . . . 73
7.2.1 Gender . . . 73
7.2.2 Age . . . 73
7.2.3 Language . . . 74
7.2.4 Mental Rotation . . . 74
7.2.5 Tetris expertise . . . 74
7.2.5.1 Self Reported . . . 75
7.2.5.2 Established by Pytris . . . 75
7.2.6 Summary . . . 76
7.3 Expected vs. Actual Results . . . 76

8 Conclusion **77**
8.1 Application of Research . . . 77
8.2 Designing Games with Psychophysical Adaptation . . . 78
8.3 Future Research . . . 80
8.3.1 Optimisation of Current Research . . . 80

8.3.2 Advancing Further . . . 80
8.4 Limitations and Open Questions . . . 82
8.5 Summary . . . 82

Bibliography **84**

List of Figures **93**

List of Tables **95**

Acknowledgements **97**

A Questionnaires and Test Material **98**
A.1 Pre-Test Questionnaire . . . 98
A.2 Consent Form Main Study . . . 100
A.3 Initial Questionnaire for Statistical Data . . . 101
A.4 Framing Texts . . . 102
A.4.1 Framed Participants . . . 102
A.4.2 Non-Framed Participants . . . 102
A.5 Data Control Sheet . . . 103

B Additional Evaluations of Studies **105**
B.1 Pre-Study . . . 105
B.1.1 Test Participants . . . 105
B.1.2 Performance-Based Results . . . 106
B.1.3 Questionnaire-Based Results . . . 107
B.1.4 Other Results . . . 107
B.2 Pilot-Study . . . 108
B.2.1 Test Participants . . . 108
B.2.2 Skill Test Results . . . 108
B.3 Main Study . . . 108
B.3.1 Population . . . 108
B.3.2 Results According to Hypotheses . . . 111
B.3.2.1 Performance Based Hypotheses . . . 111
B.3.2.2 User Experience Based Hypotheses . . . 116
B.3.2.3 Expertise Based Hypotheses . . . 119
B.3.2.4 Intersectional Hypotheses . . . 121
B.3.3 Side Effects . . . 123
B.3.3.1 Gender . . . 123
B.3.3.2 Age . . . 125
B.3.3.3 Language . . . 126
B.3.3.4 Mental Rotation . . . 128
B.3.3.5 Self-Reported Expertise . . . 128
B.3.3.6 Pytris Established Expertise . . . 129

C German Translations of the Material **130**
C.1 Consent Form . . . 130
C.2 Framing Texts . . . 131
C.2.1 Framed Participants . . . 131
C.2.2 Non-Framed Participants . . . 131
C.3 Questionnaires . . . 132
C.3.1 Statistical Data . . . 132
C.3.2 After every Game . . . 134
C.3.3 Game Experience Fragebogen . . . 135
C.3.4 Post-Game Fragebogen . . . 138

Declaration of Authorship **140**

Für meine Tante Gerti.
Weil sie uns trotz allem noch das Schöne im Leben gezeigt hat.
Und TETRIS.

Chapter 1

Aspects of Games, Technology and Adaptation

Playing games is one of the most loved recreational activities. When games became digital, new qualities emerged. It has been made possible to play with other players you never had and never will have physical contact with, but it is also possible to have social play only with artificial agents instead of human players. New technologies and new ways of playing emerged as well. Technology makes potentially costly or otherwise physically annoying games (like TETRIS) enjoyable, fun or even possible[1].

With the first screen was attached to a computing machine, digital games like SPACEWAR!, ADVENTURE and, finally, PONG were also introduced [cf. Lowood, 2009]. Since then, they have been a popular recreational activity. When playing digital games players not only interact with the particular hardware, they also interact with a game and the way it has been presented before it is actually played. This presentation often includes a storyline, a demonstration of the look and feel, an introduction to the character design, a certain style of advertisement and expected focus groups, but technologies to enhance the gameplay are also included. The latter can consist of, e.g., graphic design, construction of the artificial intelligence or novel techniques for adaptation. All these parameters add to the expectations a potential player has of a game, how to interact with it and the overall gameplay experience. As well as this gameplay experience, it is of interest to the work presented here to investigate if and how a player's performance is altered and how their eye movements differ when they are told that a game uses eye movement based adaption compared to actually playing a game using this technology.

[1] see for example `http://notsonoisy.com/tetris/index.html` where TETRIS amongst other games has been implemented physically. However, real time performances can only be simulated by this point.

There are two different types of approaches for games to calculate their next steps as agents: They could be optimising, which means, aiming for the best possible outcome, or satisficing, which means, they try to take the best option available to them until a satisfactory outcome, one that is 'good enough,' can be achieved. This procedure is more closely related to what humans do, whereas the act of optimising can be seen as mechanic and artificial [see also Simon, 1956]. Satisficing game agents – as opposed to optimising game agents – have been described as desirable by Stirling and Goodrich [1999] amongst others. A game agent can be as simple as the random number generator in combination with the time constraints used in Tetris [see Fahey, 2012]. This algorithm, in and of itself, is neither optimising nor satisficing. Satisficing implementations that are adaptive to their players [see for example Poloni, 2012, trying to create the worst possible game for a player] do exist, however.

In order to adapt well to an individual player and hence create the best gameplay experience for that player, it is desirable to collect as much information as possible and reasonable about them, extract the important data and adapt accordingly. General, individual and situational aspects of playing a game have to be considered in order to make well-done adaption within a game possible [see also Bertel, 2014]. Additionally to the way players execute commands and interact with the game directly, indirect information about their emotional state, well-being and strategic considerations can potentially be acquired by recording psychophysical data such as eye movement tracking. Accordingly, previous efforts have been undertaken to create games that adapt to a player's eye movements [see Wetzel et al., 2014].

Framing on the other hands refers to how games are presented. For example, in research discussions, some publications focus on harms of games [e.g. Griffiths, 1999] whereas others try to illuminate their usefulness [e.g. Prensky, 2005]. Both approaches create different emotions and feelings towards games in the respective research communities and certain media outlets. The same accounts for games themselves. Certain styles of presentation of the games target certain player groups; when the advertising is changed, this might open up the same game for different player groups while it has not been fundamentally changed. Framing refers to any type of conscious or subconscious presentation of an object.

This thesis investigates the question of how the suggestion – or framing – of adaptivity through eye movement derived data shapes the player's gameplay as well as their experience and how this relates to actually employing eye movement based adaption mechanisms within a game. The game of reference used in this context is TETRIS. To answer the question posed, two adaptive versions of TETRIS were developed: one using only an analysis of the player's current performance and one additionally using eye movement data.

This thesis contributes to the research community in several aspects by touching several areas of interest. The most important of these are:

- Investigation into placebo effects of the suggestion of technological advance
- An example for the mainly data driven development of a game using eye movement based adaption
- A thorough discussion of results of an extensive user study of systems with and without eye movement based adaption

Methodologically, a user study investigating a framing effect on the gameplay experience is the core point of a mainly quantitative analysis. This study is done in a mixed method setup within and inbetween test participants. The data-driven and quantitative approach for the development of the prototypes presents a generalisable approach for how eye movement based or psychophysical adaption in general can be employed in games and integrated into game design.

After an overview of the current stage of research (Chapter 2) and the main goals of the work presented here in Chapter 3, the various implementations of TETRIS used throughout this work are shown in Chapter 4. The design of the main user study is discussed according to results of a pre-test and a pilot study of TETRIS in Chapter 5. Then the results of the study are shown (Chapter 6) and discussed in Chapter 7. Finally, all findings are reviewed in relation to current research, together with a deeper look into the process of adapting games to psychophysical data and other options of real-world application of the findings (Chapter 8).

Chapter 2

On the Shoulders of Giants – Foundations of this Work

The question of framing user experience in games with eye movement based adaptation touches several issues that have already been addressed in various other areas. This chapter aims to give an overview of the existing work on these different issues and how they have been discussed individually.

Games in general and how they are played has been a topic in several different fields (Section 2.1). Detection of expertise in games in Section 2.1.1 is of special interest here. Expertise will be discussed for general skill acquisition (Section 2.1.1.1) and with a special focus on TETRIS in Section 2.1.1.2. There is also a vibrant discussion on how user experience in games can be measured (Section 2.1.2). After a thorough investigation aiming to reveal the terms of *presence*, *involvement*, *flow* and *immersion*, focus is given to questionnaires used in previous studies in Section 2.1.2.2.

A short overview of eye tracking in general and interpretation of eye movements relevant to this work is presented in Section 2.2. When discussing the process of adapting software (Section 2.3), the difference between adaptivity and adaptability is the main concern. Finally, framing is discussed in Section 2.4. Different types of research into framing as well as ethical issues attached are of core interest. Section 2.5 brings all these topics, questions and insights together again rephrasing questions of framing of user experience of games with eye movement based adaptation.

2.1 Playing Games

> *"Noone can expect, that everyone uses the same identical term of play as a single word in the same way, just like every language has a word for hand or foot. Here, it's not that easy."*[p.34 Huizinga, 1956]

When we try to talk about games we find ourselves in the curious situation of everyone knowing what is meant but not necessarily sharing the same definition. Games have fascinated cultural scientists for years regarding their social functions and the reasons for the seemingly unreasonable action of games (e.g. Caillois and Halperin [1955], Caillois [2001], Buytendijk [1976], Pias [2002], Deterding [2008]). Recently, *gamification* has been discussed as a way to make everyday experiences more enjoyable (e.g. McGonigal [2011], Deterding et al. [2011] or with a more critical perspective Bogost [2011]). When we talk about games in a scientific reference frame, the definitions of what we mean have to be even more precise, especially when they are so fluid in everyday use. Juul [2011] [cited in Koster, 2013] tries to do so by giving a very generic definition of games: "A game is a rule-based formal system with a variable and a quantifiable outcome, where different outcomes are assigned different values, the player exerts effort in order to influence the outcome, the player feels attached to the outcome and the consequences of the activity are optional and negotiable." However, this statement could be used to describe many things sharing the same attributes. For instance, it applies to research by substituting the word 'player' with 'scientist'.

In this thesis, there is a focus on a specific type of game: digital games and more concisely: digital games without a narrative. This means, that the main goal of interaction is recreational without a clear secondary goal. The game also does not have a narrative storyline. While whether a narrative game has to have a storyline has been debated (cf. Aarseth [2012]), we discard the personal storyline a user creates while interacting with the game in this definition.

2.1.1 Expertise

Since TETRIS is fairly well known and has attracted different players at different levels, a wide range of expertise can be expected among the test participants. However, TETRIS also exists since 1985, so there are expert players who have not played the game for quite a while.

2.1.1.1 Acquiring a Skill

According to Fitts and Posner [1967, cited in Anderson [2000]] there are three stages of skill acquisition.

1. a cognitive stage in which a declarative encoding for the task is created (often by repeated rehearsals).
2. an associative stage in which a learner develops a deeper understanding of the task leading to a more successful performance via iterations.
3. an autonomous stage in which the task is solved in a more automated way usually including faster performance.

The three stages have different durations for a given individual, but according to Anderson [2000], skill development over time can be expressed as a logarithmic function. It takes less time to get from a beginner to a novice level, however, to reach an expert level in a task, much more time is required (see Figure 2.1).

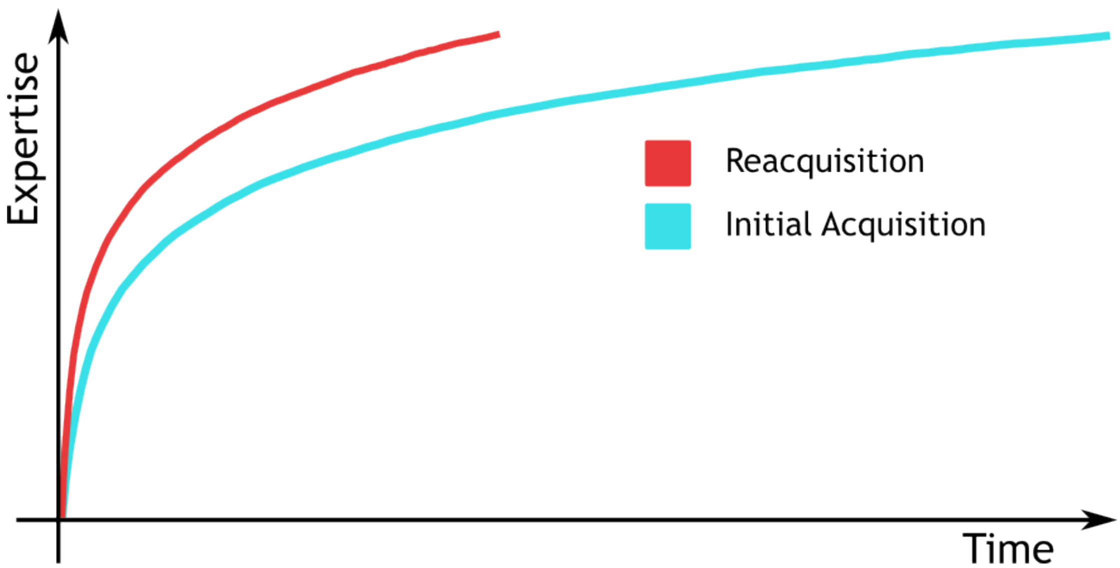


FIGURE 2.1: Development of expertise over time during initial skill acquisition (red) and reacquisition of a skill (blue) adapted from Anderson [2000]

Kolers [1976, cited in Anderson [2000]] showed how test participants acquired the skill of reading an inverted text without prior knowledge and then again a year later. They were able to establish that a previously acquired and then dormant skill can be retrained significantly faster than the initial learning process.

Tetris players are likely to have previously achieved some expertise level. In order to reach that again, they will need to retrain their skills to reacquire this former level (see Section 5.2.3 for a discussion of the solution for this problem that was used in this context).

2.1.1.2 Tetris Experts

Expertise in Tetris has been discussed, with the game used as an example for more general ramifications. Kirsh and Maglio [1994] distinguished *epistemic* from *pragmatic* actions. While pragmatic actions are goal oriented, epistemic actions are performed in order to gain more information about an environment. In the case of Tetris epistemic actions can be seen for example when a player orients their piece, before dropping it, by translating it to an edge and back or rotating the piece more than would be required in order to bring it to a desired position. It has also been claimed by Maglio and Kirsh [1996] that the use of epistemic actions increases for more experienced players.

However, Destefano et al. [2011] conducted tests that showed the opposite for a seemingly wider range of expertise. It might be likely that the use of epistemic action rises during skill acquisition and then declines again as schematically shown in Figure 2.2. They also pointed out that especially in Tetris the classification of epistemic action compared to e.g., switching ones goal can be arbitrary. Epistemic actions, while an interesting concept, are not helpful in establishing a player's expertise level.

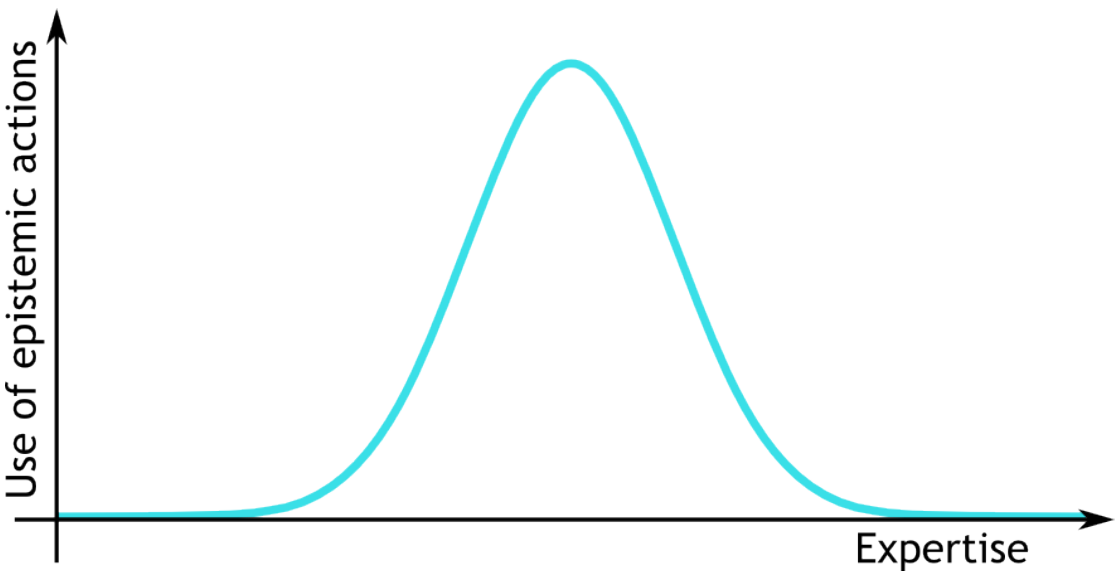


Figure 2.2: Use of epistemic actions according to level of expertise

Another study by Lindstedt and Gray [2013], investigating the relationship of the TETRIS game state and expertise, was not conclusive. While analysing game state helps artificial intelligence agents playing the game (e.g., Böhm et al. [2005] or Flom and Robinson [2005]), it appears to be less helpful when trying to determine a player's expertise.

Jermann et al. [2010] took a different approach: looking at the eye movements of experts and novices playing TETRIS in a collaborative setup. They found out that the ratio of fixations on the current piece to fixations on the contour of the active game (see Figure 2.3) is significantly higher for novices than experts (see for a general discussion of eye movements Section 2.2).

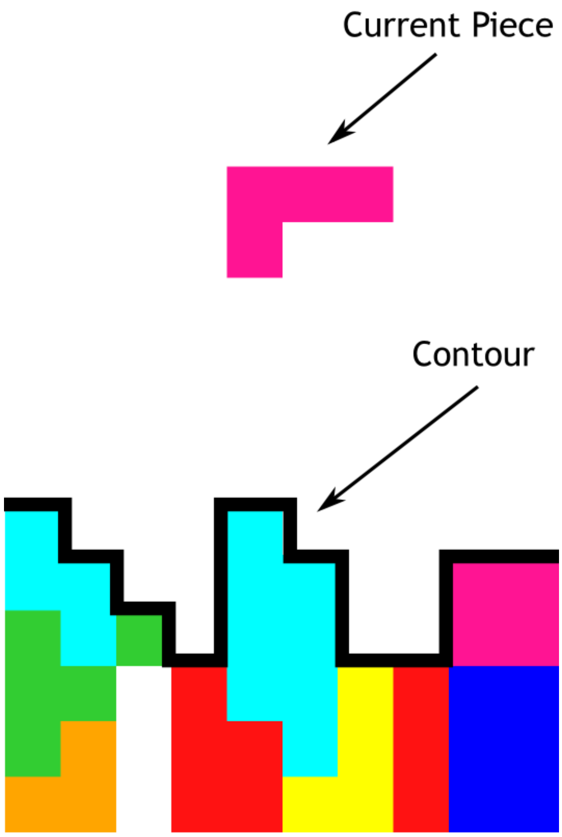


FIGURE 2.3: Areas of interest for a player of TETRIS

It has furthermore been established that playing TETRIS does support spatial expertise, but only for those objects that are integrated in the game [Sims and Mayer, 2002]. From this, it can be concluded that expertise in TETRIS does not generalize and is, hence, a type of very specialised skill or "extreme expertise" [cf. Lindstedt and Gray, 2013].

2.1.2 User Experience in Games

Within this work, the experiences a player has during a game have to be quantified, although "[q]uantifying game play is one of the most challenging research endeavors to attempt" [Appelman, 2007, p.815]. Luckily, within a game of TETRIS, the general task of logging the game state and analysing is fairly trivial (cf. Section 4.1.2). This helps in assessing the objective measures of the experience of playing games. However, since the experience is largely subjective, self-reported data is essential.

Unfortunately, a player's experiences during game play lack one distinctly used descriptive term. Current research uses *User Experience* (e.g., Korhonen et al. [2009] and Nacke et al. [2010]), *Game Experience*[1] (e.g., Poels et al. [2007]), *Player Experience* (e.g., Isbister and Schaffer [2008] and Nacke et al. [2009b]); occasionally also by the same researchers. In this work, the term *User Experience of a Game* is used in order to refer to the experiences of the user interacting within a game. Abstracting the player to a general user might be difficult in other contexts, but the approach presented here and the results of the study are supposed to be generalisable to other user experiences as well.

2.1.2.1 Presence, Involvement, Flow, Immersion

When trying to define user experience in games, a lot of terms are used as reference. The term user experience "[...] can be seen as an umbrella term used to stimulate research in HCI to focus on aspects which are beyond usability and its task-oriented instrumental values" [Hassenzahl, 2005, cited in Bernhaupt [2010]]. Other terms used to describe the specific experience of playing a game are not necessarily clearly defined and distinguished. There have been several attempts to fix this issue within the research community. For example, Brown and Cairns [2004] try to put the term *immersion* into perspective with other terms used. They establish three stages of immersion: *engagement*, *engrossment* and *total immersion*, which, according to them, can be interpreted as the experience that is referred to as *presence* in other works.

Furthermore, *immersion* has been established as a "fundamental component [...] of the gameplay experience" [Ermi and Mäyrä, 2005]. Ermi and Mäyrä's *Gameplay Experience Model* establishes three aspects that are important to children playing games: audiovisual quality and style, the level of challenge and the imaginary world and fantasy. However, the audiovisual quality and style seem to be less important, since "[...] children thought that the emotional immersion and involvement in fiction was typically

[1]This is an especially difficult term, since it can refer to the experience a player has while playing or the playing experience (= expertise) a player has.

stronger for them while reading a good book or while watching a movie" (p.6) putting a heavy focus on the narrative quality of games. This does not generalise well to strongly or purely ludic games (see for a distinction amongst others Aarseth [2012]). However, this critique can be applied to many other attempts to describe the experience (e.g. to Calvillo-Gámez et al. [2010]).

Especially when the focus is on *presence*, this seems to be the case. One example can be found in Takatalo et al. [2006] who requires not only perceptual realness but also a spatially complex enough environment to create spatial awareness for a player. Furthermore, social richness and realism appear to be aspects of presence and involvement in digital games. However, this experience might occur in more abstract games like TETRIS as well, but is not covered by their theory.

Another term that is often used to describe a specific state during playing is the concept of *flow* as developed by Csikszentmihalyi [1991]. However, the concept only covers single moments that occur during gameplay [another one being e.g. frustration, see Gilleade and Dix, 2004] making it only of limited use as a single component of analysis for game experience research [cf. IJsselsteijn et al., 2007].

All in all, the terms immersion, flow in games and presence are used interchangeably and without clear distinctions. In this thesis, user experience in games is seen as an experience that can be measured only partly. Since it's a subjective experience, self-reported measures are required. These rely on language. Even though, language as a medium does not always cover all the aspects of any experience [2], it's a valid measure as the experiences as such are subjective as well.

2.1.2.2 Questionnaires

While flow can also be seen in psychophysical measures (cf. Peifer [2012]) and can even be established by analysis of a players eye movements (cf. Jennett et al. [2008]), questionnaires are still the most reliable tool that a researcher has – with all the issues of self-reported measures (see also for a discussion of skill assessment Section 6.3).

Several questionnaires have been developed in order to assess different aspects of the experience of playing games. Most of them only assess a singular aspect or are only applicable to a certain type of game. All of them, however, share assumptions about games. These assumptions are discussed together with the most popular questionnaires used to assess game experiences.

[2] see for example in film theory the reference to somatic experience, cf. Pantenburg and Schlüter [2014]

The most widely used questionnaire to assess flow, especially in a more physical context, is the Flow State Scale developed in Jackson et al. [1996] and further improved in Jackson and Eklund [2002]. According to Nacke and Lindley [2008] this is the most widely used questionnaire to assess flow in games. Although Kivikangas et al. [2006] showed that the questionnaire is generally usable for game related research, it has been designed with a focus on measuring optimal experiences in physical activities. While there do exist digital games with physical aspects (e.g., within the WII SPORTS series), the questionnaire should be used with care.

Witmer and Singer [1998] presented a presence questionnaire in order to assess the feeling of being within a virtual environment. Methodologically, the questionnaire has encountered harsh critique by Slater [1999]. However, there are also conceptual problems, that make it difficult to use for game research. First, the questionnaire assumes that there is an actual environment, but not every game has a world to speak of. Especially puzzle like games (also like TETRIS) do not rely on graphical prowess. Second, as with the concept of flow, presence is only one possible aspect that can occur while playing digital games. Hence, the use of the questionnaire is limited to only these aspects. However, one very valuable contribution has been their Immersion Tendency Questionnaire which can be used as a normaliser for other measures of immersion. It calculates a value for how immersible a person is in general and puts their immersive experience in perspective.

Another example with a focus on virtual environments has been presented by Takatalo [2002]. While it has not been widely used, it accounts for both flow and presence as experiences in virtual environments. However, the ludic aspects of a specific game experience are not part of this questionnaire in order to make it suitable for a wider approach of digital experiences. This leads to the same problems for game research as both single focus questionnaires above, since an environment is expected, within the assessment of presence and, when assessing flow experiences which only might occur during part of the game are covered.

Vorderer et al. [2004] developed a questionnaire that focuses on presence grounded in a theory of *Spatial Presence*. It is a generic questionnaire for all possible media experiences ranging from passive consumption like books to active participation like digital games. Parts of it might be more suitable to game research than the other questionnaires – at least for certain games. The questionnaire consists of seven parts, each of them signifying one value on a 4-item, 6-item or 8-item scale. The categories of attention allocation and higher cognitive involvement are especially suitable for any game, whereas domain specific interest and suspension of disbelief are tied to narrative games only. The other categories (spatial situation model, spatial presence and visual spatial imagery) expect games that have an environment. Helpfully, the questionnaire is not only available in

English, but has been professionally translated into German, Portuguese and Finnish. However, since TETRIS is neither a narrative game nor provides the player with an immersive environment, the questionnaire has been deemed too limited for the current research.

A 19-item questionnaire with a focus on absorption, flow, presence and immersion in video games has been offered by Brockmyer et al. [2009]. However, different aspects are given differing importance. For example, there exists only one question that the authors attribute to immersion while several questions ask about the temporal perception during play. Additionally, the questionnaire was created with having a focus on violent semi-realistic video games and their effects on younger players. While the questions might be suitable for TETRIS at least in a general way, at least one item ('The game feels real') would have to be removed because the question does not make any sense to be asked in a purely ludic game with only an abstract environment. Also, the questionnaire is not available in shorter versions that would make it possible to ask questions for episodes of play.

A single parameter or term to describe user experience in games does not fully cover the whole spectrum of experiences that occur while playing games (cf. Nacke et al. [2009b] or Poels et al. [2007]). Questionnaires that only focus on flow or presence as a singular experience are limited in their use for this research, since flow is not a long term experience and presence is related to semi-realistic graphics or descriptions (see for a discussion of this issue in text based games Spiel [Chapter 2, 2012]) that represent an explorable world. Multi-parametric measures are more fruitful than those focusing on a single aspect of the experience. Nacke et al. [2009a] combines the concepts of immersion, presence and flow and influenced IJsselsteijn et al. [2013]'s Game Experience Questionnaire (GEQ).

The GEQ consists of a set of modules: the core module, the in-game module, the social presence module and the post-game module. The core module (33 items) and the in-game module (14 items) are closely related; the latter being a shortened version of the first in order to ask for experiences between episodes of play. Both of them evaluate seven components, albeit with a different number of items. These components are *competence*, *sensory and imaginative immersion*, *flow*, *tension/annoyance*, *challenge*, *negative affect* and *positive affect*. With flow describing the relationship between the challenge that is being offered to a person compared to their abilities and challenge being again its own category, some categories might have a slight overlap. This might be negligible seeing as the GEQ as a whole tries to cover multiple aspects of the experience of playing digital games. The social presence module (17 items) is only relevant to games with a social aspect, be it multiplayer or interaction with non-player characters (NPCs). Since both are not the case in TETRIS, it is not of interest for this research. The post-game module (17 items) then asks for how the player felt after the game but also about playing from a hindsight perspective. It can be used to get more data about the experiences after a player had time to reflect them more. The components of this module are *positive experience*, *negative experience*, *tiredness* and *returning to reality*.

Next to the already mentioned issue of overlapping components, the GEQ also has assumptions about a game. The GEQ expects a game to have a narrative structure (item 1 on the in-game module and item 3 in the core module: "I was interested in the game's story") and – to a somewhat lesser but still palpable extent – to have a graphical interface. The items tend to fit better when the game consists of an explorable world. These are issues when using the questionnaire for TETRIS. So while this is the best option for a self-reported experience measure, slight modifications are necessary to make it work with TETRIS (see for a description of these modifications Section 5.2.2).

2.2 Tracking Eyes

Where people look and how they move their eyes has been of interest to researchers for about 300 years starting with Porterfield and Wells in the middle of the 18th century [Wade, 2000]. Since then it has been established that there are two basic types of eye movements: *fixations*, which happen when the eye is relatively still and focused on a location, and *saccades*, which are swift switches between fixations. While fixations last between 200-300 ms [Holmqvist et al., 2011, p.23] (although ranges of 150-600ms seem to be reported as well [Duchowski, 2003, p.47]) saccades are very short ranging from 30-80ms [Holmqvist et al., 2011, p.23]. When there is moving target followed at semi-constant speed, this is called *smooth pursuit*. During TETRIS there is no smooth pursuit, since the pieces move in a step-wise fashion.

Fixations are of special interest to researchers looking into eye movements, because they tell us about where a person's attention is – at least to some degree. Dispersion of visual and attentional focus only rarely occurs [see for a more detailed discussion Bertel, 2010, Chapter 4].

There also has been interest in how to measure these eye movements ranging from simple observation over electrooculography scleral search coils[3] to the state-of-the-art eye tracking method today: video-occulography [see for a more complete and detailed history of eye movement research Wade and Tatler, 2005]. With this method a remote

[3] According to Bertel [2010], Chapter 5, still the frame of reference in terms of precision in eye movement recording, albeit a bit cruel to test participants, since a wire is physically attached to their eyes.

or head-mounted eye tracker records the movement of the eyes with a high frequency camera. For detection of every tremor of the eye, this frequency should be at least 250 Hz, although different eye trackers for different research purposes and use cases have frequencies ranging from 25-2000Hz [Holmqvist et al., 2011, p.30]. In order to account for head movements, the Purkinje reflection of an emitted infrared light next to the camera is commonly used. When this reflection can be detected, its local relationship in comparison to the pupil can be used to account for small head movements. While most eye tracking is done with a single eye – often the dominant eye [cf. Chaurasia and Mathur, 1976] – bioccular eye tracking is also common and used within this research whenever the Purkinje reflection was not clear enough for the eye tracker. For a more complete discussion of eye movements and eye tracking in general, please confer Duchowski [2003] and/or Holmqvist et al. [2011].

Eye Movement	Description	Analysis	Interpretation	Source
Fixation	relatively stable focus (150-600ms)	number of fixations	experts have overall less fixations	Megaw and Richardson [1979]
Focus Area	point of interest on fixation	closely inspected areas	experts have more relevant fixations	Singer et al. [1996]
Transition	change of point of interest	number of changes occurring	the more changes, the more uncertain a user	Goldberg and Kotval [1999]
Saccade	quick change of focus (30-80ms)	saccadic amplitude	larger amplitude indicates better understanding	Goldberg et al. [2002]
Scanpath	direction of the change of focus over time	direction of saccade (horizontal vs. vertical)	experts show more horizontal eye movements in TETRIS	Underwood [2005]

TABLE 2.1: Interpretation of Eye Movements as Relevant for Adapting TETRIS and the Analysis of the User Study (see also Chapter 4 and Chapter 6)

Within this work, analysis of *fixations*, *focus areas*, *transitions*, *saccades* and *scanpaths* (see Table 2.1) are of special interest, because they are either used for adaptation of TETRIS or for the analysis of the user study. For an overview of how certain eye movements and their derivatives might be interpreted is given by Poole and Ball [2006] and – more extensively – in Holmqvist et al. [2011].

Eye movements in computer science have been commonly used as a usability measure (exemplary, Cowen et al. [2002] or Goldberg and Wichansky [2003]) or input device, if there are no other input options available (exemplary, Huckauf and Urbina [2008] or Glücker et al. [2014])

2.3 Adapting Software

When research speaks of adapting software, this usually refers to the concept of *adaptivity* of a system in contrast to the *adaptability* of a system as described by Oppermann and Rasher [1997]. Adaptivity and adaptability are – according to them – two ends of a scale of user involvement in individualizing an interactive system. An adaptive system or the adaptive parts of the system are not user controlled whereas an adaptable system or adaptable parts of a system lie under the full control of a user. Examples for adaptable parameters are e.g. the change of a background colour or – specifically in games – the choice of an avatar and/or its appearance.

For games, this distinction holds for some types of adaptivity. All methods described by Charles et al. [2005] can be analysed with this frame. Adaption to the game environment or state happens via system initiated models. While the player is the source of the adaptive calculations, they do not control them. In the terms provided by Oppermann and Rasher [1997], this is a highly adaptive system, since the player can derive information about the adaption process via the game state. Adaption to non-player characters can be either fully adaptive (as a reaction to player performance) or fully adaptable (via player-set settings such as *easy*, *medium* or *hard*). Adaption to a player's character, however, falls outside of this scheme. A first reason is, that adapting to a player's character can influence that player's behaviour and hence their character and then again the adaption process. This results in a loop of mutual influence. A second reason is that players can possibly influence the way a system perceives their character by deciding upfront which kind of character they want to exhibit as soon as they know the game adapts to that. This is not a direct influence and players are not directly informed about the system changes. Hence, the scale of adaptability vs. adaptivity is insufficient to describe what happens in this case.

Adapting on psychophysical data falls into a even more difficult category. But not every psychophysical measurement is the same. While it is e.g. hard, but possible to control one's heart rate [cf. Hirsch et al., 1981], skin conductance cannot be influenced consciously. Eye movements are again special in this regard. Most eye movements are not consciously controlled, but rather directed by attention. However, viewing strategies exist and can be used. In competitive games, players sometimes try to hide where they are focusing on in order to not tell their opponent too much about their own strategies and, hence, try to use their eye movements most efficiently [Reingold et al., 2001]. As well as such an obvious attempt at control, viewing strategies exist also for areas such as art [Zangemeister et al., 1995]. Adapting a game like TETRIS to eye movements follows the adaption to player character more than any other adaption, but still has its own momentum by being only partly controllable by the player.

In order to describe the adaptivity of eye movements productively and having the nomenclature of Oppermann and Rasher [1997] in mind, a new term and possibly also a new dimension have to be added to the concept. If divided into user control and user information, user involvement could be differentiated more effectively.

2.4 Framing Experiences

The effect of framing on people in general has been studied in detail within social science and psychology. Most of the research in sociology is investigating political communication within a society that has mass media (e.g. Scheufele [1999] or Iyengar and Kinder [2010]). Psychology focuses more on the generalisation of "frames as informationally equivalent labels" [Scheufele and Iyengar, 2012, p.2]. In general, there are three different types of framing according to Levin et al. [1998].

The first type of framing is *risky choice framing*. This investigates risk preference [see also Kahneman, 2011, pp. 334] under a different highlighting of prospects. For example two frames could be 'You are going to get 1000$ with a 20% chance' vs. 'You are going to lose 100$ with an 80% chance' show statistically different outcomes in how people react to the choice they have.

The second type of framing is *attribute framing*. Here, researchers show an item in either a positive or a negative light and see how the evaluation of the item changes. Cases – similar to the research presented here – where certain attributes of an item are highlighted or not, without a judgement of positivity or negativity, are rare.

The third type of framing is more directly tiered to what sociology does. It's called *goal framing* and is interested in whether people adopt behaviour or opinions after they have been exposed to persuasive examples.

Attribute framing is the closest description of what is presented here. The test persons are informed about all means of adaption (including eye movements) or only partly (not mentioning eye movements). Similar research into framing interactive experiences has been done by Hartmann et al. [2008] for websites.

Ethical issues arise. Since some test persons are informed about how the software presented to them works and some are only partly informed, it is important to be sensitive toward the test persons and ensure full disclosure at the end of a test session [see also for a more general discussion Berg and Lune, 2004, Chap. 3]. Test participants still have the option of not having their data used after the full purpose of the study has been revealed to them.

2.5 Framing the User Experience of Games with Eye Movement Based Adaptation

This chapter has shown how the questions investigated in this work touch several subjects in different fields of research and science such as cultural theory, psychology, eye movement research, usability and computer science. We know now how to deal with the question of what a game is in this context, even though there isn't necessarily a clear answer for that question. However, games can be approached as a pleasing interactive experience that does not follow a clear goal.

Furthermore, we discussed how expertise and especially expertise in Tetris is formed and can be detected. With the Game Experience Questionnaire by IJsselsteijn et al. [2013], there is a tool to measure the self-reported game experiences a player has during play, even though not all assumptions the test has about games are met by Tetris.

There has been a short presentation of eye movements, how they are measured and how they can be interpreted for the purpose of this research. There has also been a presentation of the process of adaptation. Eye movement based adaptation in this context can hardly be influenced by the player and follows more the concept of adaptivity. And finally, there has been a short wrap up about types of framing and how this research follows into a type of attribute framing.

This knowledge helps comprehending the nature of the research question on the effect of framing the user experience of games with eye movement based adaptation compared to the actual use of eye movement based adaptation better.

Chapter 3

Research Hypotheses

Within a general discourse of growth and development in computer science as well as the tech sector, there is often a very positive attitude towards new ideas incorporating additional technology. However, within this positivity, critical analysis of actual use and effect of these technologies on its users appears to be neglected. This work asks the question on whether the simple suggestion of a new use of technology is enough to increase the performance and experience of a player and whether their viewing and acting patterns differ from other players who have not been informed about the use of technology. Furthermore, the actual effect of the implementation of this technology is analysed by seeing how both groups of players perform, experience and act in games with eye movement based adaption and in games without eye movement based adaption.

3.1 Performance Based Hypotheses

While performance is not the only reason for a joyful game experience, it heavily contributes to it (see Wetzel et al. [2014] for a discussion on their study). Hence, the effect of both framing as well as eye movement based adaption on performance has to be considered.

Does framing the adaptivity of the game have a positive effect on a player's performance?
H-P1$_{\mathrm{RH}}$: Framed players perform better than non-framed players.

Does eye movement based adaptation have a positive effect a player's performance?
H-P2$_{\mathrm{RH}}$: Players perform better in eye movement based adapted games than those playing a conventionally adapted game.

Which has a greater effect on a player's performance – adapting or framing a game?

H-P3$_{\mathrm{RH}}$: There is a greater effect on the performance of framed players than that of non framed players even even if they play a conventionally adapted game.

Does eye movement based adaptation also improve a framed player's performance?

H-P4$_{\mathrm{RH}}$: Framed players perform better in games with eye movement based adaptation than in games with conventional adaptation.

H-P5$_{\mathrm{RH}}$: The delta in performance of framed players in games with eye movement based adaptation compared to conventionally adapted games is smaller than the delta of performance of non-framed players in games with eye movement based adaptation compared to conventionally adapted games.

3.2 Gameplay Experience Based Hypotheses

While the gameplay experience, as it consists of multiple aspects of several experiences, is hard to assess adequately (see Section 2.1.2.2), players are expected to report on changes in their gameplay experience with positively or negatively connotated consequences. This is supposed to change according to whether they play games with or without eye movement based adaption and also whether they know about the eye movement based adaption or not.

Does framing have a positive effect on a player's gameplay experience during the game?

H-U1$_{\mathrm{RH}}$: Framed players have a better gameplay experience than non-framed players.

Does adapting the game based on eye movements have a positive effect on a player's experience during a game?

H-U2$_{\mathrm{RH}}$: Players of a game adapted on eye movements have a better gameplay experience than those playing a conventionally adapted game.

Which has a greater effect on a player's experience during a game – adapting or framing?

H-U3$_{\mathrm{RH}}$: Framed players have a better gameplay experience while playing than non-framed players, even if they play a conventionally adapted game.

Does eye movement based adaptation also improve a framed player's gameplay experience?

H-U4$_{\text{RH}}$: Framed players have a better gameplay experience in games with eye movement based adaptation than in games with conventional adaptation.
H-U5$_{\text{RH}}$: The gameplay experience of framed players does not improve as much as that of non-framed players when playing a game with eye movement based adaptation compared to a conventionally adapted game.

3.3 Expertise Based Hypotheses

Furthermore, whether the type of adaption, and whether players know about it, influences their interaction with the game is of interest within this work. It might be, that players show different behaviour when interacting with different types of games under different pretenses. One aspect of this different behaviour is expert behaviour: do players interact more confidently with the game they are playing?

Does framing support players to show more expert behaviour?
H-E1$_{\text{RH}}$: Framed players will show more expert behaviour than non framed players.

Does an eye movement based adaptation of a game support the development of expert behaviour for a player?
H-E2$_{\text{RH}}$: Players of games which are adapted according to eye movements show more expert behaviour than those of conventionally adapted games.

Which has a greater effect on the expert behaviour of players – adapting or framing?
H-E3$_{\text{RH}}$: Framed players show more expert behaviour than non-framed players, even if they play a conventionally adapted game.

Do framed players exhibit more expert behaviour when playing a game with eye movement based adaption?
H-E4$_{\text{RH}}$: Framed players will show more expert behaviour in games with eye movement based adaptation than in games with conventional adaptation.
H-E5$_{\text{RH}}$: Framed players will have less of a delta in expert behaviour than non-framed players when playing a game with eye movement based adaptation compared to a conventionally adapted game.

3.4 Intersectional Hypotheses

While this research has a focus on the effects of eye movement based adaption in games and their framing, in some cases one cannot say where an effect actually comes from.

In order to have a well-rounded picture of the influence of framing and eye movement based adaption at their own terms, some questions have to be asked that investigate the relationship between the hypotheses groups above.

Does a better performance positively influence a player's experience during a game?
H-I1$_{\mathrm{RH}}$: The better the performance of a player the better their experience.

Does a better performance correlate with more expert behaviour?
H-I2$_{\mathrm{RH}}$: The better the performance of a player, the more expert behaviour they show.

Do players who show more expert behaviour enjoy the game more than others?
H-I3$_{\mathrm{RH}}$: The more expert behaviour can be detected for a player, the better their gameplay experience.

Chapter 4

Tetris

As the reference game there will be a general description of TETRIS (4.1) including the mathematical attributes (4.1.1) and specific options of game state analysis (4.1.2). Subsequently, different versions of block selection algorithms are discussed and compared (4.2). These are NICETRIS (4.2.1), GRAB BAG (4.2.2), TRUE RANDOM (4.2.3), SKEWED RANDOM (4.2.4) and BUST HEAD (4.2.5). Furthermore, three different implementations (4.3), PYTRIS (4.3.1), NEMTRIS (4.3.2.1) and EMTRIS (4.3.2.2) are presented.

4.1 General Description

TETRIS was developed by Alexey Pajitnov and Vadim Gerasimov during 1985/1986. It is seen as an important part of computer game history [Sheff, 1993]. With its easy rules, even novice players can quickly understand the game.

The main game focuses on a field that is 20 blocks high and 10 blocks wide. Small tokens called *tetrominoes* appear on top of the game field and fall down stepwise (see Figure 4.1). The time frame from appearance of the block until it settles in a place is called an *episode*. The blocks have to be arranged in such a way that they create lines which increase the score and are then removed from the field (cleared rows). The more lines removed in one episode, the higher the score awarded. At most four rows can be cleared at once. When this happens, a *tetris* occurred. As soon as the pile reaches the top of the field, the player loses. During gameplay, the speed with which tetrominoes appear and move down increases with each row cleared.

A TETRIS player can perform one of three actions: *rotating* the currently falling tetromino clockwise or counter-clockwise, *translating* the current tetromino to the left or the right or *dropping* the piece, where the current piece falls down as far as it can onto

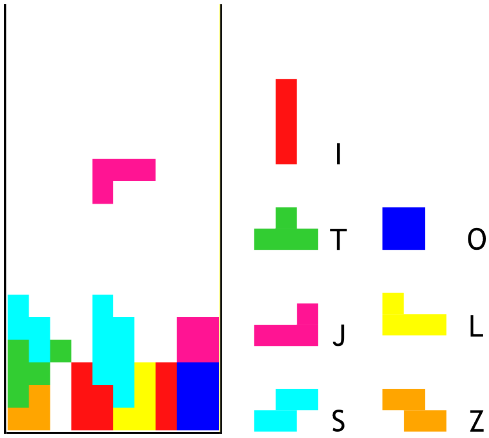


FIGURE 4.1: Illustration of a game of TETRIS in play (left) and the tetrominoes that might fall down during the game (right)

the contour of the pile. This means that the player is limited to a very small range of possible interactions with the game making it easier for a researcher to observe the game state.

Furthermore, TETRIS is fairly well-known, which makes it easy to find a sufficiently large population of interested test participants with different levels of expertise.

4.1.1 Mathematical Attributes

The question of whether you can win at all when playing a game of TETRIS has not only been asked by the players themselves, but also answered by Brzustowski [1992]. They showed that there exists no winning strategy in TETRIS. Since the only winning strategy could be an endless game, they have effectively proven, that there exists at least one kill sequence that ends a game of TETRIS for a player. This has been generalised by Burgiel [1997] showing that this happens independently from the interplay between computer and player. The computer does not react to the player's move and, hence, can be assumed as true for a wide variety of implementations of the game as long as they contain ß[1] pieces.

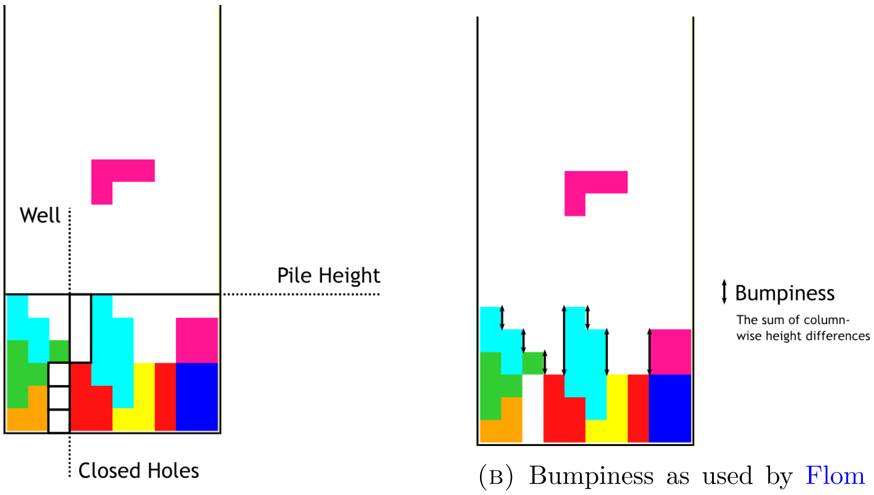
Furthermore, Demaine et al. [2003] proved the NP-completeness of various strategies a player of TETRIS could have. These include maximising the number of cleared rows, maximising the number of tetrises, minimising the pile height or maximising the number of pieces appearing before the game ends.

[1]The German letter ß is used to denote S and Z pieces as per Figure 4.1. Whenever this letter is used, it refers to both, S and Z pieces.

TETRIS is thus a game that will end eventually, even if it is not defined when. There exists no winning strategy for a player. Furthermore, determining optimal play according to several strategies is NP-complete.

4.1.2 Game State Analysis

While TETRIS is a game that works with a temporal concept of real time, the process is divided into distinct episodes for each block. This makes logging a game and analysing it particularly easy, because snapshots of the game state are enough to extract the most important parameters. But which are those?



(A) Parameters used by Fahey [2012]

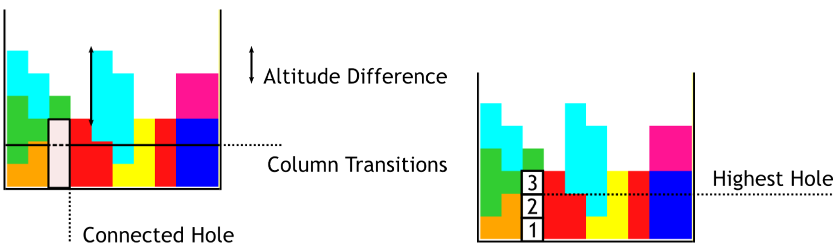
(B) Bumpiness as used by Flom and Robinson [2005]

FIGURE 4.2: Parameters for TETRIS Game State Analysis

In order to enable artificially intelligent agents to play TETRIS, Fahey [2012] established *pile height*, which describes the height of the highest point of the contour, *number of closed holes*, which counts the number of unreachable areas with size $1x1$ under the contour, and *number of wells*, which are deep narrow spaces that can only be filled by an I block in order to clear one or more rows, as defining aspects of the quality of a game (see also Figure 4.2a). Flom and Robinson [2005] added a measure of *bumpiness* calculating the height differences along the contour (see also Figure 4.2b). They also considered *lines cleared* up to that point as a parameter that describes the quality of a game.

A more fine-grained approach can be found with Böhm et al. [2005]. As well as all of the parameters described so far (except bumpiness), their algorithm considers

- *connected holes*, where holes that border on holes are counted as one hole instead of two (cf. Figure 4.3a),
- *altitude difference*, the difference between the highest and the lowest point in the contour (cf. Figure 4.3a),
- *maximum well depth*, the depth of the deepest well,
- *landing height*, the height at which the last block has been placed,
- *occupied*[2], the number of currently occupied cells,
- *weighted occupied*, where occupied cells on a higher level are counted with a higher coefficient than occupied cells on a lower level,
- *row transitions*, calculating the sum of all transitions between occupied and unoccupied cells on a horizontal axis,
- *column transitions*, calculating the sum of all transitions between occupied and unoccupied cells on a vertical axis (cf. Figure 4.3a).



(A) Selection of parameters used by Böhm et al. [2005]

(B) How Shahar and West [2010] deal with holes

FIGURE 4.3: Parameters for TETRIS Game State Analysis Cont.

Shahar and West [2010] additionally used the concepts of *weighted holes*, in which the holes are weighted according to their height, *highest hole*, which records the height of the highest placed hole (cf. Figure 4.3b), and *game status* which indicates whether the game is still running or has been lost already.

All of these parameters only depend on the current state of the game or its past (for lines cleared), so they only have to be calculated once per episode. This not only allows for episode-wise analysis of a game, but also for a comparatively fine-grained adjustment of the game while not disrupting a single episode. Since players expect a change of

[2]This attribute is called *blocks* in the original source. However, this term has been deemed too ambiguous in this context, so it was changed.

difficulty only after they have cleared a row, it might otherwise be confusing, if the game obviously reacts to anything other than placing a tetromino.

In order to analyse expertise in TETRIS, Lindstedt and Gray [2013] divided their metrics into different scales of human action following Newell [1990]. They divided possible metrics into *global metrics* like all of the ones above, *local metrics*, evaluating the position in which the current tetromino has been placed, and *immediate interaction metrics*, accumulating information about the interaction during an episode. While the metrics in the last group are calculated throughout an episode, the actual analysis of all these metrics can only be performed per one or more episodes. All in all they use eight different global metrics – some of them equivalent to the ones above –, six different local metrics and seven different immediate interaction metrics. The most expressive of those were none of the global metrics, *matched edges* effectively decreasing bumpiness, *filled wells*[3] and *filled overhangs* on the local scale and finally *total rotations*, *drop ratio* (comparing the original speed with the actual speed achieved via player interaction) and *drop latency*, the time measured from the appearance of a tetromino until it is being dropped by the player.

While there have been lots of approaches on how to actually measure the game board, with all of their findings taken together, they are non-conclusive in their usefulness for actual game state analysis, which is why some have been analysed within the Pre-Study according to how they correlate with the actual performance of a player (see Section 5.1).

4.2 Choosing Blocks

The heart of any TETRIS game is the algorithm choosing which blocks are spawned. Next to obvious solutions like a random choosing of all available blocks (see also below 4.2.3), there have also been propositions of using TETRIS games as means of communication by encoding information in how the blocks are chosen. This selection of blocks is then still supposed to appear random or at least semi-random to an unknowing player (see for an example, e.g. Ou and Chen [2011]).

In this section, five algorithms are presented with the general idea behind how they choose blocks together with a sample implementation in Python.

[3] named *uncovered pits* in the original source; changed here for consistency

4.2.1 Nicetris

In order to offer players an algorithm that helps them learning the game and play it with a minimum amount of challenge, the NICETRIS algorithm has been designed using the inverted principles of the BUST HEAD (see below 4.2.5) algorithm. By analysing the situation of the current game board and especially the contour, this approach ensures that the player always has an edge-fitting option to place the current tetromino and, hence, enable them to clear rows quickly. `situation[4]` in Listing 4.1 refers to an array containing all shapes, that fit into the current pile. If the contour is really not fitting for any piece (which is theoretically impossible), the choice is made from the set of generally well fitting blocks (O, I, and L-blocks).

```
nice_bag = []
for element in bag:
    if element in situation[4]:
        nice_bag.append(element)
if len(nice_bag) > 1:
    return choice(nice_bag)()
else:
    return choice([I,O,L])()
```

LISTING 4.1: Sample Implementation of the NICETRIS Algorithm in Python

4.2.2 Grab Bag

According to Khandaker [2011], this is the original TETRIS algorithm. All possible tetrominoes are put in a bag and drawn randomly one after the other without replacement, which means, until the bag is empty. Then, another bag is opened for the next seven pieces (cf. Listing 4.2).

```
if len(ungrabbed_bag) == 0:
    ungrabbed_bag = [ O, I, S, Z, L, J, T ]
block = choice( ungrabbed_bag )
ungrabbed_bag.remove(block)
return block()
```

LISTING 4.2: Sample Implementation of the GRAB BAG Algorithm in Python

This should create a fair random game. With this algorithm there are at most 12 pieces between two I tetrominoes and a maximum of four ß pieces can come in a row. Hence, the chances of encountering a run of the same (bad) pieces are lowered.

4.2.3 True Random

The most basic algorithm for choosing blocks in a game of TETRIS is the TRUE RANDOM version (cf. Listing 4.3).

```
        return choice( O,I,L,J,T,S,Z )()
```

LISTING 4.3: Sample Implementation of the TRUE RANDOM Algorithm in Python

The piece selection is random and independent. Very fortunate and very unfortunate series of blocks are equally likely.

4.2.4 Skewed Random

In order to increase the likelihood for a kill sequence as described by Burgiel [1997], SKEWED RANDOM assigns a 50% chance to either of the ß pieces instead of the likelihood of $2/7 = 28.57\%$ like this is the case for TRUE RANDOM (cf. Listing 4.4).

```
        if random.randint(0,1) == 0:
            return choice([S,Z])()
        else:
            return choice([O,I,L,J,T])()
```

LISTING 4.4: Sample Implementation of the SKEWED RANDOM Algorithm in Python

Mild Skewed Random

A milder version of this algorithm adapts SKEWED RANDOM as such, that the likelihood for ß pieces is at 39%. This version is expected to still be more difficult than the previously described algorithms (NICETRIS, GRAB BAG and TRUE RANDOM) while being less harsh than the 50% version.

4.2.5 Bust Head

This algorithm is inspired by the BASTET game developed by Poloni [2012]. While their version relies on well analysis, the version used in this research is based on contour analysis. First, the algorithm checks which pieces do not fit the contour (fitting pieces are recorded in an array in `situation[4]` in Listing 4.5). Then, a bag of these pieces plus the O piece is used to randomly choose the next tetromino. If every possible tetromino could fit the contour, the undesirable combination of O and ß pieces is used as a bag for the next elements. This procedure should increase the likelihood of a kill sequence as described by Burgiel [1997].

```
    tiny_bag = [O]
    for element in bag:
        if element not in situation[4]:
            tiny_bag.append(element)
    if random.randint(0,2) > 0:
        if len(tiny_bag) == 1:
            return choice([O,S,Z])()
        else:
            return choice(tiny_bag)()
    else:
        return choice(bag)()
```

LISTING 4.5: Exemplary Implementation of the BUST HEAD Algorithm in Python

While BASTET and BUST HEAD share the same core idea, which is to create a really difficult game of TETRIS, their methods of achieving this differ. Hence, the algorithm used here is differently named, but phonetically similar in order to keep the roots of the idea in mind.

4.3 Implementations

In order to fulfill different purposes for different test scenarios, three special versions of TETRIS have been implemented. These have been designed having broad data recording in mind, so that there is data available for research questions that might arise later or come from a different angle.

4.3.1 Pytris

PYTRIS has been implemented in PYTHON 2.7 using the PYGAME library[4]. This made it possible to account for the desired goal of having test participants create a realistic game environment on their own computer.

4.3.1.1 Game Mechanics

PYTRIS offers all game mechanics found in common versions of TETRIS. Players are able to translate, rotate and drop tetrominoes coming down. Additionally, they can listen to the original TETRIS-score while playing, if they desire to do so.

[4]see http://www.pygame.org/news.html

Upon starting the script, PYTRIS captures the screen in a full-screen mode so that players are not visually distracted by other programmes. Graphically there are also no further distractions from the game itself. It does not show a score or a number of lines made indicating performance directly (see also Figure 4.4).

The initial speed of the game is set to 400 ms, which means the tetromino is moved every 400 ms. Speed increases in steps of 5 ms happen whenever a line is removed. If more than one line is removed at the same time, the speed increase is multiplied by the number of lines removed. The minimum speed is set to 75 ms. These parameters can be changed easily within a config file, if needed.

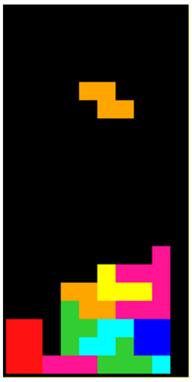


FIGURE 4.4: Graphical interface of PYTRIS during an active game

For each game, the algorithm for choosing pieces is chosen at random without replacement from a bag holding all five basic algorithms described above twice. This means, each player plays ten games of TETRIS, two per algorithm, in a random order.

A game ends after five minutes or when a player loses – whichever event comes first. The script pauses at that point in order to give players time to fill in a questionnaire or have a self paced break between games. However, during these breaks the screen is still captured preventing players to do any other computer based activity, at least at the machine they are playing PYTRIS on. Due to the time restrictions for each individual game, players play a maximum of 50 minutes during one session.

4.3.1.2 Data Recording

PYTRIS has been designed in order to record a variety of data. For each game in a session, a log file is created. After a test session concludes, these files can be retrieved in a folder named by the participant ID.

Every entry in a log file has a time stamp. A log file consists of a header for general test data such as the participant ID defined in the player setup, the number of the game (ranging from 1 to 10), which algorithm has been used in this game, whether the player had the original TETRIS score activated and finally, when the game started. The initial speed and every speed change are recorded as well.

For each new block, a situational analysis is performed. This consists of a count of lines made so far into the game, the current pile height, the current bumpiness measure, the current number of closed holes and for which type of tetromino there are possible placements on the current contour. Then the chosen block and the current grid are recorded as well as every keyboard interaction of the player.

4.3.1.3 Version for Skill Determination

PYTRIS has been slightly modified in order to use it for skill determination in the main study (see also Section 5.2.3). For that, the program is split into two parts. The first part lets a player play with the same constraints as described above, but only three games and only using the GRAB BAG algorithm. Additionally, the speed increases faster than in the original version, so that players are challenged sooner (starting at 400ms intervals and increasing by steps of -10ms). Games are also shorter, only lasting 150s. The second part does the same with only two games, giving them the game identifiers 4 and 5.

4.3.2 (N)EMtris

For the main research study two versions of adaptive TETRIS have been implemented. Possible adaption modes in TETRIS are the choice of the algorithm for choosing blocks and the speed with which the tetrominoes fall. Both modes are used in both versions of (N)EMTRIS.

While NEMTRIS considers the current state of each episode and, hence, the previous and current actions of a player, EMTRIS additionally considers deviated measures from fixations during the last episode. Both have been implemented in C# and incorporated into an in-house framework for eye tracking related research.

4.3.2.1 Adaptivity of NEMtris

Adaptivity in NEMTRIS relies only on the current state of the board after an episode. As we can see later on (see Section 5.1.3), the values for bumpiness and pile height describe the difficulty of the current situation best.

Value for Bumpiness	Algorithm
<= 14	Bust Head
15|16	Skewed Random
17|18	Mild Skewed Random
19|20	True Random
>= 21	Grab Bag

Table 4.1: Adapting to Bumpiness - relationship of numeric values to algorithms

Table 4.1 shows how bumpiness influences the algorithm. The specific numeric values have been derived in a data-driven fashion from values that were found in the pre-study by taking the mean for each algorithm (cf. Table 5.4) and adjusting within the global minimum minus the standard deviation and maximum plus the standard deviation. Since the bumpiness is strongly influenced by the pieces that are available to the player it is more suitable for algorithm adaption than speed adaption.

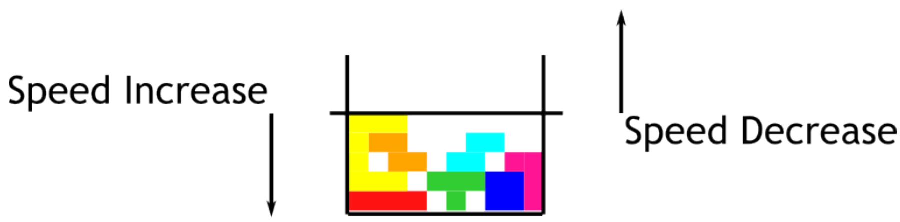


Figure 4.5: Adaptation to Pile Height - abstract illustration of concept

The speed difficulty becomes even harder when the pile rises to a certain height. While the criticalness of this height might differ for each player, the principle holds. In NEMTris adaption to pile height only starts when the pile has reached a height of five rows in the tallest column. This value has also been determined by the pre-study considering the minimum for the easiest algorithm minus the standard deviation (cf. Table 5.3). After this threshold value has been reached, speed adjustments happen whenever the pile height changes in a reciprocal relationship. This means that speed increases when the pile height decreases and speed decreases when the pile height increases (see also Figure 4.5). The steps for this adjustment are two milliseconds, while this value gets multiplied with the number of rows that contribute to the increase or decrease of the pile height. Additionally, whenever a player removes a line, there is an obvious change of speed of ten milliseconds (multiplied with the number of lines removed), so that the game exhibits expected behaviour for the player. This means that there is a speed increase of twelve milliseconds (times number of rows removed) whenever a player removes rows and a speed decrease of two milliseconds (times the pile height difference) whenever the pile rises at its maximum.

4.3.2.2 Adaptivity of EMtris

While EMTRIS makes use of the adaptivity mechanics of NEMTRIS, it additionally uses data derived from eye movements in order to adapt the speed of the game and the algorithms for choosing blocks. While the derived measurements have been established in a theory-driven way, the threshholds have been determined according to data from the pilot study.

Transition Value	Algorithm
$<= 0.00414594$	BUST HEAD
0.00666522 - 0.00414594	SKEWED RANDOM
0.01032776 - 0.00666522	MILD SKEWED RANDOM
0.02058082 - 0.01032776	TRUE RANDOM
> 0.02058082	GRAB BAG

TABLE 4.2: Adaption to Transition Value - mapping of percentiles to algorithms

The transition value describes the relationship of fixation transitions between areas of interest (AOIs). There are in total eleven AOIs on a field of (N)EMTRIS (see also Figure 4.6). Eight of them are simply the available background space tiled into equal parts. Additionally there are three special AOIs: two around the contour and one as a bounding box of the tetromino. The region of the contour AOIs – which are divided at the center of the board into *contour left* and *contour right* – is determined by taking the pile height of a column and adding the cells above and below the contour to the AOI dynamically after every episode. The piece AOIs are determined dynamically at every point in time where the tetromino changes position on the board be it via rotation or actual movement. The AOI contains the piece itself and all cells that surround it. This means, that differently shaped tetrominoes have different AOI sizes. While the T, L, J, S, Z and I shapes have AOIs of a size of 18 cells, the AOI containing the O shape only has a total of 16 cells. Due to its condensed form, this is, however, unavoidable and a negligible difference for the actual consideration of tetromino AOIs. While the whitespace around the board has not been deemed an AOI in and of itself, transitions from and to it are also counted for the transition value.

Mathematically, the transition value is determined by the relationship $\frac{\text{number of transitions}}{\text{seconds}}$. A low value means that there have not been a lot of transitions during this episode and that the player, hence, was focused on items with a longer dwell time. According to Goldberg and Kotval [1999] this then indicates an efficient and focused use of eye movements. When players show more transitions during an episode, the opposite is

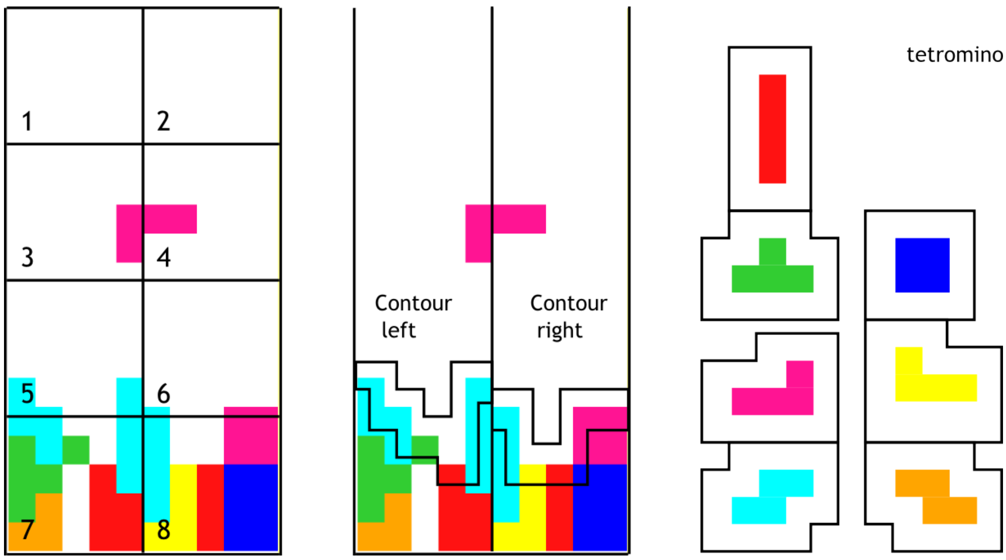


FIGURE 4.6: Areas of Interest in EMTRIS as considered for the calculation of the transition value

the case and the search and processing is less efficient and, hence, as an exhibition of uncertainty.

In order to determine the values of Table 4.2, the distribution of transition values for the pilot study have been taken and divided into fitting percentiles at 20%, 40%, 60% and 80% to determine delimiters. Since the nervousness of a player is most likely influenced by the current game state, which can be shaped by the choice of pieces, the transition value influences the block choosing algorithms.

Since EMTRIS also uses the game state analysis based adaption mechanisms NEMTRIS uses, Figure 4.7 shows, how the full algorithm adaption is calculated in EMTRIS. Each algorithm is represented by a number from one to five, where GRAB BAG is represented by the one and BUST HEAD by five. The values for both algorithm adaptions are calculated separately and then summed up and divided by two. If both come to the same conclusion, the value stays unchanged. However, if according to the bumpiness of the game it is supposed to be very hard, but according to the transition value, the player is coping inefficiently with the situation, and would make the game very easy, the harsh judgement of the bumpiness gets smoothed by the eye movement based adaption.

With TETRIS there are clear regions of interest for players to look at. These are the contour and the tetromino of the current episode. Jermann et al. [2010] showed that experts at TETRIS look at these regions more regularly than novice players, which follows comparable studies [e.g. Singer et al., 1996]. This means that a measurement derived from where people look tells us about their competence. In comparison, the transition values measure how players look at the game and, hence, measure their current emotional state.

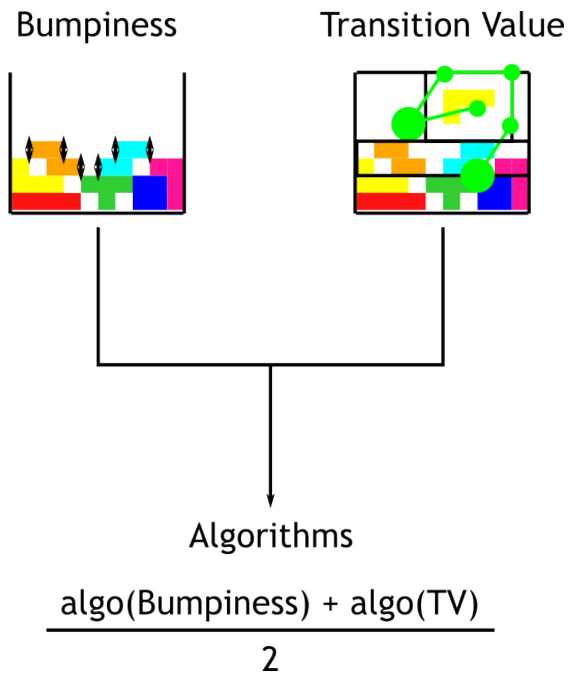


FIGURE 4.7: Adaption of Algorithms in EMTRIS - combination of game state adaption and eye movement based adaption

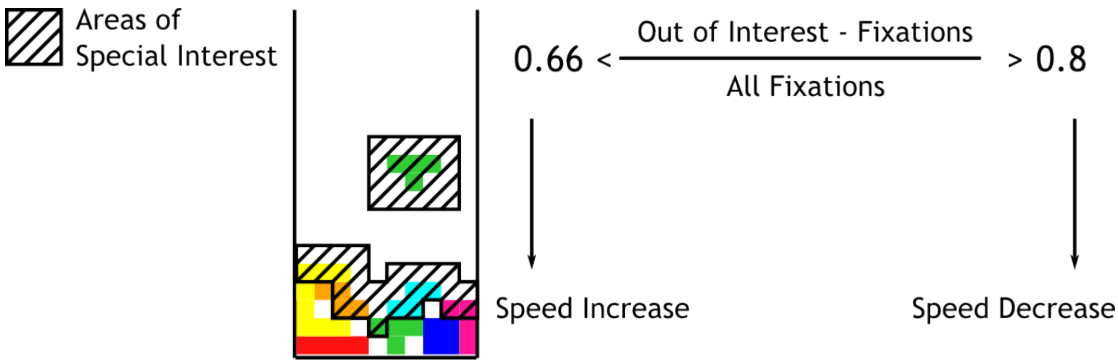


FIGURE 4.8: Adaption to Out-of-Interest Fixations - abstract illustration of concept

The Out-Of-Interest(OOI) Fixations ratio is calculated by $\frac{\text{Number of Out-Of-Interest-Fixations}}{\text{Number of All Fixations}}$. An OOI-Fixation for this calculation is every fixation that occurs in an area that does not belong to the AOIs of the contour or the current tetromino. The values for the thresholds as can be seen in Figure 4.8 have again been determined by the pilot study. Since experts perform a given task they are experts in faster than novices, the eye movement based speed adaption relies on the OOI-Fixations ratio. OOI-Fixations are especially useful, since as a tool for analysis, they give additional information about how meaningful and relevant a given fixation is. The speed adaption then relies on the actual value achieved for the OOI-Fixations ratio. This also means that in this way, adaption for decrease happens slightly faster than adaption for increase, because the relationship of OOI-Fixations ratio to the speed adaption is again one of reciprocity.

With regards to speed adaption, EMTRIS incorporates the pile height measure of NEMTRIS.

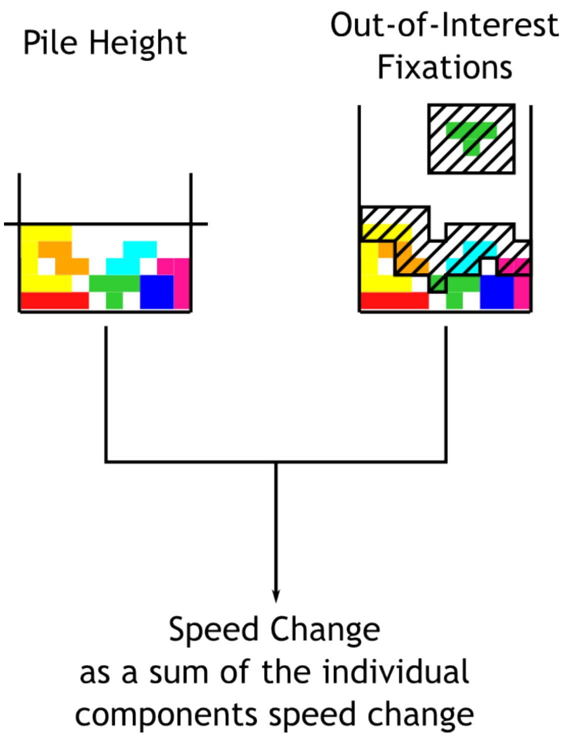


FIGURE 4.9: Adaption of Speed in EMTRIS - combination of game state adaption and eye movement based adaption

Speed is fully adapted by the sum of the individual components calculated by the pile height and OOI-Fixations ratio (see also Figure 4.9). This way, if the pile height suggests a fast game, but the OOI-Fixations ratio establishes that the player exhibits low expert behaviour in their fixations, the speed adjustment levels each other out. However, if both measures establish the same direction (decrease or increase), the values are amplified.

4.3.2.3 Data Recording

As well as the logs automatically created by the EYELINK II by SR RESEARCH that have been used for the study in this context, (N)EMTRIS also does its own data recording for general information about the game and its states next to eye movement events and calculations performed on them. Since the in-house framework would enable researchers to use different eye trackers according to their needs[5], the log files provided by each eye tracker might differ or even be non-existant.

(N)EMTRIS records every key input a player makes that results in an action together with that action. Whenever a fixation occurs, the data of the fixation is recorded

[5]Currently, the following options are provided: EYELINK II, THE EYE TRIBE, a self constructed version of the ITU-TRACKER [see for the concept Hansen et al., 2004] and a Mousetracker simulating eye movements. It is foreseeable that the software will be enhanced to also consider other and more modern eye trackers as they become available to the laboratory.

together with its AOI and whether the fixation is in an Area of Special Interest or OOI. Furthermore, the eye movement events of blinks, gazes in general and saccades are recorded.

At the start of a game, a new line is appended indicating a new game section and the participant identifier is recorded as well as the game type (NEMTRIS or EMTRIS). At the end a line is written to indicate the end of the game. Every time a player cleared a line, the number of lines cleared so far and the score resulting from that are recorded. Note that the score can differ, since it depends on the number of lines cleared at the same time (see for more details Section 6.4.1).

At the start of an episode of a tetromino, its shape is registered. Whenever an episode ends, the current game state is fully registered. This means, as well as pile height, the height difference between the previous pile height and the current pile height, the number of closed holes and the current sum for bumpiness, the logger also records the eye movement derived measures of OOI-Fixations ratio, the number of transitions and the transition value itself. In order to be able to reconstruct all steps of adaption, the speed adjustments of both NEMTRIS and EMTRIS and the resulting speed adjustment (according to version) are noted as well as the respective data for choice of algorithm. Finally, a current representation of the board is printed.

With this data recording mechanism at hand, all requirements according to the design of the user study (see Chapter 5) were fulfilled and the hypotheses (see Chapter 3) formulated within this research project can be investigated. Furthermore, the data recording enables researchers to develop their own questions along the data set produced by the study (see Chapter 6).

Chapter 5

Pre-Study, Pilot Study and Design of User Study

In order to investigate the research questions described in Chapter 3 and test the hypotheses derived from them, a user study has been conducted (see Section 5.4 for the resulting operational hypotheses). To prepare it, a pre-study investigating different TETRIS block choosing algorithms (5.1) and a pilot study (5.3) have been conducted as well. The design of the user study is covered in Section 5.2.

The reasoning for which test is used when can be found in Section 6.2.

5.1 Pre-Study

The pre-study has been conducted using the PYTRIS implementation (see 4.3.1) of TETRIS. Its main purpose was to investigate intrinsic differences between the five types of algorithms established in the corresponding section above (4.2).

5.1.1 Leading Questions

The pre-study was designed to find out whether different algorithms for choosing blocks in TETRIS have intrinsically different difficulties and whether they are also perceived by players as having different difficulties. Data recording was done in such a way that the conventional adaptation in NEMTRIS and EMTRIS can be built upon a data driven foundation.

As a hypothesis, there was the expectation that the algorithms would be ranked as follows from most to least fun, perceived easiest to hardest and best performance to worst:

1. NICETRIS → 2. GRAB BAG → 3. TRUE RANDOM
→ 4. SKEWED RANDOM → 5. BUST HEAD

5.1.2 Procedure

Test participants were recruited amongst peers on a voluntary basis. They were free to choose between a lab setting, a home setting and a lab setting on their own computers. The settings are, hence, not comparable. This was intentional, since test participants were encouraged to create a 'natural' gaming situation for them in order to counteract effects that might occur when gaming in laboratory settings (see for a discussion of this issue Ladouceur et al. [1991]).

After giving their consent to the use of data in oral form, participants first filled in a questionnaire asking for statistical data and self-rated TETRIS expertise. Then they started playing the ten TETRIS games as defined in PYTRIS (see Section 4.3.1). After every game, there was a short questionnaire asking how players rated fun and difficulty of the game they just played. The full questionnaire concept can be found in Appendix A.1. Pauses between games were self-paced; however, it was suggested that after five games test participants should take a longer pause, since playing for 50 minutes straight can be tiring.

The recorded data included the log data provided by PYTRIS and the completed questionnaires.

5.1.3 Results

In total, 16 test participants took part in the pre-study, resulting in 160 games played, 32 per algorithm. The gender of test persons was equally distributed along a binary type categorisation (male, female)[1]. Test participants' ages ranged from 22 to 34 years (mean: 26 years, median 25.5 years).

[1]As can be seen in Appendix A.1, the attribute of 'gender' has been enquired via a free form field. However, none of the test participants violated social boundaries. Everyone filled in either male or female according to external and internal attribution. See also for a general discussion amongst others Butler [2011].

Eleven of the test participants were educated or worked in Computer Science, five had other occupations. All of them knew of Tetris and had played it before. Along a self-estimated Likert scale ranging between [1..10], test participants rated their competence in playing Tetris on an average of 5.867 (range: [3..9], median: 6).

For the full range of descriptive statistics on the test participants, see Appendix B.1.1.

5.1.3.1 Performance Based Results

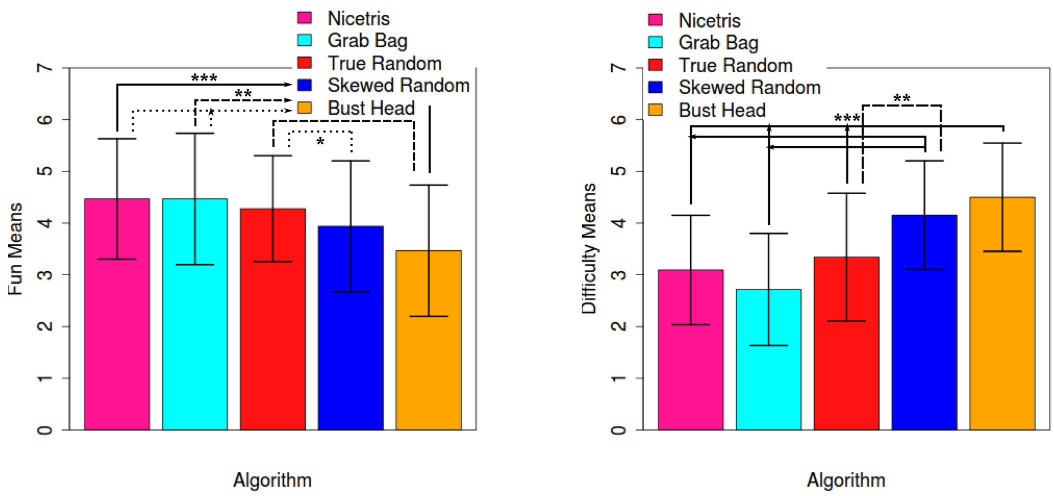
Table 5.1 shows how players performed in the pre-study according to the number of lines cleared during the game. This measure reflects the *intrinsic* difficulty of the algorithms, whereas the questionnaire data reflects the *perceived* difficulty and fun of the algorithms. The difference between Nicetris and Grab Bag appears to be negligible. Between True Random and Skewed Random there is a larger difference than between the other algorithms. The mean of lines cleared over all games played regardless of algorithm lies at 27.08 lines. The differences in performance are significant with a large effect of the algorithm on performance ($p < 0.001$, Spearman's $\rho = 0.524$).

Algorithm	Mean	Median	sd	Significance	Rank
Nicetris	34.19	32	13.585	$p < 0.001$	2
Grab Bag	34.22	32	14.524	$p < 0.001$	1
True Random	31.97	35	14.901	$p < 0.001$	3
Skewed Random	21.66	19.5	12.936	$p < 0.001$	4
Bust Head	13.38	10.5	9.366	$p < 0.001$	5

Table 5.1: Lines Cleared According to Algorithms in Pre-Study. Significance has been calculated using the Wilcoxon-Mann-Whitney test. Rank (according to mean) goes from most lines cleared(1) to least lines cleared(5).

Since speed increases with every line made, it seems natural that over all games as well as individually (see Table B.2), the measure of speed correlates negatively and strongly with the number of lines made ($p < 0.001$, Spearman's $\rho = -1.000$). This demonstrates the equivalence of the two measures.

With regards to duration of play, the expected ranking appeared (see Table B.3). The values for Nicetris (mean: 276.6s, median: 285s, sd: 33.46s) and Grab Bag (mean: 272.8s, median: 298.5s, sd: 45.14s) were again closer to each other than for other algorithms.



(A) Reporting on Fun

(B) Reporting on Difficulty

FIGURE 5.1: Perceived Difficulty and Fun for Different Types of Algorithms for Choosing Blocks in TETRIS as Reported in the Questionnaires. The level of significance is denoted as $* \to p < 0.05, ** \to p < 0.01, *** \to p < 0.001$. Significance has been tested using the Wilcoxon-Mann-Whitney test .

5.1.3.2 Questionnaire Results

In general, NICETRIS and GRAB BAG were ranked similarly by the test participants regarding the fun and difficulty. The differences between the other algorithms were stronger (cf. Appendix B.1.3). The questionnaire data follows the trend of the performance based results, where NICETRIS and GRAB BAG were virtually the same. Players also did not see a big subjective difference between them as well. As can be seen in Figure 5.1b, alternately to the initial hypothesis, NICETRIS has even been rated as being more difficult than GRAB BAG.

5.1.3.3 Other Results

The analysis of closed holes[2] (see also Table 5.2) shows a ranked difference between GRAB BAG, TRUE RANDOM, SKEWED RANDOM and BUST HEAD. According to this data, NICETRIS places between TRUE RANDOM and SKEWED RANDOM. This means, that it is the only algorithm which does not sort itself consistently into the algorithm order. The global median for closed holes over all games lies at 9.0. A Kruskal-Wallis analysis shows that the differences between the algorithms is significant ($\chi^2 = 81.02, df = 52, p < 0.01$). After a Bonferroni Correction leading to $\alpha = 0.002$, significant differences are individually only between GRAB BAG and BUST HEAD (Wilcoxon-Mann-Whitney: $U = 261, Z = 3.382, p < 0.001, r = 0.423$).

[2]For a definition of closed holes please consider Section 4.1.1.

Algorithm	Mean	Median	sd	Significance(p)	Corr.	Rank
NICETRIS	8.31	8.5	4.63	< 0.001	-0.42	3
GRAB BAG	7.00	6	3.96	< 0.001	-0.37	1
TRUE RANDOM	7.44	8	3.84	< 0.001	-0.26	2
SKEWED RANDOM	8.91	8	4.52	< 0.001	-0.25	4
BUST HEAD	10.50	11	3.77	< 0.001	-0.14	5
Over all	8.43	9	4.29	< 0.001	-0.44	

TABLE 5.2: Maxmimum Number of Closed Holes per Game According to Algorithms in Pre-Study. Significance has been done using the Wilcoxon-Mann-Whitney test. Correlation means the correlation of the maximum number of closed holes with lines made calculated with Spearman's ρ. Rank (according to mean) goes from lowest maximum (1) to highest maximum(5) of closed holes found.

Similarly, according to the pile height analysis (see Table 5.3), the algorithms are all distinct. The general order of the algorithms again follows previous findings. The differences between algorithms are significant according to a Kruskal-Wallis analysis ($\chi^2 = 114.33, df = 52, p < 0.001$). After a Bonferroni correction leading to $\alpha = 0.002$ significant differences can be found individually between NICETRIS and SKEWED RANDOM (this and following with Wilcoxon-Mann-Whitney: $U = 271, Z = 3.236, p < 0.001, r = 0.404$), NICETRIS and BUST HEAD ($U = 166, Z = 4.646, p < 0.001, r = 0.581$), GRAB BAG and SKEWED RANDOM ($U = 232.5, Z = 3.753, p < 0.001, r = 0.469$), GRAB BAG and BUST HEAD ($U = 127, Z = 5.170, p < 0.001, r = 0.646$), TRUE RANDOM and BUST HEAD ($U = 226.5, Z = 3.834, p < 0.001, r = 0.479$) as well as SKEWED RANDOM and BUST HEAD ($U = 280.5, Z = 3.108, p < 0.002, r = 0.389$).

Algorithm	Mean	Median	sd	Significance(p)	Corr.	Rank
NICETRIS	7.15	6.75	1.72	< 0.001	-0.32	2
GRAB BAG	6.69	6.57	1.81	< 0.001	-0.46	1
TRUE RANDOM	7.53	7.37	2.00	< 0.001	-0.56	3
SKEWED RANDOM	8.47	8.87	1.30	< 0.001	-0.48	4
BUST HEAD	9.45	9.81	1.58	< 0.001	-0.51	5
Over all	7.86	7.99	1.95	< 0.001	-0.66	

TABLE 5.3: Mean Pile Height per Game According to Algorithms in Pre-Study. Significance has been done using the Wilcoxon-Mann-Whitney test. Correlation means the correlation of the mean pile height with lines made calculated with Spearman's ρ. Rank (according to mean) goes from lowest mean (1) to highest mean(5) of mean pile heights.

Bumpiness (see Table 5.4) also shows a strong correlation with lines made, individually and over all algorithms. The differences between algorithms are significant with a Kruskal-Wallis analysis ($\chi^2 = 99.87, df = 52, p < 0.001$). After a Bonferroni correction leading to $\alpha = 0.002$ significant differences can be found individually between

Nicetris and Bust Head (this and following with Wilcoxon-Mann-Whitney: $U = 288, Z = 3.478, p < 0.001, r = 0.435$), Grab Bag and Skewed Random ($U = 203, Z = 4.149, p < 0.001, r = 0.519$) as well as Grab Bag and Bust Head ($U = 181, Z = 4.444, p < 0.001, r = 0.556$).

Algorithm	Mean	Median	sd	Significance(p)	Corr.	Rank
Nicetris	17.37	17.09	2.66	< 0.001	-0.34	2
Grab Bag	16.48	16.67	2.58	< 0.001	-0.50	1
True Random	17.66	17.59	2.99	< 0.001	-0.44	3
Skewed Random	19.27	19.38	2.43	< 0.001	-0.67	4
Bust Head	19.95	19.77	2.94	< 0.001	-0.39	5
Over all	18.14	17.93	2.98	< 0.001	-0.61	

Table 5.4: Mean Bumpiness According to Algorithms in Pre-Study. Significance has been done using the Wilcoxon-Mann-Whitney test. Correlation means the correlation of the time needed with lines made calculated with Spearman's ρ. Rank (according to mean) goes from lowest bumpiness (1) to highest bumpiness (5).

5.1.4 Consequences for Further Development

One of the two core results from the pre-study that influenced further development on (N)EMtris was that the algorithms differ significantly from each other. However, while the effective differences between Nicetris and Grab Bag are questionable, there appears to be a gap between True Random and Skewed Random. For the adaption along block choosing algorithms, the results are important for two different reasons. First, an in-between version like Mild Skewed Random as described in Section 4.2.4 above smooths the transition between True Random and Skewed Random during a live game. Second, since Nicetris does not produce as consistent results as the other algorithms, it should not be used as a block choosing algorithm.

The other core result from the pre-study is that the situation analysis based measurements can be used as indicators for performance. This means that the number of closed holes, bumpiness and pile height are all individually suited as parameters on which classical live-adaption without eye movements could be based. Of these three, bumpiness and pile height show the strongest correlation with lines made and also the most significant inner categorical differences. Since (N)EMtris alters its speed and algorithms for adaptation, pile height and bumpiness, which both show strong effects, were considered for adaptation.

5.2 Study Setup

Since the study setup was almost the same for the pilot study and the main study, the general setup for both is presented together.

5.2.1 Test Setup

In order to be able to see and test for a framing effect, test subjects have to be primed differently. Hence, there are four groups into which the test subjects were sorted (see Table 5.5). The four groups are given identifying group names to make later analysis easier.

	eye movement based adaption	conventional adaption
player framed	FE	FC
player not framed	NE	NC

TABLE 5.5: Classes of Test Subjects. The test is sorted in a mixed setting between-subjects (rows) and within-subjects (columns). F/N distinguishes framing and no-framing, E/C distinguishes eye-movement based adaption and conventional adaption

The test participants were divided into two groups (between-subjects) according to how they were being framed. Both groups then played games that which were adapted to eye movements and such which have only been adapted conventionally (within-subjects). Players in neither framing group were informed that two versions exist beforehand and were not aware during play, which version of (N)EMTRIS they were currently playing.

Due to the variance of the population from which the test subjects were drawn, the results of the studies will also be tested for effects of *gender*, *age*, TETRIS *expertise*, *presumed* TETRIS *expertise* as reported by the test participants and their *spatial abilities* to limit the potential of reporting false effects.

5.2.2 Study Procedure

Every test person followed the same procedure. Before anything happens, test participants signed a consent form (see Appendix A.2). As a first step, they filled in an initial questionnaire recording descriptive statistical data (see Appendix A.3). Then they performed a TETRIS skill test (see Section 5.2.3) using the respective version of PYTRIS (see Section 4.3.1.3). This was followed by the framing of the players as an introduction to (N)EMTRIS (see Appendix A.4).

The core part of the test consisted of players playing four games at random (two of each version of (N)EMTRIS) with an optional pause in-between. Every test participants was being tracked during this part. The In-Game GEQ was used as individual game experience measure for each game.

The full version of the GEQ (with the exception of questions regarding narrative) was filled in when players finished their four games. This was followed by a pause of at least five minutes, where players were offered a snack. In order to enable them to reflect their post game experience and report on it, this pause is required. The In-Game GEQ, the GEQ and the Post-Game Questionnaire are all taken from IJsselsteijn et al. [2013], albeit with removal of the question "I was interested in the game's story", because TETRIS has no story to speak of. The questions hinting to a semi-realistic environment and enquiring about the quality of the graphical presentations were left in.

As a last formal step, participants were asked to do the revised Vandenberg & Kuse Mental Rotations Tests in the MRT-A form [Peters et al., 1995] to assess their spatial abilities. The full test setup and its purpose are revealed in the end.

The setup was fully available in English and German (see Appendix C) due to the nature of the pool of test participants (see Section 6.3). To ensure that everything was properly prepared and conducted for every test participants, a data control sheet (see Appendix A.5) was used.

5.2.3 Determining Expertise

In order to establish the expertise of a test participant according to a performance measure (see also Section 2.1.1.1) next to a self-reported one, test participants were initially asked to play the skill determining version of PYTRIS (see Section 4.3.1.3) three times in a row. Their gameplay was then analysed according to their performance; the lines made in that time. If performance increases from the first to the second game, the delta is analysed (how many more lines have been made). The performance increase is quantified as a relationship of $\frac{\text{lines made/time} - \text{lines made previously/time}}{\text{lines made previously/previously used time}}$. Time normalisation is needed to account for players who lose within the given time frame.

The potential performance increase is again measured from second to third game. If the performance difference in the third game was more than 25%, the test participant was asked to play 2 more games. It is assumed that by this point they have reached the apex of their skill level for the test day.

5.3 Pilot Study

The setup for the pilot study was slightly different. Test participants were informed beforehand that there were two different versions of (N)EMTRIS that they would encounter during the test sessions. They also only filled in the questionnaire asking for statistical data. The mental rotation test was not part of the pilot study.

5.3.1 Purpose and Leading Questions

The purpose of the pilot study was playtesting the implementation of (N)EMTRIS. Since it cannot be expected that developers can test all edge cases of a given software, having this extra test helped to find unintended behaviour of the software, that would stay undetected otherwise and might influence the results of the main study in unwanted ways.

A second purpose of the pilot study was to confirm the suitability of the initial skill test that determines a player's expertise. Having a larger audience exposed to the skill test version of PYTRIS shows whether the approach works and whether the speed adjustment works for differently skilled participants.

Furthermore, it was also important to acquire eye movement data for Out-of-Interest fixations and transition values during a game of TETRIS to ensure the viability of the eye movement adaptation. In the pilot study setup those values were set to theoretically driven ranges between [0.0..1.0].

Lastly, it was also important that the two versions of (N)EMTRIS are indistinguishable to players on a conscious level. If their perceived difficulty is different, players would know that they encountered two types of games, although they only have been framed for one. The question here was, whether players could determine which game uses their eye movement data and which one does not.

5.3.2 Results

There were six test participants who volunteered for the pilot study. They were between 18 and 27 years old (mean: 23.8 years, median: 24.5 years), five of them had a background in computer science and two of them identified as female. Four of them were shortsighted, however, eye tracking was possible for all participants. All of them played games fairly regularly and knew TETRIS. They rated their confidence between 4 and 9 on a Likert scale ranging in [1..10] (mean: 6, median 5.5). More detailed information about the individual participants can be found in Appendix B.2.1.

During the first tryouts, several bugs within the implementation of (N)EMTRIS were detected, localized and could subsequently be fixed for the main study. Since these bugs included an issue with how the random number generator for the choice of blocks worked, the pilot study was especially important, since it showed that there could be games that mainly consisted of ß blocks, the worst of them only having three other blocks, two of those being t-tetrominoes. Over several players, it also showed, that ß blocks were generally more likely than with a plain TETRIS using the GRAB BAG algorithm; however, that was mathematically to be expected.

As with the skill tests, it became obvious, that all participants did not show a difference in performance compared to the last game of more than an absolute of 17.6% according to lines made over time during the game as defined in Section 5.2.3. Table B.8 depicts the individual performances in more detail. A closer analysis reveals that the standard deviation for the best and average performances are lower, which means that the analysis of these results in accordance to the actual performance of a player might be more stable.

	OOI-ratio	transition value
shapiro	$W = 0.928 \| p < 0.001$	$W = 0.077 \| p = 0.0$
mean	0.689	0.0177
median	0.75	0.008
20% percentile		0.004
33% percentile	0.667	
40% percentile		0.007
60% percentile		0.010
66% percentile	0.8	
80% percentile		0.021

TABLE 5.6: Analysis of eye movements relevant to eye movement interaction in EMTRIS. Values taken from pilot study. Note that the concept for EMTRIS as described in Section 4.3.2.2 requires different percentile splits for the two ratios. The line named 'shapiro' indicates a test on whether the data is normally distributed.

As can be seen in Table 5.6, neither the Out-of-Interest (OOI) ratio nor the transition value were normally distributed over all test participants. In order to account for this, not a deviation from the mean according to a normal distribution but rather fitting percentiles have been chosen as meaningful boundaries for the eye movement based adaption of EMTRIS.

When test participants tried to guess which version of (N)EMTRIS they were currently playing, they had a hard time deciding and only did so after several prompts. One participant only made a statement for two of the four games. Players were unable to guess which game they were playing, since none of the results were significantly different from a random distribution (NEMTRIS: $p = 0.346$, EMTRIS: $p = 0.072$ – using Wilcoxon-Mann-Whitney).

5.3.3 Consequences for Main Study

Following the results of the pilot study, some adjustments and further developments were made for the (N)EMTRIS prototype.

A hard ß block has been implemented. Following the realisation that ß pieces come more often on average in (N)EMTRIS no matter what, the algorithm was changed so that only a maximum of five ß pieces can come in a row. Otherwise, a ban on ß pieces comes into place: For the next x pieces, ß tetrominoes are excluded from the choice of tetrominoes. Here, x depends on the algorithm that is currently active. It is set to five for GRAB BAG down to one for BUST HEAD to retain the model of difficulty for the algorithms.

The skill test showed that it was suitable for leveling players to their daily performance level and also to establish an objective measure about player expertise in TETRIS. Thus it was deemed suitable for the main study without any changes.

The delimiters for eye movement based adaption according to the transition value and the Out-of-Interest fixation ratio have been adjusted following the results of the pilot study. Since the data was not normally distributed, percentiles have been used as meaningful boundaries for adaptation.

Since the two versions of (N)EMTRIS were indistinguishable for the test participants, they were deemed equally difficult and, hence, suitable for the user study.

5.4 Operational Hypotheses

With a test layout that has been designed with the research hypotheses of Chapter 3 in mind, operationalised hypotheses arise from these. They are still partly abstract and usable for other research with similar questions, but more concrete in terms of what an indicator for a parameter is (such as score as a measure of performance).

5.4.1 Performance Based Hypotheses

H-P1$_{\mathrm{OH}}$: Framed players have a higher final score than non framed players.

H-P2$_{\mathrm{OH}}$: Players of eye movement based adapted games have a higher final score than those playing a conventionally adapted game.

H-P3$_{\mathrm{OH}}$: Framed players have a higher final score even if playing a conventionally adapted game.

H-P4$_{\mathrm{OH}}$: Framed players have a higher final score in games with eye movement based adaptation than in games with conventional adaptation.

H-P5$_{\mathrm{OH}}$: The difference between final scores according to type of adaptation of a game of framed players is smaller than the difference of final scores of non-framed players when playing a game with eye movement based adaptation compared to a conventionally adapted game.

5.4.2 Gameplay Experience Based Hypotheses

H-U1$_{\mathrm{OH}}$: Framed players report a better gameplay experience in questionnaires than non-framed players.

H-U2$_{\mathrm{OH}}$: Players of a game adapted on eye movements report a better gameplay experience in questionnaires than those playing a conventionally adapted game.

H-U3$_{\mathrm{OH}}$: Framed players report a better gameplay experience in questionnaires than non-framed players, even if they play a conventionally adapted game.

H-U4$_{\mathrm{OH}}$: Framed players report on a better gameplay experience in questionnaires when playing games with eye movement based adaptation than in games with conventional adaptation.

H-U5$_{\mathrm{OH}}$: The difference between reported values on gameplay experience with regards to a game's adaptation type of framed players is smaller than the difference of reported values on gameplay experience of non-framed players when playing a game with eye movement based adaptation compared to a conventionally adapted game.

5.4.3 Expertise Based Hypotheses

H-E1$_{\mathrm{OH}}$: Framed players show more expert behaviour in their eye movements and their interaction with the game than non framed players.

H-E2$_{\mathrm{OH}}$: Players of games with eye movement based adaptation show more expert behaviour in their eye movements and their interaction with the game than those of conventionally adapted games.

H-E3$_{\mathrm{OH}}$: Framed players show more expert behaviour in their eye movements and their interaction with the game than non-framed players, even if they play a conventionally adapted game.

H-E4$_{\mathrm{OH}}$: Framed players show more expert behaviour in their eye movements and their interaction with the game in games with eye movement based adaptation than in games with conventional adaptation.

H-E5$_{\mathrm{OH}}$: The difference in expert behaviour of framed players that can be extracted from their eye movements and their interaction with the game will be larger than the one of non-framed players when playing a game with eye movement based adaptation compared to a conventionally adapted game.

5.4.4 Intersectional Hypotheses

H-I1$_{\mathrm{OH}}$: Players who receive a high final score also report positively on the game experience questionnaire.

H-I2$_{\mathrm{OH}}$: The higher the final score of a player, the more expert behaviour can be detected in their eye movements and their interaction with the game.

H-I3$_{\mathrm{OH}}$: The more expert behaviour which can be detected via a player's eye movements and their interaction with the game, the better their self-reported gameplay experience.

Chapter 6

Results of User Study

After a short presentation of the technical setup and the final procedure of acquiring these results (Section 6.1), a reasoning for the use of statistical tests analysing the results is provided in Section 6.2. General descriptive statistics about the test participants are shown in Section 6.3 followed by the results (Section 6.4), which are discussed along statistical hypotheses derived from the operational ones in Section 5.4.

6.1 Procedure

The study was conducted over two weeks. Test participants were recruited amongst peers via email notices on a voluntary basis. They received cake as a thank you for participating in the study. All of the test participants came into a lab within the university in which the full test session was conducted in one appointment such as described in Section 5.2.2. The EyeLink II by SR Research was used for eye movement recording.

Even though the issues with gaming in laboratory settings [Ladouceur et al., 1991] have been raised previously, the lab setting was chosen to ensure similar light conditions to reduce noise recording of the eye tracker. The study room had no windows which made this task easier. Recording was done at a frequency of 250 Hz. Since The Eye Tribe, the only suitable alternative to the EyeLink II available at the time, with its maximum of 60 Hz does not offer a high enough temporal resolution for this research, the potential disruptive effect of a head-mounted eye tracker was unfortunate, but not avoidable.

6.2 Interpretation of Data

In order to not appear arbitrary in the choice of tests that are used on reporting results throughout this thesis, this section shows why certain tests are used and under what circumstances alternatives are chosen.

6.2.1 Test on Normal Distribution

In general, tests that check whether the data is normally distributed have been done using the Shapiro-Wilk test. This test appears to be especially suitable for the analysis of data with $n < 50$ [Field et al., 2012, cf.].

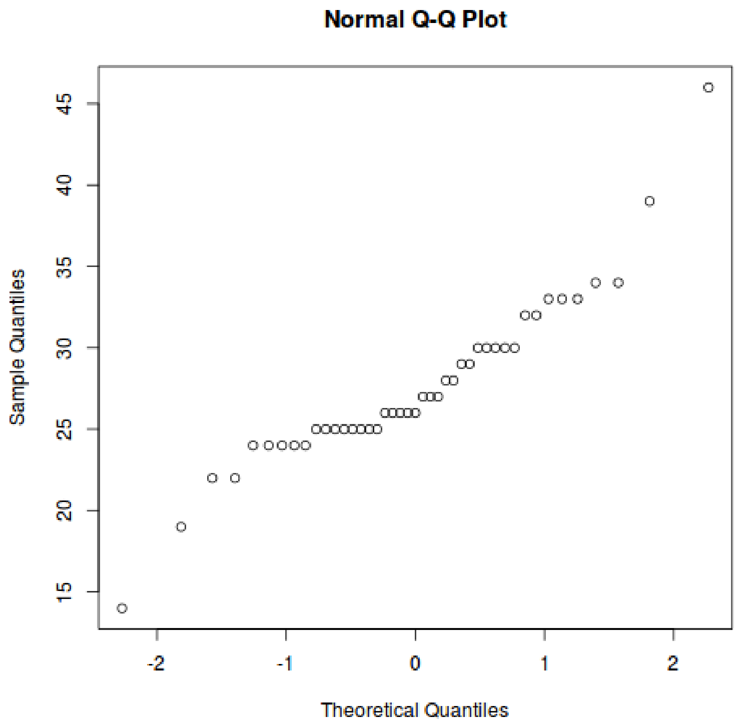


FIGURE 6.1: Q-Q Plot of Age Variable Indicating No Normal Distribution

Whenever the mathematical tests on normal distribution indicated edge cases, Q-Q plots such as in Figure 6.1 investigating the norm have been used to verify results with a visual inspection of the data.

6.2.2 Tests on Significance of Difference

Tests on significance of the difference of two or more samples are chosen depending on whether data was normally distributed or not. Student's t-test for two samples or the ANOVA for more than two samples both assume that the data is normally distributed [Field et al., 2012, cf.]. They only have been used, when previous tests indicated, that the data in fact is normally distributed.

Their equivalent for not normally distributed data are the non parametric tests by Wilcoxon-Mann-Whitney (also called U-Test) for two sample sizes and the Kruskal-Wallis analysis for more groups of analysis.

Whenever an ANOVA or a Kruskal-Wallis analysis yielded a significant difference somewhere in the analysed data, a pairwise comparison with either a t-test or a U-test (depending on normal distribution) with a Bonferroni correction (cf. Abdi [2007]) adjusting the α level for which the p value has been reported as actually significant has been applied.

6.2.3 Determination of Effect Size

The determination of effect sizes also depends on whether the data is normally distributed or not. It also depends on the type of data. The tests how they have been used here are shown in Table 6.1 and ties back to Field et al. [2012].

	Data Linear	Data Categorial
Normal Distribution	Pearson's r	Cohen's d
No Normal Distribution	Spearman's ρ	Wilcox Z

TABLE 6.1: Determination of Effect Sizes for Different Types of Data and Different Distributions of Data

An alternative for linear data with no normal distribution is the non parametric test of Kendall's τ. According to Colwell and Gillett [1982], however, Spearman's ρ is less sensitive to the order of the given data, which is why it has been chosen here. It usually reports on higher values though and should be interpreted carefully [Hauke and Kossowski, 2011, cf.]. Effect sizes are deemed small whenever they cross a threshold of $r >= 0.1$ or $d >= 0.2$, medium when $r >= 0.3$ or $d >= 0.5$, large, when $r >= 0.5$ or $d >= 0.8$ and very large when $r >= 0.8$ or $d >= 1.3$ [Ellis, 2010]. Effect sizes calculated with ρ are interpreted according to r values; although a visual inspection is used before reporting on effects too confidently.

6.3 Population

The user study was conducted with 43 participants in total. Of them, 21 were not informed about the use of eye movements and 22 were. The language of these tests was mostly German (83.72%), but seven tests were conducted in English. The age of test participants ranged from 14 to 46 (mean: 27.63, median: 26). With two participants not disclosing their gender identity, 17 of the test participants identified as female and 24 of them identified as male. Most of them worked or studied in the field of computer science (27), however four participants came from social studies, another four from humanities and eight came from other disciplines (such as design or architecture) (see for more detail Table B.9).

All of the test participants were familiar with digital games, albeit with different levels of competence (ranging in $[0, .., 10]$ with mean: 4.94 and median: 4) and regularity in interacting with them (17 less than once a month). Furthermore, they all knew of TETRIS, albeit two of them only as a pop-cultural reference; they had never played the game themselves. The actual TETRIS expertise measured as $\frac{\text{lines made}}{\text{minutes}}$ ranged in $[1.45..19.07]$ (mean: 9.13, median: 8.39). The values for self-reported values on TETRIS expertise range in $[0, .., 10]$ (mean: 5.51, median: 5) (see for more detail Table B.10).

Addendum to competence:
While the data of the self reported TETRIS is not normally distributed according to the Shapiro-Wilk test ($W = 0.93, p < 0.05$), a Q-Q plot on a normal distribution shows a fairly linear tendency. Since the actual skills measured as $\frac{\text{lines made}}{\text{minutes}}$ is normally distributed (Shapiro-Wilk $W = 0.98, p = 0.48$), a t-test on the normalised reveals that they are significantly different ($t = 2.40, df = 83.78, p < 0.05$) and do correlate negatively with a small effect size ($r = -0.22$). This shows how the self-reported measure is less meaningful in this context than the objectively measured TETRIS skills and also, that both measurements do not measure the same thing. This follows the effects reported by Kruger and Dunning [1999] (see also Figure 6.2).

The test participants also exhibited different levels of competence with regards to their mental rotation skills. The results of the redrawn Vandenberg-Kuse test ranged in $[3, .., 22]$ (mean: 12.19, median: 12) (individual values can be found in Table B.9), a maximum score of 24 was technically possible to achieve. Test participants were given three minutes to solve twelve items of the test. A point was only awarded, if both rotation alternatives were identified correctly.

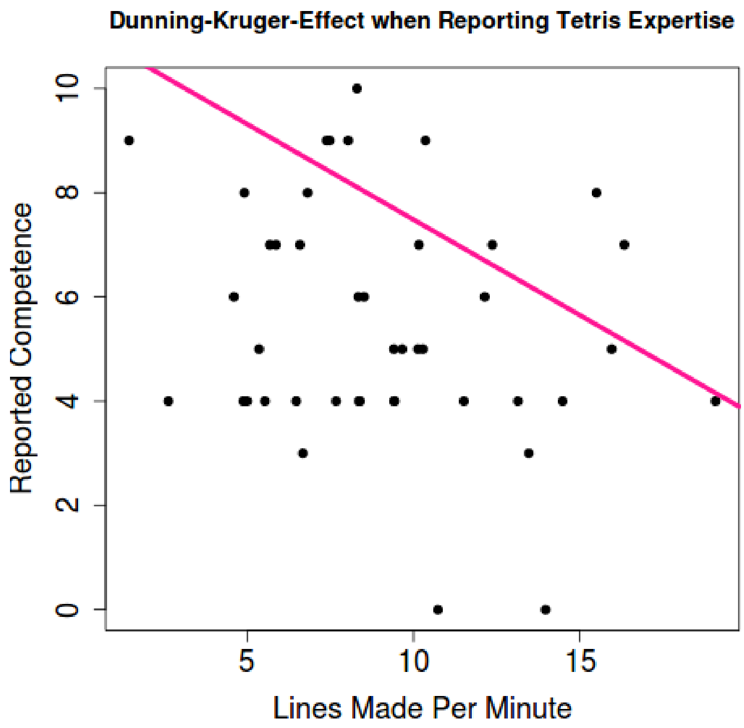


FIGURE 6.2: Reestablishment on the Dunning-Kruger Effect on Self-Assessed TETRIS Expertise vs. Play Performance in Skill Test

6.4 Results According to Hypotheses

The research hypotheses as described in Chapter 3 and operationalised in Section 5.4 have been reformed into statistical hypotheses including formulated null-hypotheses in order to later be able to provide a meaningful interpretation of the results of the user study. Results are, hence, first presented according to the established hypotheses.

In the analysis, N refers to non-framed participants and F refers to framed participants, C refers to conventionally adapted games and E refers to games with eye movement based adaptation. When numbers for any categories are given, they refer to the means for this category.

6.4.1 Performance Based Hypotheses

The score in these hypotheses refers to a calculation that considers how many lines are made at the same time. This score increases exponentially. If a player clears one line, they receive one point; if they clear two, they receive three points; if they clear three, they receive six points and if they clear four lines, they receive ten points.

6.4.1.1 Framed vs. Non-Framed Players

H-P1$_0$: There are no significant differences in performance measures between framed players.

H-P1$_{SH}$: Framed players finish a game with more cleared lines than non framed players (1). This coincides with a higher score (2), a higher mean for speed at the end (3) and on average a higher class of algorithms when the game ends (4).

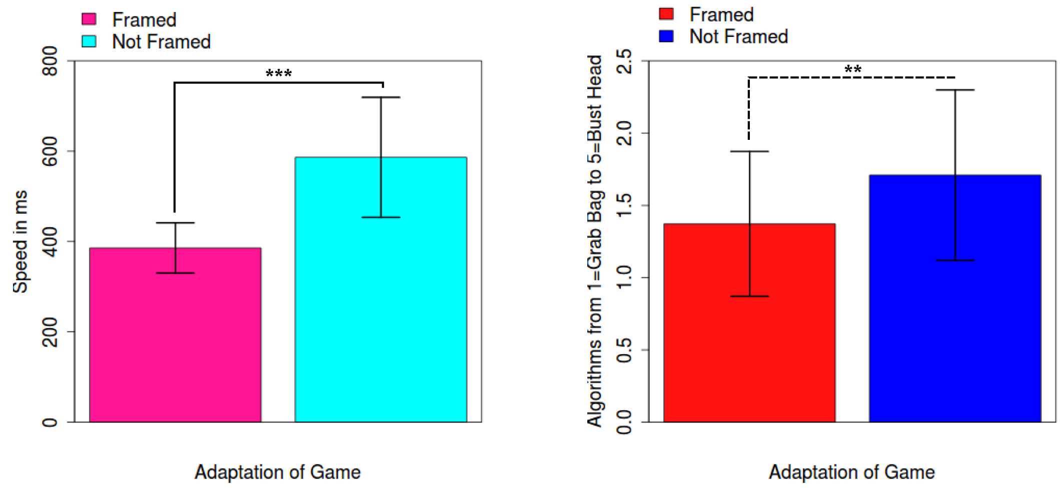
While the two categories produce differences in the number of cleared lines, this difference is not statistically significant ($p = 0.389$) (see Table B.11). However, the means of the non-framed players are actually a bit higher than those of framed players ($N = 14.36$, $F = 12.88$). The scores follow this trend ($N = 18.11$, $F = 16.57$, $p = 0.520$, see Table B.12) as does the measure of speed at the end ($N = 473.1ms$, $F = 498.5ms$, $p = 0.354$, see Table B.13) and algorithms ($N = 1.57$, $F = 1.51$, $p = 0.617$, see Table B.14). This leads to a rejection of H-P1 and H-P1$_0$ has to be favoured.

6.4.1.2 Type of Adaptation

H-P2$_0$: There are no significant differences in performance measures between eye movement based adaptation and conventional adaptation.

H-P2$_{SH}$: The mean of the number of lines cleared of eye movement based adapted games is higher than that of conventionally adapted games (1). This coincides with a higher score (2), a higher mean for speed at the end (3) and on average a higher class of algorithms when the game ends (4).

While there is no significant difference between the types of adaption concerning lines made ($C = 13.66, E = 13.65, p = 0.86$, see Table B.15 or scores ($C = 17.69, E = 16.95, p = 0.94$, see Table B.16), in both cases the values for games with eye movement based adaptation were closer to a normal distribution. For speed, eye movement based games showed a normal distribution, while conventionally adapted games did not (see Table B.17). However, conventionally adapted games ended significantly faster than games with eye movement based adaptation with a very strong effect ($C = 385.7ms, E = 586.5ms, p < 0.001, d = 1.971$, see also Table B.17 and Figure 6.3a). This is the exact opposite of H-P2$_{SH}$(3). For algorithms, the hypothesis holds with a medium effect ($C = 1.37, E = 1.71, p < 0.01, d = 0.622$, see also Table B.18 and Figure 6.3b). This means, H-P2$_{SH}$ is only likely (4). However, H-P2$_0$ has to be rejected as well. For point (3), an inverse hypothesis is likely.



(A) Speed Differences (B) Algorithm Differences

FIGURE 6.3: Effect of Type of Adaptation on Speed and Algorithms. Significance is denoted as $** \rightarrow p < 0.01, *** \rightarrow p < 0.001$.

6.4.1.3 Effect of Framing on Conventionally Adapted Games

H-P3$_0$: There is no significant difference according to framing in conventionally adapted games.

H-P3$_{SH}$: The mean of the number of lines cleared of framed players is higher than that of non-framed players even if they play a conventionally adapted game (1). This coincides with a higher score (2), a higher mean for speed at the end (3) and on average a higher class of algorithms when the game ends (4).

Non framed players actually performed better measured in lines made and scores than framed players on average when playing conventionally adapted games (lines made: $N = 15.88, F = 11.55$, scores: $N = 20.43, F = 15.07$), however, the differences were not significant (lines made: $p = 0.132$, scores: $p = 0.229$). As can be seen in Tables B.19 and B.20, the performance of framed players was closer distributed to a normal distribution than the performance of non-framed players. For speed ($N = 396.7ms, F = 375.3ms, p = 0.081$, Table B.21) and algorithms ($N = 1.43, F = 1.32, p = 0.798$, Table B.22) the results are similar. Hence the null hypothesis for H-P3 has to be accepted.

Although not being part of any hypothesis, the comparison has been done for games with eye movement based adaptation as well. However, the results are reverse there with framed players performing better than non-framed players, albeit again, without significant differences. The values for framed players for lines made and scores are normally distributed whereas those for non-framed players are not (see also Table B.23).

6.4.1.4 Effect of Type of Game on Framed Players

H-P4$_0$: There is no significant difference according to the type of game for framed players.

H-P4$_{SH}$: The mean of the number of lines cleared of framed players is higher in games with eye movement based adaptation compared to conventionally adapted games (1). This coincides with a higher score (2), a higher mean for speed at the end (3) and on average a higher class of algorithms when the game ends (4).

By inverting the point of view of H-P3 and seeing how the type of game affects framed and – even though not covered by the hypothesis – non-framed players, it shows, that the speed at the end is significantly different for non-framed ($C = 373ms, E = 564ms, W = 70, p < 0.001, Z = 3.79, r = 0.09$) and framed players ($C = 369ms, E = 589ms, W = 4, p < 0.001, r = 0.13$) alike. For framed players even the algorithm at the end of the game is significantly different ($C = 1.32, E = 1.71, W = 159, p = 0.041$), however, after a Bonferroni correction, the significance level has to be lowered to $\alpha = 0.025$ in order to be a reliable measure. With a $p = 0.040$, the difference between algorithms cannot be reported as significant anymore. Lines made and scores were not significantly for framed players (lines: $C = 11.55, E = 14.2, W = 205.5, p = 0.398$, scores: $C = 15.07, E = 18.07, W = 209, p = 0.445$) or non-framed players (lines made: $C = 15.88, E = 13.07, W = 255, p = 0.392$, scores: $C = 20.43, E = 15.79, W = 254.4, p = 0.399$). Following this, H-P4$_{SH}$(3) can be accepted albeit inversely. However, for all other aspects, H-P4$_0$ is more likely.

6.4.1.5 Framing and Player Improvement

H-P5$_0$: There is no significant difference between the differences in performance of types of game according to framing category.

H-P5$_{SH}$: The difference between the means of the number of lines cleared according to type of adaptation of a game of framed players is smaller than the difference for non-framed players when playing a game with eye movement based adaptation (1). Same accounts for the difference in the scores (2), the means for speed (3) and class of algorithm when the game ends (4).

The differences in performance are as expected for lines made ($F = 6.84, N = 10.43, p = 0.593$, see also Table B.24) and scores ($F = 9.00, N = 13.93, p = 0.527$, see also Table B.25), albeit not significantly different according to framing categories. An inverse trend occurs with speed ($F = 251.5, N = 164.1, p < 0.05$, see also Table B.26), but the significance does not hold after a Bonferroni correction. Furthermore, the differences for algorithms are not significant ($F = 0.61, N = 0.57, p = 0.940$, see also Table B.27). Hence, H-P5_0 is more likely to be true than H-P5_{SH}.

6.4.1.6 Summary

Table 6.2 shows all performance based hypotheses and what the data reveals about their likelihood. All in all, algorithms are only different for different types of games and speed is significant in the opposite direction than expected according to type of adaptation and especially again for framed players.

Hypothesis	Result
H-P1_{SH}	✗
H-P2_{SH}	(4) and inverse (3)
H-P3_{SH}	✗
H-P4_{SH}	inverse(3)
H-P5_{SH}	✗

Table 6.2: Summary of Performance Based Hypotheses

6.4.2 Gameplay Experience Based Hypotheses

Gameplay experience was measured using the Game Experience Questionnaire (GEQ) developed mainly by [IJsselsteijn et al., 2013]. When a hypothesis refers to desired experiences, this means high values in the aspects of *Flow*, *Challenge*, *Competence*, *Immersion* and *Positive Affect* combined with low values in the aspects of *Tension/Annoyance* and *Negative Affect* in the In-Game as well as Core modules of the GEQ and high values in the aspect of *Positive Experience* combined with low values in the aspect of *Negative Experience* in the Post-Game module of the GEQ. The higher the difference between values for positive and negative aspects, the more desired the experience. The aspects of *Tiredness* and *Returning to Reality* in the Post-Game module of the GEQ are not covered by the following hypotheses, but recorded regardless for future analysis.

6.4.2.1 Effect of Framing

H-U1$_0$: The differences in reported GEQ values are not significant according to framing.

H-U1$_{\mathrm{SH}}$: Framed players report a more desired gameplay experience than non-framed players.

For the results reported here, the Core element of the GEQ as well as its Post-Game module are relevant. However, as Table B.28 shows, none of the reported values were significant. The only value coming close is *challenge* ($F = 2.6, N = 2.22, p = 0.074$) so there might be an inverse trend. For now H-U1$_0$ has to be accepted.

6.4.2.2 Effect of Type of Adaptation

H-U2$_0$: There are no significant differences in reported GEQ values according to type of adaptation.

H-U2$_{\mathrm{SH}}$: Players of a game adapted on eye movements report a more desired gameplay experience than those playing a conventionally adapted game.

For the results reported here, the In-Game module of the GEQ was relevant as an inter categorical measure between individual games. Table B.29 shows that none of the reported were significant. This indicates no significant effect of the type of adaption on the gameplay experience of players. H-U2$_0$ is more likely to be accepted.

6.4.2.3 Effect of Framing on Conventionally Adapted Games

H-U3$_0$: There are no significant effects for conventionally adapted games between framed and non-framed players.

H-U3$_{\mathrm{SH}}$: Framed players report a more desired gameplay experience than non-framed players, even if they play a conventionally adapted game.

According to Table B.30 there is only one significant difference in the reported GEQ-values between framed players and non-framed players playing conventionally adapted games and that is the one for challenge with a medium to strong effect ($F = 3.38, N = 2.77, p < 0.05, d = 0.718$, see also Figure 6.4). This means, framed players experienced a higher challenge in their game than non framed players even if playing a conventionally adapted game. For games with eye movement based adaptation, framing had no significant influence ($F = 3.17, N = 2.77, p = 0.151$). All in all, H-U3$_{\mathrm{SH}}$ is likely only for the value of challenge.

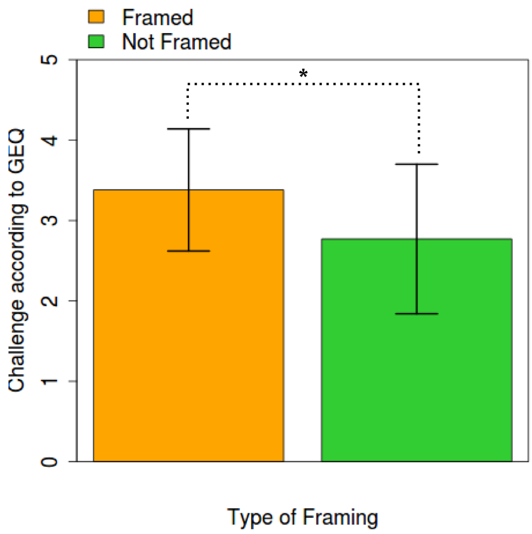


FIGURE 6.4: Challenge for Framed Players in Conventionally Adapted Games. Significance is denoted as $* \rightarrow p < 0.05$.

6.4.2.4 Effect of Type of Game on Framed Players

H-U4$_0$: There is no significant difference in reported gameplay experience for framed players according to type of adaptation.

H-U4$_{SH}$: Framed players report a more desired gameplay experience when playing games with eye movement based adaptation than in games with conventional adaptation.

While the type of adaptation doesn't influence framed players in most of the categories inquired by the GEQ In-Game module (see Table B.31), their perceived competence does change significantly with a medium effect ($C = 2.06, E = 2.55, p < 0.05, d = 0.669$). When they play a game with eye movement based adaptation, they report having more competence than in conventionally adapted games. Hence, H-U4$_{SH}$ can be tentatively accepted, but only for the aspect of competence.

6.4.2.5 Framing and Difference in Reported Experience

H-U5$_0$: There is no significant difference between the deltas of values of gameplay experience for type of adaptation of framed and non-framed players.

H-U5$_{SH}$: The difference between reported positive values on gameplay experience according to type of adaptation of a game of framed players is smaller than the difference of reported positive values on gameplay experience of non-framed players when playing a game with eye movement based adaptation.

Table B.32 shows that there are no significant differences between the deltas of reported values for gameplay experience on any aspect the GEQ in-game module tests for. This means, H-U5_0 has to be accepted.

6.4.2.6 Summary

Table 6.3 presents a summary of gameplay experienced hypotheses. Only in very distinct circumstances, framing or type of adaptation partly influence this experience.

Hypothesis	Result
H-U1_{SH}	✗
H-U2_{SH}	✗
H-U3_{SH}	for challenge
H-U4_{SH}	for competence
H-U5_{SH}	✗

TABLE 6.3: Summary of Gameplay Experience Based Hypotheses

6.4.3 Expertise Based Hypotheses

When a hypothesis refers to a distinct fixation, it refers to all fixations which are not of the type OOI-Fixation as defined in Section 4.3.2.2.

Epistemic actions are defined after Kirsh and Maglio [1992] as amount of rotations over the bare minimum to see the tetromino in every possible position and as amount of superfluous translations from and to the border. While Lindstedt and Gray [2013] pointed out, that such actions could also mean a change of strategy, this issue has been ignored, since it is very likely that the change in strategy is related to the translations in that it has been discovered by it. Hence, the change of strategy then results from an epistemic action.

Due to the fact that data loss occurred in the log files implemented by the internal framework, the log files supplied by the EYELINK II have been used for eye movement analysis instead.

6.4.3.1 Effect of Framing

H-E1$_0$: There are no significant differences according to framing for the measured eye movements or epistemic actions.

H-E1$_{SH}$: The mean of the number of fixations over all games is lower for framed players than non framed players. This coincides with over all longer fixations, a higher percentage of horizontal eye movements, larger saccadic amplitudes and a higher mean of the ratio of epistemic actions over all actions.

Table B.33 shows that the expertise based measures show no statistically significant differences for any of them. This means, H-E1$_0$ has to be accepted.

6.4.3.2 Effect of Type of Adaptation

H-E2$_0$: There are no significant differences for any of the recorded eye movements or epistemic actions depending on the type of game adaptation.

H-E2$_{SH}$: The mean of the number of fixations is lower for players of games with eye movement based adaptation compared to conventionally adapted games. This coincides with over all longer fixations, a higher percentage of horizontal eye movements, larger saccadic amplitudes and a higher mean of the ratio of epistemic actions over all actions.

The type of adaptation yields no significant differences between performance measures as calculated for this hypotheses as can be seen in Table B.34. Accordingly, H-E2$_0$ holds with a higher likelihood.

6.4.3.3 Effect of Framing in Conventionally Adapted Games

H-E3$_0$: Framing has no significant effect on players' eye movements and epistemic actions when they play conventionally adapted games.

H-E3$_{SH}$: The mean of the number of fixations is lower while the number of OOI-Fixations is lower for framed players than non-framed players, even if they play a conventionally adapted game. This coincides with over all longer fixations, a higher percentage of horizontal eye movements, larger saccadic amplitudes and a higher mean of the ratio of epistemic actions over all actions.

Since OOI-Fixations were used as an input-variable to adaptation for EMTRIS, it can only be used as an analytical tool for conventionally adapted games. It is also the only measure which shows significant differences according to the framing category when

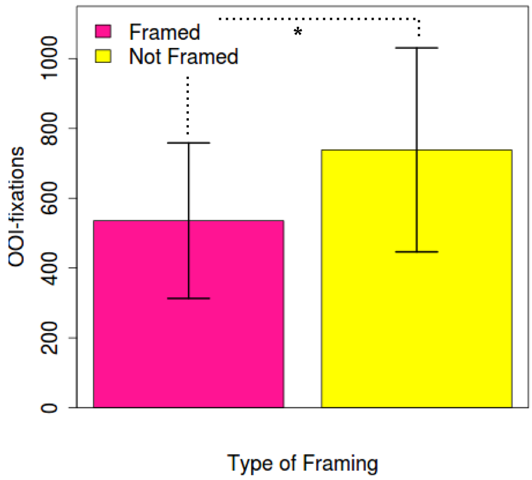


FIGURE 6.5: OOI-Fixations for Framed Players in Conventionally Adapted Games. Significance is denoted as $* \rightarrow p < 0.05$.

compared within conventionally adapted games with a medium to strong effect (F : $535.8, N : 738.5, p < 0.05, d = 0.78$, see also Figure 6.5 and Table B.35). This means, players had less fixations outside of the focus regions within a game when they were framed. H-E3$_{SH}$ can only be accepted with higher confidence for OOI-Fixations.

6.4.3.4 Effect of Type of Adaptation on Framed Players

H-E4$_0$: The type of adaption has no significant effect on framed players' eye movements and epistemic actions.

H-E4$_{\text{SH}}$: The mean of the number of fixations is lower for framed players playing a game adapted on eye movements compared to a conventionally adapted game. This coincides with over all longer fixations, a higher percentage of horizontal eye movements, larger saccadic amplitudes and a higher mean of the ratio of epistemic actions over all actions.

According to Table B.36, the type of adaptation did not yield significantly differences for framed players. Thus, H-E4$_0$ is accepted with higher confidence.

6.4.3.5 Player Improvement

H-E5$_0$: Framing has no significant effect on the difference between eye movements or epistemic actions between the type of adaptation.

H-E5$_{\text{SH}}$: The difference between the mean of the number of fixations, fixation duration, the ratio of horizontal eye movements, saccadic amplitudes and the ratio of epistemic actions according to type of adaptation of a game is less for framed players than non-framed players when playing a game with eye movement based adaptation.

This investigation also let to no significant findings (see Table B.37). H-E5_0 is more likely.

6.4.3.6 Summary

Table 6.4 summarises the findings for expertise based hypotheses. Only the measure of focus of fixations with the number of OOI-fixations yields significant differences for framed players playing conventionally adapted games.

Hypothesis	Result
H-E1_{SH}	✗
H-E2_{SH}	✗
H-E3_{SH}	for OOI-fixations
H-E4_{SH}	✗
H-E5_{SH}	✗

TABLE 6.4: Summary of Expertise Based Hypotheses

6.4.4 Intersectional Hypotheses

Intersectional hypotheses were analysed in order to examine mutual impacts of the measured variables.

6.4.4.1 Effect of Cleared Lines on Gameplay Experience

H-I1_0: There is hardly any effect between the number of cleared values and reported gameplay experience.

H-I1$_{SH}$: There is a strong effect between the number of cleared lines and high values for a desired gameplay experience.

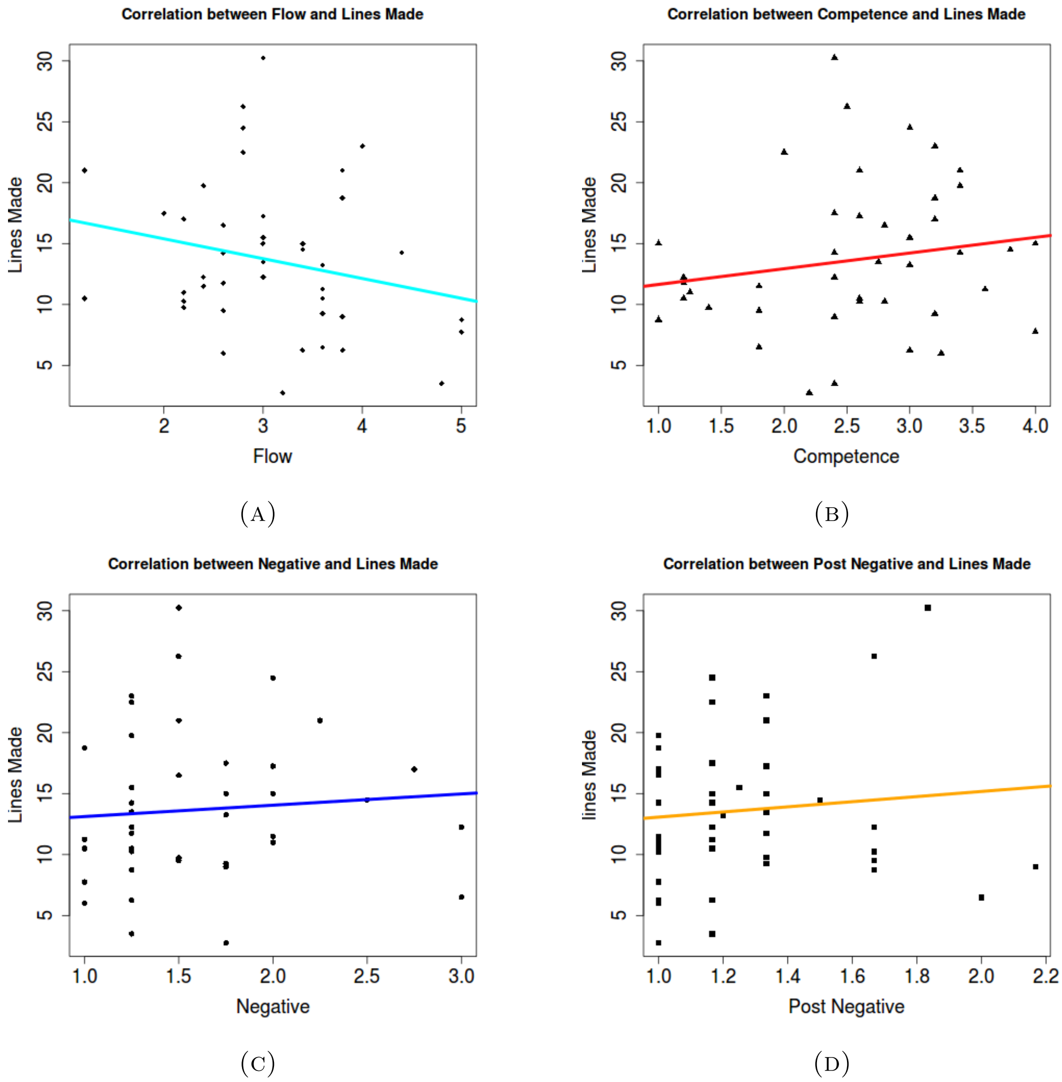


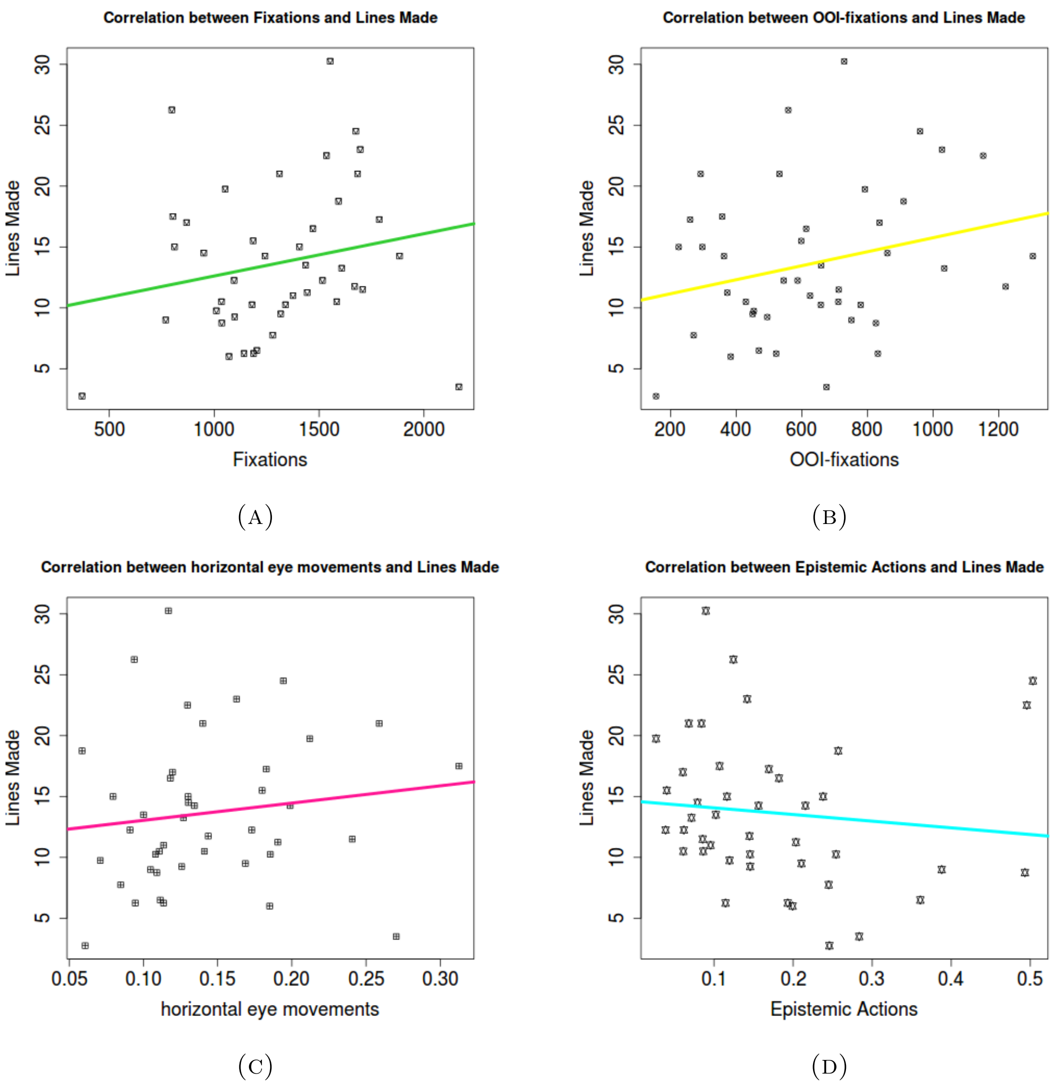
FIGURE 6.6: Correlations of Lines Made on GEQ Values. Only for values, for which the correlation was $r, \rho > 0.1$

Table B.38 and Figure 6.6 show the correlations between lines made and GEQ values. Flow ($r = 0.230$), competence ($r = 0.171$), negative($\rho = 0.208$) and post negative($\rho = 0.106$) are the only values where the correlation is notable, but weak. Visual inspection shows, that potential effects are rather small. Following this, H-I1$_0$ can be accepted with low confidence.

6.4.4.2 Effect of Cleared Lines on Eye Movements and Epistemic Actions

H-I2$_0$: There is hardly any effect between the number of cleared lines and a player's eye movements and epistemic actions.

H-I2$_{SH}$: There is a strong effect between the number of cleared lines and the number of fixations, the ratio of meaningful fixations, the ratio of horizontal eye movements, saccadic amplitudes and epistemic actions over all actions.



(A) (B) (C) (D)

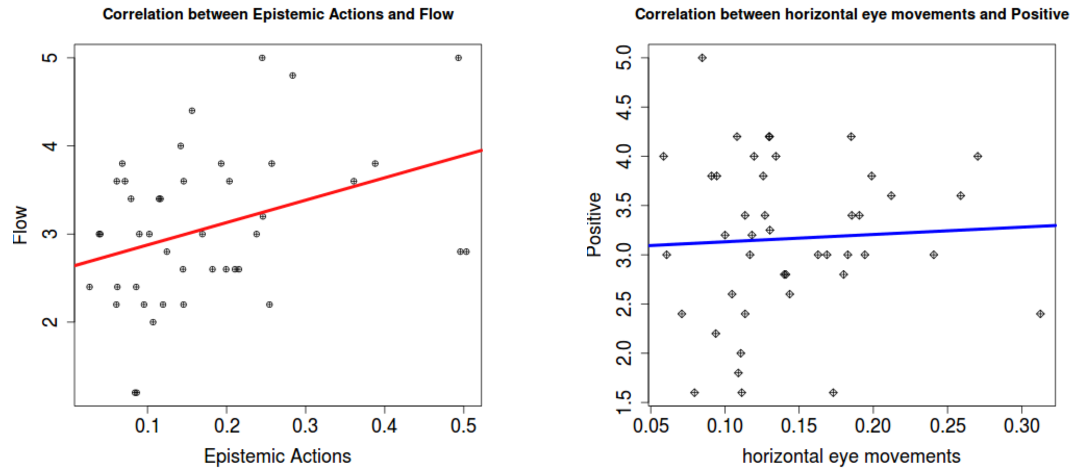
FIGURE 6.7: Correlations of Lines Made and Expertise Measures. Only for values, for which the correlation was $r, \rho > 0.1$

When there is an effect, it is again only due to a weak correlation (see Table B.39 and Figure 6.7). Since the number of OOI-fixations ($r = 0.258$) is the only expertise related measure that yielded any significant results, it is especially interesting to this work. However, no medium or strong correlations can be found, so H-I2$_0$ can be accepted.

6.4.4.3 Correlation between Eye Movements/Epistemic Actions and GEQ Values

H-I3$_0$: There is hardly any correlation between GEQ values and a player's eye movements or epistemic actions.

H-I3$_{SH}$: There is a strong effect between high values for a desired gameplay experience and the number of fixations, the ratio of meaningful fixations, the ratio of horizontal eye movements, saccadic amplitudes and the ratio of epistemic actions over all actions.



(A) (B)

FIGURE 6.8: Correlations of GEQ Values and Expertise Measures. Only for values, for which the correlation was $\rho > 0.3$

Only small correlations can be established between GEQ values and a player's eye movements or their epistemic actions (see also Figure 6.8 and Table B.40). Hence, those can be seen as fairly independent values and H-I3$_0$ holds.

6.4.4.4 Summary

Table 6.5 shows how all intersectional hypotheses have to be rejected with the data produced by the conducted study.

Hypothesis	Result
H-I1$_{\mathrm{SH}}$	✗
H-I2$_{\mathrm{SH}}$	✗
H-I3$_{\mathrm{SH}}$	✗

TABLE 6.5: Summary of Intersectional Hypotheses

Chapter 7

Discussion

This chapter first discusses interpretations of results according to the hypotheses in Section 7.1. In order to understand the results in a well rounded matter, they also have to be tested against other side effects such as age and language (Section 7.2). In Section 7.3 there is a take on what the results and their interpretation essentially mean when taken together and how they relate to the previously formulated expectations.

7.1 Interpretation According to Hypotheses

For the interpretation following the previously established hypotheses (see Chapter 3), the grouping is not according to measured variables but according to axes of analysis. This might yield a better understanding of how certain measures contribute to certain differences.

7.1.1 Framing

None of the framing related hypotheses show any significant effect. Neither H-P1 nor H-U1 or H-E1 lead to any significant differences between framed or non framed players. This means, that – on a global scale – simple framing does not have an influence on players by itself, although this is different when only looking at conventionally adapted games (see Section 7.1.3).

7.1.2 Use of Eye Movement Based Adaptation

While there was no effect of eye movement based adaptation on players' gameplay experience as established by the GEQ (H-U2) or their eye movements and epistemic actions

(H-E2), there are significant differences with strong and medium effect sizes for the speed and algorithms at the end of eye movement based adaptation. This means, that difficulty is formed differently for the two types of games and that the adaptation actually is different. In consequence, it can be stated that eye movement based adaptation is distinctly different than conventional adaptation – at least in this setup. While in conventionally adapted games, players got to a faster game albeit with an easier algorithm, players played games with a harder algorithm slower when they played games with eye movement based adaptation. Essentially, in games with eye movement based adaptation, the cognitive difficulty in terms of choosing block algorithm was higher, but the physical difficulty in terms of speed lower than in conventionally adapted games. This gives an indication towards the effects of eye movement based adaptation for future game designs. If a designer intends to make a game harder on a cognitive level, they can employ eye movement based adaptation in the fashion it has been done here (according to expertise). If they want it to be physically easier, they can adapt to the state of nervousness a player has.

It is also notable, that the performance measures of lines made and score were following a more normal distribution, which means, that eye movement based games might be easier and steadier to learn with no large gaps in between performances. This also would result in a more comparable performance, which might be desirable for comparative high score measures to keep players going and feel enabled to climb the high score ladder step by step. Since multiple test persons voiced the question for a high score, this can add to the effect of games with eye movement based adaptation, but should not be overstated as a fact. For educational games or educational software, this could prove to be a more directly relevant question and should, hence, be tested separately.

7.1.3 Framing in Conventionally Adapted Games

There was no significant difference in performance metrics (H-P3) for framed and non-framed players of conventionally adapted games. However, framed players felt more challenged than non-framed players (H-U3). This means, that framing does have an effect on the gameplay experience, albeit only in a limited way. While the actual challenged posed to the players did not differ, their perceived challenge did. This framing effect does not repeat for games with eye movement based adaptation. Hence, when a game is presented as more sophisticated, players might be influenced to expect a larger challenge even if it isn't there, but feel equally challenged when it is actually presented.

There was also a significant difference between Out-of-Interest Fixations for framed and non framed players in conventionally adapted games (H-E3). Since this is an input variable adapting in games with eye movement based adaptation, this value has only been taken for analysis in this specific setting. This means, that framed players show more TETRIS expert behaviour as defined by Jermann et al. [2010] in EMTRIS than in NEMTRIS.

Since the two variables are independent, they appear in the same situation, but are not related. This means, that a higher perceived challenge does not automatically lead to an expert viewing strategy; however, eye movement based adaptation does.

7.1.4 Adapting Games for Framed Players

Framed players played significantly faster in conventionally adapted games than in games with eye movement based adaptation (H-P4). This is a repetition of what we have seen in Section 7.1.2. However, the difficulty of algorithms is not significantly different between the games for framed players. This shows again how framing does have a smaller influence on a player in general than the actual difference between NEMTRIS and EMTRIS.

Another significant difference for the game type for framed players is competence as measured by the GEQ (H-U4). Framed players felt more competent in games with eye movement based adaptation than in conventionally adapted games. This might be because the game behaved more accordingly to their expectations and gave them the feeling that they were better prepared for it. Since the difference does not hold for non-framed players, this is a unique aspect for framed players.

A player's eye movements and their epistemic actions are not affected either way for framed players when they play differently adapted games (H-E4).

7.1.5 Framing and Adaptation Differences

The differences (or improvements) between games were not significant between framed and non-framed players for any of the recorded measures be it for performance (H-P5), gameplay experience (H-U5) or expert behaviour (H-E5). Considering the previous results, this is not unexpected. It again shows the limited general effect of framing on games.

7.1.6 Intersectional Influences

There were only small correlations between the values, which means, that they can be seen as independent measures. This demonstrates the validity of recording the variety of measures as has been done in this user study.

7.2 Side Effects

As stated in Section 5.2.1, external effects which have no hypotheses attached have to be investigated as well in order to report on a well-rounded basis. This section, hence, looks at several categories and assesses their influence on performance, user experience and eye movements respectively.

7.2.1 Gender

As Table B.41 shows, no significant influence of gender is given on the performance measures. Note, that for this calculation, only those participants who identified themselves on a binary gender schematic, were considered. This result is contrary to the findings of Okagaki and Frensch [1994]. This can either mean that adaptation in (N)EMTRIS leads to mitigation of gender effects or that they were not there in the first place.

Similarly, gender did not play a role in how players experienced (N)EMTRIS as can be seen in Table B.42. The same accounts for eye movements and epistemic actions (see Table B.43). According to these results, gender is an irrelevant category.

As an addendum: There was also no significant difference in the mental rotation performance according to gender ($f = 10.76, m = 12.75, p = 0.166$) contrary to a variety of published research (amongst others: Collins and Kimura [1997] or Parsons et al. [2004] (for paper-based tests)).

7.2.2 Age

Correlations for age and any performance variables were only with a small effect indicating trends (see Table B.44). The same is the case for expert behaviour (see Table B.46) with the exception of horizontal eye movements ($\rho = 0.390$).

For the GEQ results, Table B.45 indicates how there are several correlations of medium strength for flow ($\rho = -0.365$), challenge ($\rho = -0.318$) and tension ($\rho = -0.340$). Since they are not significantly different to each other according to framing categories

($F = 27.68, N = 27.57, p = 0.328$), the previous findings for challenge in conventionally adapted games are still valid.

7.2.3 Language

There were only seven test participants for whom the whole test setup was performed in English. All others have been conducted in German. The sample sizes are hence, rather different.

No statistically significant differences according to language have been found for performance measures (see Table B.47) or for any relevant aspects of the GEQ (see Table B.48). This shows, that the questionnaires were roughly equivalent. We decided against calculating Cronbach's [cf. Cronbach, 1951], because there was only a small amount of items per dimension available through the GEQ. Validation via internal consistency is hardly ever applicable with such small categories.

In expertise based measures, the difference between the language category was significant, albeit with a very small effect size ($U = 52, p < 0.05, Z = 2.43, r = 0.057$), which makes this finding irrelevant. Also no significant reports have been made with this measure in the main hypotheses.

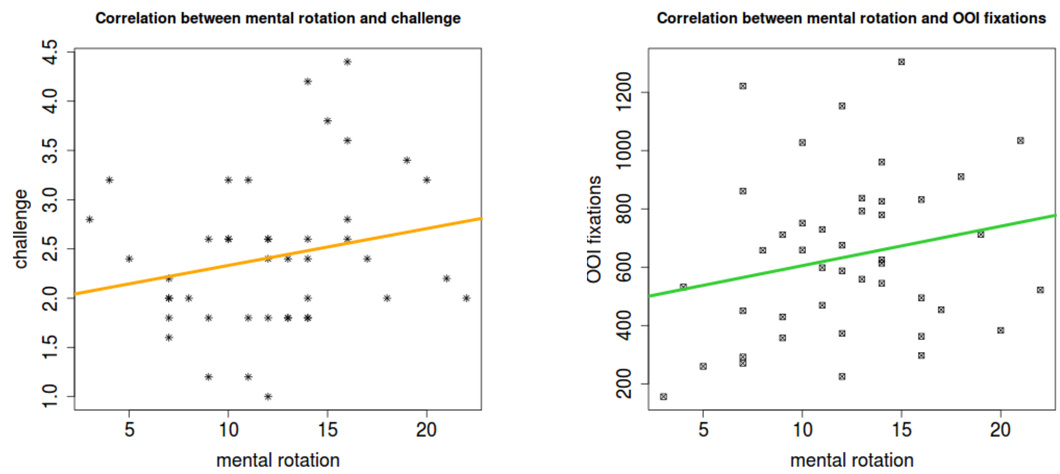
7.2.4 Mental Rotation

Mental rotation skills only showed weak correlations with the measured variables for performance (see Table B.50), gameplay experience (see Table B.51) or expert behaviour in the form of eye movements or epistemic actions (see Table B.52). The strongest relationship exists for the perceived challenge ($r = 0.219$) and OOI-fixations ($r = 0.218$), however those are still rather week (see also Figure 7.1).

While the correlation of perceived challenge can only be interpreted speculatively, it does make sense, that players with high mental rotation skills also know where to look at in Tetris. Since the differences in mental rotation according to framing are not significant ($F : 12.05, N : 12.33, p = 0.836$), the effect reported for $H - E3$ still holds.

7.2.5 Tetris expertise

Due to the Dunning-Kruger effect discussed in Section 6.3, the self reported values and the values established via Pytris have to be analysed separately.



(A) (B)

FIGURE 7.1: Correlations for Mental Rotations and Challenge/OOI Fixations.

7.2.5.1 Self Reported

Since the correlations between self reported values and performance values (see Table B.53), gameplay experience (see Table B.54) or expert behaviour as measured here (see Table B.55) are all very weak, it can be negated. The Dunning-Kruger effect also repeats with a small effect on actual performance.

7.2.5.2 Established by Pytris

The medium correlation between expertise as established by PYTRIS and actual performance measures (see Table B.56) shows in hindsight, how important it was, to have an objective measure for experience. In this case, the data validates the approach. Since there is no strong effect between performance in (N)EMTRIS and performance in classical TETRIS, this can also be taken as a mild proof of concept for (N)EMTRIS as a game, mitigating effects of previous expertise, in that the challenge is adapted.

Since the perceived challenge does not correlate with previous expertise (see Table B.57), these two measures appear to be independent. However, it is important to point out, that more experienced players felt less flow and more negative (also afterwards) about the gameplay experience, but again only with small effects.

Taking a look at the expert behaviour (see Table B.58), it is striking, that OOI-fixations and previous expertise correlate on a medium level but negatively. However, lines made and OOI-fixations also correlate with $\rho = 0.258$.

Since expertise does not significantly differ between the framing categories ($p = 0.380$), it can be expected that its effect are distributed equally along them.

7.2.6 Summary

All in all, the side effects play only a small role for the research presented here. While gender and language categories had absolutely no influence, age and mental rotation had occasional weak correlations. However, both attributes are distributed over the between-subject category of framing in such a way, that there are no significant differences for either statistical attribute between the framing categories. The same accounts for expertise – at least as established by Pytris. For self reported expertise, the Dunning-Kruger effect emerges again.

7.3 Expected vs. Actual Results

While many of the hypotheses have to be rejected after an investigation into the data, the discussion showed insights beyond the hypotheses. It can be said, that framing does shape the gameplay experience in terms of challenge when conventionally adapted games are played. The subjective interpretation of the challenge a game provides can, hence, be different when certain attributes of a game are only implied and not actually at hand. This means, that future research should carefully test for framing effects when it deals with games or technological enhancement. Framing also has an effect on eye movements in that there were significantly less OOI-fixations for framed players than those who did not know about eye movement based adaptation. The reason for this is unclear and could be investigated further. Maybe an insight into how concentrated the players were in their game in connection with expert eye movements could help shed some light onto this matter.

Furthermore, it has also been established, that eye movement based adaptation is different from conventional adaptation – at least in how it has been done here. It could have been, that there was no significant difference at all between the games, but that speed and algorithm at the end differ according to type of adaptation shows, that this is not the case. Even if there were not many differences along the dimensions of the GEQ, framed players also felt more competent when playing games with eye movement based adaptation compared to the conventionally adapted versions.

All in all, although few of the aspects investigated here yielded significant differences, those which did, did so with large effect sizes. This means, they should be taken seriously for future game development and further research.

Chapter 8

Conclusion

This work concludes with the presentation of the real life usefulness of research for games, educational software and adaptive applications in Section 8.1. More general ramifications about the possibilities of generalisation of the steps taken here to designing a game with psychophysical adaptation are presented in Section 8.2. As with any research, the work here is not exhaustive and hopefully opens up further productive questions (Section 8.3). Some of the limitations and more direct open questions are shown in Section 8.4. Finally, a summary (Section 8.5) reiterates the goals and results of this thesis and where its core contributions lie.

8.1 Application of Research

The implications of the results as discussed in Chapter 7 lead to the unsurprising conclusion, that the actual implementation of eye movement based adaptation in games might have more of an effect on players than simply framing them. These results can be used in hands-on software development of games and adaptive software, in general.

First of all, however, it has been shown that eye movement based adaptation can be brought outside of the lab setting as well. As Johansen et al. [2011] shows, low-cost eye trackers become usable in many settings where eye tracking is needed. With eye trackers becoming smaller, more efficient and less costly, it appears to be viable to look into their possible use within at home settings within an entertainment context or otherwise. The Eye Tribe[1] was the first low-cost eye tracker commercially available. While its temporal resolution is very low (either 30 Hz or 60 Hz), its accuracy (between 0.5-1°) is acceptable for most use cases in usability testing. While it was not used within the

[1] See https://theeyetribe.com/

work presented here, it shows how development in this technological area is very active and can lead to applications with eye movement based adaptivity.

For game development we were able to establish, that eye movement based adaptation as implemented here added to cognitive difficulty while simultaneously lowering the physical difficulty. While this has to be verified further, it was also dependent on the type of adaptation. Since algorithms were adapted to an expertise indicator and speed was adapted to an emotional indicator, game developers can take this knowledge and apply it to active game design while conducting playtesting sessions in order to verify this for their specific game.

Developers of adaptive software in general can now use these results knowing that eye movement based adaptation makes a difference, even if the measured time frame is comparatively short; here it was only one tetromino episode. Eye movement based adaptation could be used in a modular way, where it establishes a user's needs ad hoc whenever this is needed. With this, eye movement based adaptation can be a powerful and flexible tool at the hands of developers.

8.2 Designing Games with Psychophysical Adaptation

With the example design of (NEM)TRIS as a game with psychophysical adaptation based on eye movements, the question arises of how the process of developing games with psychophysical adaption can be put into a schematic which others can easily follow. Together with the previous work on eye movement based adaption in games [see Wetzel et al., 2014], it becomes obvious, that the process of designing such games is necessarily iterative (see also Figure 8.1). For best results, theory and data driven approaches should complement each other. This means, that for the development of psychophysically adapted games the following steps are recommended. The focus lies on adaption with eye movements, general examples are given as well.

1. prototype: **suitable game**
   A suitable game can either be created or an existing one used. It should be checked for its suitability in terms of the employed psychophysical adaption. For example, when using eye movements multidimensional visual features that demand a distributed attention on at least one dimensional axis are suitable. When using heart rate, temporal features might be more important.

2. theory: **meaningful adaption**
   It is furthermore important to adapt in meaningful ways to the game. Therefore,

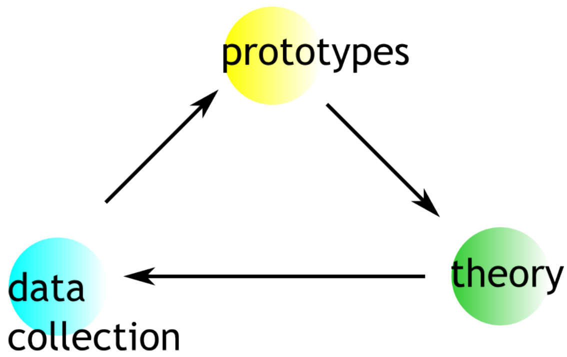


FIGURE 8.1: The Iterative Scheme behind the Process of Designing Games with Psychophysical Adaptation

a developer should create a theoretical model of what affects a game's difficulty and how these parameters can be changed.

3. data collection: **verification I**
   In order to verify that the adaption as planned in step two actually influence difficulty for players in the expected way developers should conduct user studies to verify the theory. If the theory is not verified, further adjustments and potentially studies are required in order to ensure suitable adaption mechanisms.

4. prototype: **theory driven prototype**
   With a first prototype employing the plain game and some adaption mechanisms outside of psychophysical measures, developers should find out, how the psychophysical data is distributed for that game. This means, that exemplary data has to be created in order to establish the suitability of the psychophysical attribute on which to adapt.

5. theory: **establish meaningful attributes**
   Direct or derived measures of the psychophysical data should be established in a theory driven way.

6. data collection: **verification II**
   In order to verify the suitability of the whole approach, user studies should be conducted in order to verify whether the adaptive mechanisms work for the user. Repeat the last two steps until user studies indicate that the adaption process functions as intended.

7. prototype: **data driven prototype**
   This version of the game can be seen as the first conceptual draft of a psychophysical game.

8. theory: **formulate hypotheses**
   Hypotheses about the effect of psychophysical adaptation for this game have to be formulated.

9. data collection: **user study**
   A user study verifies or counteracts the hypotheses and helps in further steps in prototyping, theorising and data collection.

This work did not follow all of these steps, but with it contributed to the list by being critically analysed. The list is also a collection of suggestions on how to learn from previous mistakes and do it better.

8.3 Future Research

Like with any research, there are aspects which could be optimised further and new directions that are offered up from here for new research projects.

8.3.1 Optimisation of Current Research

In the current versions of (N)EMTRIS, the eye data has been used only for the spans of episodes. A new version, where adaptation according to eye movements and their development over a game are considered together might result in a better approach in adaptation and generally result in a better gameplay experience for players.

Furthermore, this research was focused on a quantitative data analysis. Due to experiences in interaction with software in general and especially experiences with games, a qualitative investigation into the question of how users perceive games might be additionally required to get a more well-rounded picture of the gameplay experience in games with eye movement adaptation.

In order to find out how the type of eye tracker (head-mounted vs. remote) might influence the results, a future study should try to conduct this research with a high-quality remote eye tracker. It might additionally be helpful to attempt to create a more natural gaming environment for players than a lab setup.

8.3.2 Advancing Further

There are also new questions that arise after the presentation of this research. Those (together with other research) open up new avenues for further research.

With this work and the work of Wetzel et al. [2014] eye movement based adaptation has been implemented in two different games in two different ways. Whereas here, game mechanics have been accessed immediately and the game is meant as a solo activity, HEX, the game Wetzel et al. [2014] used, was antagonistic in nature and the adaptation mechanisms were indirectly influencing an AI game engine. The difference in eye movements and eye movement based adaptation between antagonistic games and single player games would be an option for further investigation according the axis of formal development as well as the axis of gameplay experience. There might be an intrinsic difference in eye movements for antagonistic and single player games.

One could also think about adapting to eye movements in a different way. There are all kinds of players with different experiences and different levels of expertise out there. There are also different levels of adaptation, one could think of Bertel [2014]. Investigating into the differences of general, individual and situational aspects of gameplay and, hence, general, individual and situational forms of adaptation via eye movements, their effects and the differences might produce valuable results, that give game developers more knowledge about the tool eye movement based adaptation. Another option would be to abstract from the live knowledge and compare it to previously acquired data about a player.

Based on this, the question arises whether it is possible to learn and predict eye movements of certain players or certain player types algorithmically. A system could extrapolate from current eye movements and react more accordingly to a player's eye movements. This might be suitable to games, but probably even more to learning software that would be expected to react preemptively to a user's problems and steer them in the desired direction. However, new issues about user direction and normalisation arise as well and should be discussed alongside the technological research. Furthermore, the question should be asked on whether adaptation can and should be adaptive itself.

From a managements perspective, it would be helpful, if a cost-benefit analysis of eye movement analysis incorporated in software yields good results. For that, the benefit of what eye movement based adaptation for example could give has to be defined first. The costs are not only in the hardware, which becomes cheaper and cheaper anyway (see also Section 8.1), but also in the cost of analysis, interpretation and acquisition. For analysis and interpretation, it could be helpful to provide intuitive interfaces, that are easily understandable for beginners in the field such as OPEN EYES [Wetzel, 2014].

Finally, there is also the question into the adequacy of quantifying approaches in game experience. After all, is it really desirable, to develop to the mean of the best game experience? Would not that create a higher likelihood for games that aim to please everyone and succeed to be boring for all? Game experiences are formed along several axes. The environment in which the gaming happens, the social level, background music, disruptions by e.g. pets all add up to the experience in ways developers cannot always forsee. Simply by the fact that some participants in the pre-study took their break as scheduled, some took more and some did not take any break, it shows, that comfort in playing might not be easily described by a quantitative user study that does not discriminate between types and does not include any qualitative descriptions or situated knowledge about individual cases. Research about the fluent developments of eye movements in different game situations for different types of players might lead to more fine-grained systems that also include potential player groups with e.g. different viewing patterns, such as players with autism in games that include characters with faces [cf. Pelphrey et al., 2002].

8.4 Limitations and Open Questions

To generalise from these results is hard, since every study has its limitations which are bound to the environment it has been conducted in and – especially within usability research – the exemplary objects used for the study. Scalability and validity is given not in the rigorous way mathematics does, but in an exploratory way, where different options are tested continuously.

Introducing a different form of eye movement based adaptation offers a new way of technological advance, that has not been done in this way before. As such, game designers have a new form of adaptation at their hand and maybe feel inspired to try out new ways for adapting to eye movements. The actual effects on performance might not change; however, the fun of games is not always connected to performance.

The approach taken here was one option out of many. The game has been directly adapted in its only core mechanisms. Partway adaptations or passive eye movement based interaction in more complex visuo-spatial games have yet to be developed and discovered. The same accounts for collaborative settings in which players are supposed to play together. They might produce different eye movements depending on circumstances such as group size and purely cooperative vs. group against group settings.

8.5 Summary

This work initially set out to ask the question: What has a larger effect on the performance, gameplay experience and expert behaviour of players – Framing the game in a way that suggests the use of eye movement based adaptation or actually implementing this type of adaptation? To answer this question, two versions of adaptive TETRIS have been implemented: NEMTRIS, which only uses game state based adaptation mechanisms and EMTRIS, which additionally uses eye movements for the adaptive calculations.

In a mixed-method study, players tested the different adaptation types in a within-subject setting whereas their framing was conducted between subjects. This means, that every player played both types of games, but under different pretenses.

Results show, that the two versions of (N)EMTRIS create significantly different games. Eye movement based adaptation increased the cognitive level of difficulty by adapting to expert behaviour and decreased the physical level of difficulty by adapting to the efficacy of visual inspection. Framed players have different experiences in that they feel more challenged when playing conventionally adapted games compared to non-framed players; they also exhibit more expert behaviour in their eye movements than non-framed players. Framed players also report on a different level of perceived competence in games with eye movement based adaptation compared to conventionally adapted games.

This means that both the type of game adaptation and the framing individually contribute to different aspects of gameplay experience. Developers should consider both effects carefully when including eye movement based adaptation in their games.

Bibliography

Espen Aarseth. A Narrative Theory of Games. In *Proceedings of the International Conference on the Foundations of Digital Games*, pages 129–133. ACM, 2012.

Hervé Abdi. Bonferroni and Šidák Corrections for Multiple Comparisons. In NJ Salkind, editor, *Encyclopedia of Measurement and Statistics*. Thousand Oaks, CA: Sage, 2007.

John R. Anderson. *Cognitive Psychology and Its Implications*. Worth Publishers, New York, 7th edition edition, 2000.

Robert Appelman. Experiential Modes of Game Play. In *Situated Play, Proceedings of the DiGRA 2007 Conference*, pages 815–822, 2007.

Bruce Lawrence Berg and Howard Lune. *Qualitative Research Methods for the Social Sciences*, volume 5. Pearson Boston, 2004.

Regina Bernhaupt. User Experience Evaluation in Entertainment. In *Evaluating User Experience in Games*, pages 3–7. Springer, 2010.

Sven Bertel. *Spatial Structures and Visual Attention in Diagrammatic Reasoning*. Pabst Science Publ., 2010.

Sven Bertel. Individual Cognitive Abilities and Styles in HCI: Three Main Challenges and a Tiered Adaption Model. In *Proceedings of the 6th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, EICS '14, pages XX–XX, New York, NY, USA, 2014. ACM.

Ian Bogost. Gamification is Bullshit. *Ian Bogost Blog*, 8, 2011.

Niko Böhm, Gabriella Kókai, and Stefan Mandl. An Evolutionary Approach to Tetris. In *6th Metaheuristics International Conference (6th Metaheuristics International Conference Wien August 22-26, 2005)*, 2005.

Jeanne H Brockmyer, Christine M Fox, Kathleen A Curtiss, Evan McBroom, Kimberly M Burkhart, and Jacquelyn N Pidruzny. The development of the Game

Engagement Questionnaire: A Measure of Engagement in Video Game-Playing. *Journal of Experimental Social Psychology*, 45(4):624–634, 2009.

Emily Brown and Paul Cairns. A Grounded Investigation of Game Immersion. In *CHI'04 Extended Abstracts on Human Factors in Computing Systems*, pages 1297–1300. ACM, 2004.

John Brzustowski. Can You Win at Tetris? Master's thesis, University of British Columbia, 1992.

Heidi Burgiel. How to Lose at Tetris. *Mathematical Gazette*, 81:194–200, 1997.

Judith Butler. *Bodies that Matter: On the Discursive Limits of Sex*. Taylor & Francis, 2011.

Frederik Jacobus Johannes Buytendijk. *Wesen und Sinn des Spiels*. Ayer Company Pub, 1976.

Roger Caillois. *Man, Play, and Games*. University of Illinois Press, 2001.

Roger Caillois and Elaine P Halperin. The Structure and Classification of Games. *Diogenes*, 3(12):62–75, 1955.

Eduardo H Calvillo-Gámez, Paul Cairns, and Anna L Cox. Assessing the Core Elements of the Gaming Experience. In *Evaluating User Experience in Games*, pages 47–71. Springer, 2010.

Darryl Charles, A Kerr, M McNeill, M McAlister, M Black, J Kcklich, A Moore, and K Stringer. Player-Centred Game Design: Player Modelling and Adaptive Digital Games. In *Proceedings of the Digital Games Research Conference*, volume 285. Citeseer, 2005.

BD Chaurasia and BBL Mathur. Eyedness. *Cells Tissues Organs*, 96(2):301–305, 1976.

David W Collins and Doreen Kimura. A Large Sex Difference on a Two-Dimensional Mental Rotation Task. *Behavioral Neuroscience*, 111(4):845, 1997.

DJ Colwell and JR Gillett. 66.49 Spearman versus Kendall. *The Mathematical Gazette*, pages 307–309, 1982.

Laura Cowen, Linden Js Ball, and Judy Delin. An Eye Movement Analysis of Web Page Usability. In *People and Computers XVI-Memorable Yet Invisible*, pages 317–335. Springer, 2002.

Lee J Cronbach. Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, 16(3):297–334, 1951.

Mihaly Csikszentmihalyi. *Flow: The Psychology of Optimal Experience*, volume 41. HarperPerennial New York, 1991.

Erik D Demaine, Susan Hohenberger, and David Liben-Nowell. Tetris is Hard, Even to Approximate. In *Computing and Combinatorics*, pages 351–363. Springer, 2003.

Marc Destefano, John K. Lindstedt, and Wayne D. Gray. Use of Complementary Actions Decreases with Expertise. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society*, pages 2709–2014, 2011.

Sebastian Deterding. Wohnzimmerkriege. Vom Brettspiel zum Computerspiel. *Strategie spielen. Medialität, Geschichte und Politik des Strategiespiels. Münster: Lit-Verlag*, pages 87–113, 2008.

Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. From Game Design Elements to Gamefulness: Defining Gamification. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, pages 9–15. ACM, 2011.

A.T. Duchowski. *Eye Tracking Methodology: Theory and Practice*. Springer, 2003.

Paul D Ellis. *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results*. Cambridge University Press, 2010.

Laura Ermi and Frans Mäyrä. Fundamental Components of the Gameplay Experience: Analysing Immersion. *Worlds in Play: International Perspectives on Digital Games Research*, page 37, 2005.

Colin Fahey. Tetris, July 2012. URL http://colinfahey.com/tetris/tetris.html.

A. Field, J. Miles, and Z. Field. *Discovering Statistics Using R*. SAGE Publications, 2012.

Paul Morris Fitts and Michael I Posner. Human Performance. 1967.

Landon Flom and Cliff Robinson. Using a Genetic Algorithm to Weight an Evaluation Function for Tetris, 2005.

Kiel M Gilleade and Alan Dix. Using Frustration in the Design of Adaptive Videogames. In *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, pages 228–232. ACM, 2004.

Hartmut Glücker, Felix Raab, Florian Echtler, and Christian Wolff. EyeDE: gaze-enhanced software development environments. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pages 1555–1560. ACM, 2014.

Joseph H Goldberg and Xerxes P Kotval. Computer Interface Evaluation Using Eye Movements: Methods and Constructs. *International Journal of Industrial Ergonomics*, 24(6):631–645, 1999.

Joseph H Goldberg and Anna M Wichansky. Eye Tracking in Usability Evaluation: A Practitioner's Guide. *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, pages 573–605, 2003.

Joseph H Goldberg, Mark J Stimson, Marion Lewenstein, Neil Scott, and Anna M Wichansky. Eye Tracking in Web Search Tasks: Design Implications. In *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*, pages 51–58. ACM, 2002.

Mark Griffiths. Violent Video Games and Aggression: A Review of the Literature. *Aggression and Violent Behavior*, 4(2):203–212, 1999.

Dan Witzner Hansen, David JC MacKay, John Paulin Hansen, and Mads Nielsen. Eye Tracking Off the Shelf. In *Proceedings of the 2004 symposium on Eye tracking research & applications*, pages 58–58. ACM, 2004.

Jan Hartmann, Antonella De Angeli, and Alistair Sutcliffe. Framing the User Experience: Information Biases on Website Quality Judgement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 855–864. ACM, 2008.

Marc Hassenzahl. The Thing and I: Understanding the Relationship Between User and Product. In *Funology*, pages 31–42. Springer, 2005.

Jan Hauke and Tomasz Kossowski. Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data". *Queastiones Geographicae*, 2: 87–93, 2011.

Judith Ann Hirsch, Beverly Bishop, et al. Respiratory Sinus Arrhythmia in Humans: How Breathing Pattern Modulates Heart Rate. *Am J Physiol*, 241(4):H620–H629, 1981.

Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van de Weijer. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press, 2011.

Anke Huckauf and Mario H Urbina. On Object Selection in Gaze Controlled Environments. *Journal of Eye Movement Research*, 2(4):4, 2008.

Johan Huizinga. Homo Ludens: Vom Ursprung der Kultur im Spiel.(1938). *Hamburg: Rowohlt*, 1956.

WA IJsselsteijn, YAW De Kort, and K Poels. The Game Experience Questionnaire: Development of a Self-Report Measure to Assess the Psychological Impact of Digital Games. Manuscript in Preparation. 2013.

Wijnand IJsselsteijn, Yvonne de Kort, Karolien Poels, Audrius Jurgelionis, and Francesco Bellotti. Characterising and Measuring User Experiences in Digital Games. In *International Conference on Advances in Computer Entertainment Technology*, volume 2, page 27, 2007.

Katherine Isbister and Noah Schaffer. *Game Usability: Advancing the Player Experience*. CRC Press, 2008.

Shanto Iyengar and Donald R Kinder. *News that Matters: Television and American Opinion*. University of Chicago Press, 2010.

Susan A Jackson and Robert C Eklund. Assessing Flow in Physical Activity: The Flow State Scale-2 and Dispositional Flow Scale-2. *Journal of Sport & Exercise Psychology*, 2002.

Susan A Jackson, Herbert W Marsh, et al. Development and Validation of a Scale to Measure Optimal Experience: The Flow State Scale. *Journal of Sport & Exercise Psychology*, 18:17–35, 1996.

Charlene Jennett, Anna L Cox, Paul Cairns, Samira Dhoparee, Andrew Epps, Tim Tijs, and Alison Walton. Measuring and Defining the Experience of Immersion in Games. *International Journal of Human-Computer Studies*, 66(9):641–661, 2008.

Patrick Jermann, Marc-Antoine Nüssli, and Weifeng Li. Using Dual Eye-Tracking to Unveil Coordination and Expertise in Collaborative Tetris. In Tom McEwan and Lachlan McKinnon, editors, *BCS HCI*, pages 36–44. ACM, 2010.

Sune Alstrup Johansen, Javier San Agustin, Henrik Skovsgaard, John Paulin Hansen, and Martin Tall. Low Cost vs. High-End Eye Tracking for Usability Testing. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pages 1177–1182. ACM, 2011.

Jesper Juul. *Half-Real: Video Games Between Real Rules and Fictional Worlds*. MIT press, 2011.

Daniel Kahneman. *Thinking, Fast and Slow*. Farrar, Straus and Giroux, New York, 2011. ISBN 9780374275631 0374275637.

Mitu Khandaker. Column: "Gambrian Explosion": Games, Randomness, and The Problem with Being Human, April 2011. URL http://www.gamesetwatch.com/2011/04/column_gambrian_explosion_game.php.

David Kirsh and Paul Maglio. Reaction and Reflection in Tetris. In *Artificial Intelligence Planning Systems: Proceedings of the First Annual International Conference (AIPS92*. Morgan Kaufman, 1992.

David Kirsh and Paul P. Maglio. On Distinguishing Epistemic from Pragmatic Action. *Cognitive Science*, 18(4):513–549, 1994.

J Matias Kivikangas et al. Psychophysiology of Flow Experience: An Explorative Study. 2006.

Paul A Kolers. Reading a Year Later. *Journal of Experimental Psychology: Human Learning and Memory*, 2(5):554, 1976.

Hannu Korhonen, Markus Montola, and Juha Arrasvuori. Understanding Playful User Experience through Digital Games. In *International Conference on Designing Pleasurable Products and Interfaces*, pages 274–285, 2009.

Raph Koster. *Theory of Fun for Game Design.* " O'Reilly Media, Inc.", 2013.

Justin Kruger and David Dunning. Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments. *Journal of Personality and Social Psychology*, 77(6):1121, 1999.

Robert Ladouceur, Anne Gaboury, Annie Bujold, Nadine Lachance, and Sarah Tremblay. Ecological Validity of Laboratory Studies of Videopoker Gaming. *Journal of Gambling Studies*, 7(2):109–116, 1991. ISSN 1050-5350.

Irwin P Levin, Sandra L Schneider, and Gary J Gaeth. All Frames are not Created Equal: A Typology and Critical Analysis of Framing Effects. *Organizational Behavior and Human Decision Processes*, 76(2):149–188, 1998.

John Lindstedt and Wayne Gray. Extreme Expertise: Exploring Expert Behavior in Tetris. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, Cooperative Minds: Social Interaction and Group Dynamics, pages 912–917, 2013.

Henry Lowood. Videogames in Computer Space: The Complex History of Pong. *IEEE Annals of the History of Computing*, 31(3):5–19, 2009.

Paul P Maglio and David Kirsh. Epistemic Action Increases with Skill. In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*, volume 16, pages 391–396, 1996.

Jane McGonigal. *Reality is Broken: Why Games Make Us Getter and How They Can Change the World*. Penguin, 2011.

ED Megaw and J Richardson. Eye Movements and Industrial Inspection. *Applied Ergonomics*, 10(3):145–154, 1979.

Lennart Nacke and Craig Lindley. Boredom, Immersion, Flow – A Pilot Study Investigating Player Experience. *Proceedings of the IADIS Gaming 2008: Design for Engaging Experience and Social Interaction*, pages 25–27, 2008.

Lennart Nacke, Sophie Stellmach, Dennis Sasse, and Craig A. Lindley. Gameplay Experience in a Gaze Interaction Game. In A. Villanueva, J. P. Hansen, and B. K. Ersbøll, editors, *Proceedings of the 5th Conference on Communication by Gaze Interaction – COGAIN 2009: Gaze Interaction For Those Who Want It Most*, pages 49–54, Lyngby, Denmark, 2009a. The COGAIN Association, The COGAIN Association.

Lennart E Nacke, Anders Drachen, Kai Kuikkaniemi, Joerg Niesenhaus, Hannu J Korhonen, van den WM Hoogen, Karolien Poels, W IJsselsteijn, and Y Kort. Playability and Player Experience Research. In *Proceedings of DiGRA*, 2009b.

Lennart E Nacke, Mark N Grimshaw, and Craig A Lindley. More Than a Feeling: Measurement of Sonic User Experience and Psychophysiology in a First-Person Shooter Game. *Interacting with Computers*, 22(5):336–343, 2010.

Allen Newell. Unified Theories of Cognition. *Cambridge, MA: Harvard University*, 1990.

Lynn Okagaki and Peter A. Frensch. Effects of Video Game Playing on Measures of Spatial Performance: Gender Effects in Late Adolescence . *Journal of Applied Developmental Psychology*, 15(1):33–58, 1994.

Reinhard Oppermann and R Rasher. Adaptability and Adaptivity in Learning Systems. *Knowledge Transfer*, 2:173–179, 1997.

Zhan-He Ou and Ling-Hwei Chen. Hiding Data in Tetris. In *2011 International Conference on Machine Learning and Cybernetics (ICMLC)*, volume 1, pages 61–67. IEEE, 2011.

Volker Pantenburg and Stefanie Schlüter. Zehn Anmerkungen zur Filmbildung. In *FilmBildung*, pages 46–49. Schüren, Marburg, 2014.

Thomas D Parsons, Peter Larson, Kris Kratz, Marcus Thiebaux, Brendon Bluestein, J Galen Buckwalter, and Albert A Rizzo. Sex Differences in Mental Rotation and Spatial Rotation in a Virtual Environment. *Neuropsychologia*, 42(4):555–562, 2004.

Corinna Peifer. Psychophysiological Correlates of Flow-Experience. In Stefan Engeser, editor, *Advances in Flow Research*, pages 139–164. Springer, 2012.

Kevin A Pelphrey, Noah J Sasson, J Steven Reznick, Gregory Paul, Barbara D Goldman, and Joseph Piven. Visual Scanning of Faces in Autism. *Journal of Autism and Developmental Disorders*, 32(4):249–261, 2002.

Michael Peters, Bruno Laeng, Kerry Latham, Marla Jackson, Raghad Zaiyouna, and Chris Richardson. A redrawn Vandenberg and Kuse Mental Rotations Test – Different Versions and Factors that Affect Performance. *Brain and Cognition*, 28(1): 39–58, 1995.

Claus Pias. Computer Spiel Welten. 2002.

Karolien Poels, Yvonne de Kort, and Wijnand Ijsselsteijn. It Is Always a Lot of Fun!: Exploring Dimensions of Digital Game Experience Using Focus Group Methodology. In *Proceedings of the 2007 Conference on Future Play*, pages 83–89. ACM, 2007.

Frederico Poloni. Notes on the Bastet Algorithm, January 2012. URL http://fph.altervista.org/prog/bastetalgo.html.

Alex Poole and Linden J Ball. Eye Tracking in HCI and Usability Research. *Encyclopedia of Human Computer Interaction*, 1:211–219, 2006.

Marc Prensky. Computer Games and Learning: Digital Game-Based Learning. *Handbook of Computer Game Studies*, 18:97–122, 2005.

Eyal M Reingold, Neil Charness, Marc Pomplun, and Dave M Stampe. Visual Span in Expert Chess Players: Evidence from Eye Movements. *Psychological Science*, 12(1): 48–55, 2001.

Dietram A Scheufele. Framing as a Theory of Media Effects. *Journal of Communication*, 49(1):103–122, 1999.

Dietram A Scheufele and Shanto Iyengar. The State of Framing Research: A Call for New Directions. *The Oxford Handbook of Political Communication Theories. New York: Oxford UniversityPress*, 2012.

Elad Shahar and Ross West. Evolutionary AI for Tetris. 2010. URL http://www.cs.uml.edu/ecg/pub/uploads/AIfall10/eshahar_rwest_GATetris.pdf.

David Sheff. *Game Over: How Nintendo Zapped an American Industry, Captured Your Dollars, and Enslaved Your Children.* Random House Inc., 1993.

Herbert A Simon. Rational Choice and the Structure of the Environment. *Psychological Review*, 63(2):129, 1956.

Valerie K. Sims and Richard E. Mayer. Domain Specificity of Spatial Expertise: The Case of Video Game Players. *Applied Cognitive Psychology*, 16(1):97–115, 2002.

Robert N Singer, James H Cauraugh, Dapeng Chen, Gregg M Steinberg, and Shane G Frehlich. Visual Search, Anticipation, and Reactive Comparisons Between Highly-Skilled and Beginning Tennis Players. *Journal of Applied Sport Psychology*, 8 (1):9–26, 1996.

Mel Slater. Measuring Presence: A Response to the Witmer and Singer Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, 8(5):560–565, 1999.

Katharina Spiel. Out of Sight –Navigation and Immersion of Blind Players in Text-Based Games, 2012.

Wynn C Stirling and Michael A Goodrich. Satisficing Games. *Information Sciences*, 114(1):255–280, 1999.

Jari Takatalo. Presence and Flow in Virtual Environments: An Explorative Study. *Master's Thesis, University of Helsinki*, 2002.

Jari Takatalo, Jukka Häkkinen, Jeppe Komulainen, Heikki Särkelä, and Göte Nyman. Involvement and Presence in Digital Gaming. In *Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles*, pages 393–396. ACM, 2006.

Jean Underwood. Novice and Expert Performance with a Dynamic Control Task: Scanpaths During a Computer Game. *Cognitive processes in eye guidance*, pages 303–323, 2005.

Peter Vorderer, Werner Wirth, Feliz R Gouveia, F Biocca, T Saari, F Jäncke, S Böcking, H Schramm, A Gysbers, T Hartmann, et al. MEC Spatial Presence Questionnaire (MEC-SPQ): Short Documentation and Instructions for Application. *Report to the European Community, Project Presence: MEC (IST-2001-37661)*, 2004.

Nicholas Wade and Benjamin Tatler. *The Moving Tablet of the Eye: The Origins of Modern Eye Movement Research.* Oxford University Press, 2005.

Nicholas J Wade. Porterfield and Wells on the Motions of Our Eyes. *PERCEPTION-LONDON-*, 29(2):221–240, 2000.

Stefanie Wetzel. Can Eye Movements Reveal Cognitive Load? – Parallel Coordinates - Establishing a Cognitive Load Measure for Information Visualizations. Master's thesis, Bauhaus-Universität Weimar, 2014.

Stefanie Wetzel, Katharina Spiel, and Sven Bertel. Dynamically Adapting an AI Game Engine Based on Players' Eye Movements and Strategies. In *Proceedings of the 6th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, EICS '14, pages 3–12, New York, NY, USA, 2014. ACM.

Bob G Witmer and Michael J Singer. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3): 225–240, 1998.

WH Zangemeister, Keith Sherman, and Lawrence Stark. Evidence for a Global Scanpath Strategy in Viewing Abstract Compared With Realistic Images. *Neuropsychologia*, 33(8):1009–1025, 1995.

List of Figures

2.1 Skill Acquisition 6
2.2 Epistemic Actions 7
2.3 Looking at TETRIS 8

4.1 TETRIS Basics 23
4.2 Parameters for TETRIS Analysis 24
4.3 Parameters for TETRIS Analysis II 25
4.4 PYTRIS 30
4.5 Adaptation to Pile Height 32
4.6 Areas of Interest in EMTRIS 34
4.7 Adaption of Algorithms in EMTRIS 35
4.8 Adaption to Out-of-Interest Fixations 35
4.9 Adaption of Speed in EMTRIS 36

5.1 Perceived Difficulty and Fun for Different Types of Algorithms 41

6.1 Q-Q Plot of Age Variable 52
6.2 Dunning-Kruger Effect with TETRIS Expertise 55
6.3 Effect of Type of Adaptation on Speed and Algorithms 57
6.4 Challenge for Framed Players in Conventionally Adapted Games 61
6.5 OOI-Fixations for Framed Players in Conventionally Adapted Games . . . 64
6.6 Selected Correlations Between Lines Made and GEQ Values 66
6.7 Selected Correlations Between Lines Made and Expertise Measures 67
6.8 Selected Correlations Between GEQ Values and Expertise Measures . . . 68

7.1 Selected Correlations for Mental Rotations 75

8.1 Designing Games with Psychophysical Adaptation 79

List of Tables

2.1 Interpretation of Eye Movements 14

4.1 Adaption to Bumpiness 32
4.2 Adaption to Transition Value 33

5.1 Lines Cleared According to Algorithms in Pre-Study 40
5.2 Analysis of Closed Holes in Pre-Study 42
5.3 Analysis of Pile Height in Pre-Study 42
5.4 Mean Bumpiness According to Algorithms in Pre-Study 43
5.5 Classes of Test Subjects 44
5.6 Establishing Eye Movement Interaction 47

6.1 Determination of Effect Sizes 53
6.2 Summary of Performance Based Hypotheses 59
6.3 Summary of Gameplay Experience Based Hypotheses 62
6.4 Summary of Expertise Based Hypotheses 65
6.5 Summary of Intersectional Hypotheses 69

B.1 Participants of Pre-Study 105
B.2 Analysis of Maximum Speed in Pre-Study 106
B.3 Analysis of Time Used per Game in Pre-Study 106
B.4 Perceived Difficulty of Algorithms in Pre-Study 107
B.5 Perceived Fun of Algorithms in Pre-Study 107
B.6 Analysis Correlation between Closed Holes and Pile Height in Pre-Study . 107
B.7 Participants of Pilot Study 108
B.8 Performance in Skill Tests of Pilot Study 108
B.9 Participants of Main Study 109
B.10 Game Experience of Participants in Main Study 110
B.11 H-P1$_{SH}$(1) 111
B.12 H-P1$_{SH}$(2) 111
B.13 H-P1$_{SH}$(3) 111
B.14 H-P1$_{SH}$(4) 112
B.15 H-P2$_{SH}$(1) 112
B.16 H-P2$_{SH}$(2) 112
B.17 H-P2$_{SH}$(3) 112
B.18 H-P2$_{SH}$(4) 113
B.19 H-P3$_{SH}$(1) 113
B.20 H-P3$_{SH}$(2) 113
B.21 H-P3$_{SH}$(3) 113

B.22 H-P3$_{SH}$(4) . . . 114
B.23 H-P3 for Games with Eye Movement Based Adaptation . . . 114
B.24 H-P5$_{SH}$(1) . . . 115
B.25 H-P5$_{SH}$(2) . . . 115
B.26 H-P5$_{SH}$(3) . . . 115
B.27 H-P5$_{SH}$(4) . . . 115
B.28 H-U1$_{SH}$ . . . 116
B.29 H-U2$_{SH}$ . . . 116
B.30 H-U3$_{SH}$ . . . 117
B.31 H-U4$_{SH}$ . . . 117
B.32 H-U5$_{SH}$ . . . 118
B.33 H-E1$_{SH}$ . . . 119
B.34 H-E2$_{SH}$ . . . 119
B.35 H-E3$_{SH}$ . . . 120
B.36 H-E4$_{SH}$ . . . 120
B.37 H-E5$_{SH}$ . . . 120
B.38 H-I1$_{SH}$ . . . 121
B.39 H-I2$_{SH}$ . . . 121
B.40 H-I3$_{SH}$ . . . 122
B.41 Gender and Performance . . . 123
B.42 Gender and Gameplay Experience . . . 123
B.43 Gender and Expert Behaviour . . . 124
B.44 Age and Performance . . . 125
B.45 Age and Game Experience . . . 125
B.46 Age and Expert Behaviour . . . 125
B.47 Language and Performance . . . 126
B.48 Language and Gameplay Experience . . . 126
B.49 Language and Expert Behaviour . . . 127
B.50 Mental Rotation and Performance . . . 128
B.51 Mental Rotation and Game Experience . . . 128
B.52 Mental Rotation and Expert Behaviour . . . 128
B.53 Self Reported Expertise and Performance . . . 128
B.54 Self Reported Expertise and Game Experience . . . 129
B.55 Self Reported Expertise and Expert Behaviour . . . 129
B.56 PYTRIS Established Expertise and Performance . . . 129
B.57 PYTRIS Established Expertise and Game Experience . . . 129
B.58 PYTRIS Established Expertise and Expert Behaviour . . . 129

Acknowledgements

This work would not have been possible without the big help of Kearsley Schieder as a constant supply of constructive criticism at all stages.

Furthermore, I have to – again – thank Prof. Sven Bertel for his great supervision. Without his constant feedback, I wouldn't have been pushed to the point of being able to deliver research at the standard it is at now. Additional thanks go to Prof. Anke Huckauf, who was willing to give an external review of this work, making the quality of judgement so much higher.

Additionally thanks go to all the critical SAFT participants, who asked interested and helpful questions that lead to direct and indirect improvements of the research presented here.

Without all the test participants in the several studies, this work would also have been impossible. I would be nothing without their time donation and their willingness to put themselves in my hands for about one and a half hours, sometimes even forced to play horribly unpleasant and frankly occasionally unplayable versions of my game. Thanks for doing that for me!

I am particularly grateful for all the proof reading that has been done by Jennifer Beckmann, Kearsley Schieder, Julia Spiel and Stefanie Wetzel. They found the nagging bits and pieces and were critical and encouraging at the right places.

Thanks go also out to the Heinrich-Böll-Stiftung who enabled me to pursue my studies the way I wanted in the first place by providing me both financial security and guidance.

And last but not least, I also have to thank Q for sitting on my laptop occasionally and forcing me to take a break, when I needed it the most.

Appendix A

Questionnaires and Test Material

A.1 Pre-Test Questionnaire

Note: This questionnaire is only available in German, since all test participants were fluent German speakers.

Statistical Data and Tetris Expertise

- Alter ______________________
- Geschlecht ______________________
- Beruf/Ausbildungsrichtung

  ____
- Digitale Spielerfahrung
  - Spiele generell ______________________
  - Spiele regelmäßig ______________________
- Kennst Du TETRIS ja○ nein○
- Wenn ja, wann hast Du zuletzt TETRIS gespielt?

  ____
- Wie kompetent glaubst du, bist du in TETRIS?

  sehr kompetent ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ohne Kompetenz || k.A.○

After the First Game

- Algorithmus ____________________
- SpielID ____________________
- Wie schwierig fandest Du das Spiel?

  sehr schwierig ○ ○ ○ ○ ○ ○ sehr einfach || keine Antwort○

- Wie viel Spaß hat Dir das Spiel gemacht?

  sehr viel ○ ○ ○ ○ ○ ○ absolut keinen || keine Antwort○

After Game 2-10

- Algorithmus ____________________
- SpielID ____________________
- Wie schwierig fandest Du das Spiel?

  sehr schwierig ○ ○ ○ ○ ○ ○ sehr einfach || keine Antwort○

- Ich fand das Spiel schwieriger || leichter als das voherige.
- Wie viel Spaß hat Dir das Spiel gemacht?

  sehr viel ○ ○ ○ ○ ○ ○ absolut keinen || keine Antwort○

- Das Spiel hat mir mehr || weniger Spaß gemacht als das vorherige.

Vielen Dank für Deine Mithilfe.

A.2 Consent Form Main Study

Thank you for participating in this test. You can stop the game as well as the whole test at any time. (It does actually make sense to still fill out the questionnaires afterwards, in case you didn't finish all games.) If you have any questions during the test, you can always ask your supervisor; however, they might not always give you an answer. If you experience problems, please contact them immediately.

Your data will be dealt with confidentially. Your name will not be recorded at all, we use an encoded identifier for the questionnaires. The data will only be used for research purposes.

Next to the questionnaires, we will also log your test session in terms of interaction as well as your eye movements for analysis. If you disagree on any of these procedures – also at a later point–, please inform your test supervisor.

We hope you'll have fun!

Please indicate your understanding and agreement to the terms of the test here.

Date, Signature ______________________________

A.3 Initial Questionnaire for Statistical Data

- Age ____________________
- Gender ____________________
- Work/Education science○ social science○ humanities○ other○
- Are you aware of any defects in your vision? yes○ no○ n.a. ○

  If yes, please specify ____________________
- Experience with Digital Games
  - Genre ____________________

    How often?

    less than once a month○ more than once a month○
    more than once a week○ daily or near-daily○

    Competence

    very competent○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ no competence —— no answer○
  - Genre ____________________

    How often?

    less than once a month○ more than once a month○
    more than once a week○ daily or near-daily○

    Competence

    very competent○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ no competence —— no answer○

- Do you know of TETRIS yes○ no○
- If yes, when did you last play TETRIS?

____

- How would you rate your competence in TETRIS?

  very competent○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ no competence —— no answer○

A.4 Framing Texts

A.4.1 Framed Participants

You will play two games of TETRIS (each at a maximum of five minutes) and, after a break, two more games. This version of TETRIS is special because the game alters its difficulty based on what is being done by a player as well as their eye movements. If you are interested in more specifics, ask your supervisor after the test.

A.4.2 Non-Framed Participants

You will play two games of TETRIS (each at a maximum of five minutes) and, after a break, two more games. Because your eye movements can tell us about your performance, we record them to see whether the mechanisms that look at what players do and adapt the game are working well and improving your play. If you are interested in more specifics, ask your supervisor after the test.

A.5 Data Control Sheet

Testperson ID ____________________

Sprache: Deutsch○ Englisch○

Bevor Testperson kommt

1. ID festgelegt○
2. ID in Skilltest definiert○
3. ID in NEMtris definiert○
4. Log Files vorbereitet○
5. Reihenfolge der Spiele festgelegt○ ____________
6. Reihenfolge der Spiele in NEMtris festgelegt○
7. Eyetracker gereinigt○

BITTE WENDEN.

Nach dem Test

1. Archivieren der Skilltest Logs○
2. Archivieren der NEMtris Logs○
3. Digitalisierung der Fragebögen○
4. Digitalisierung der Spatial Ability Tests○
5. Schreddern der Spatial Ability Tests○

Mit Testperson

1. trägt Brille? ja○ nein○
2. Begrüßt und Aufgeklärt (Willkommensblatt unterschrieben)○
3. Statistischen Fragebogen ausgefüllt○
4. Skill Test durchgeführt○
5. Einführung (N)EMtris gemacht○
6. dominantes Auge ____________________
7. EyeTracking Setup ____________________
8. Validierungsergebnis ____________________
9. Spiele mit Pause durchgeführt (viermal plus Zwischenfragebögen)○
10. Game Experience Questionnaire ausgefüllt○
11. Kuchen und Pause angeboten und durchgeführt○
12. Post Game Questionnaire ausgefüllt○
13. MRT gemacht○
14. Bedankt!○

Raum für Notizen:

Appendix B

Additional Evaluations of Studies

B.1 Pre-Study

B.1.1 Test Participants

	Age	Gender	Occupation	TETRIS?	How long ago?	Comp.
	34	m	tech	yes	$>$ 12 months	4
	29	m	tech	yes	$>$ 12 months	7
	25	m	tech	yes	$>$ 24 months	7
	25	m	non-tech	yes	$>$ 6 months	4
	30	m	tech	yes	$>$ 12 months	4
	22	f	tech	yes	$>$ 12 months	3
	25	m	non-tech	yes	$>$ 12 months	3
	24	f	non-tech	yes	$>$ 24 months	7
	26	f	non-tech	yes	$>$ 24 months	6
	22	f	tech	yes	$>$ 2 months	8
	23	f	tech	yes	$>$ 84 months	9
	26	m	tech	yes	$>$ 12 months	5
	25	f	non-tech	yes	$>$ 6 months	9
	27	f	tech	yes	$>$ 24 months	n.a.
	26	f	tech	yes	$>$ 120 months	4
	27	f	tech	yes	$>$ 120 months	8
mean	26	8 f	11 tech	100%	$>$ 31.625 months	5.867
median	25.5	8 m	5 non-tech		$>$ 12 months	6
std	3.055				$>$ 39.241 months	2.134

TABLE B.1: Description of All Participants of the Pre-Study.

B.1.2 Performance-Based Results

Differences between algorithms for lines made:

- between all: $\chi^2 = 488.6633$, df = 124, p-value < 0.001
- nice/grab: $\chi^2 = 64.6026$, df = 31, p-value < 0.001
- grab/true: $\chi^2 = 167.601$, df = 31, p-value < 0.001
- true/skew: $\chi^2 = 171.0318$, df = 31, p-value < 0.001
- skew/bust: $\chi^2 = 116.5952$, df = 31, p-value < 0.001

Algorithm	Mean	Median	sd	Significance(p)	Corr.	Rank
Nicetris	228.1	240	67.14	< 0.001	-1.00	1
Grab Bag	229.1	240	71.77	< 0.001	-1.00	2
True Random	240.0	225	74.66	< 0.001	-1.00	3
Skewed Random	291.9	302.5	64.13	< 0.001	-1.00	4
Bust Head	333.1	347.5	46.83	< 0.001	-1.00	5
Over all	264.4	265	77.06	< 0.001	-1.00	

Table B.2: Maximum Speed per Game According to Algorithms in Pre-Study. Significance has been done using the Mann-Whitney-Wilcoxon test. Correlation means the correlation of the maximum speed with lines made calculated with Spearman's ρ. Rank (according to mean) goes from highest speed (1) to lowest speed (5). The scales on speed are inverted, because they are expressed inverted as time intervals in milliseconds, hence the correlation is always negative.

Algorithm	Mean	Median	sd	Significance(p)	Corr.	Rank
Nicetris	276.6	285	33.46	< 0.001	0.27	1
Grab Bag	272.8	298.5	45.14	< 0.001	0.24	2
True Random	258.0	289	58.39	< 0.001	0.55	3
Skewed Random	212.1	211.5	62.27	< 0.001	0.72	4
Bust Head	170.3	167.5	58.81	< 0.001	0.65	5
Over all	238.0	259.5	66.27	< 0.001	-0.09	

Table B.3: Time Used (in seconds) per Game According to Algorithms in Pre-Study. A possible maximum lies at 300 seconds plus time to drop the last tetromino. Significance has been done using the Mann-Whitney-Wilcoxon test. Correlation means the correlation of the time needed with lines made calculated with Spearman's ρ. Rank (according to mean) goes from most time used (1) to lowest time used (5).

Algorithm	Mean	Median	sd	Significance(p)	Rank
Nicetris	3.094	3	1.058	< 0.001	4
Grab Bag	2.719	2.5	1.085	< 0.001	5
True Random	3.344	3	1.234	< 0.001	3
Skewed Random	4.156	4	1.051	< 0.001	2
Bust Head	4.562	5	1.076	< 0.001	1

Table B.4: Perceived Difficulty of Algorithms in Pre-Study. Likert Scale had six item and was encoded on the [1..6] range. Significance has been done using the Mann-Whitney-Wilcoxon test. Rank (according to mean) goes from hardest(1) to easiest(5).

Algorithm	Mean	Median	sd	Significance(p)	Rank
Nicetris	4.469	4.5	1.164	< 0.001	2
Grab Bag	4.469	5	1.270	< 0.001	1
True Random	4.281	4	1.023	< 0.001	3
Skewed Random	3.938	4	1.268	< 0.001	4
Bust Head	3.469	3	1.270	< 0.001	5

Table B.5: Perceived Fun of Algorithms in Pre-Study. Likert Scale had six item and was encoded on the [1..6] range. Significance has been done using the Mann-Whitney-Wilcoxon test. Rank (according to mean, then median) goes from most fun(1) to least fun(5).

B.1.3 Questionnaire-Based Results

B.1.4 Other Results

Algorithm	Spearman's ρ	p	S
Nicetris	0.78	< 0.001	1187.43
Grab Bag	0.76	< 0.001	1331.61
True Random	0.41	0.004	2771.95
Skewed Random	0.39	0.029	3343.27
Bust Head	0.29	0.106	3865.74
Over all	0.61	< 0.001	267009

Table B.6: Analysis Correlation between Closed Holes and Pile Height in Pre-Study.

B.2 Pilot-Study

B.2.1 Test Participants

	Age	Gender	Occupation	TETRIS?	How long ago?	Comp.
	26	m	tech	yes	> 1 months	4
	18	m	other	yes	> 12 months	4
	22	f	tech	yes	> 1 month	5
	27	f	tech	yes	> 12 months	8
	23	m	tech	yes	> 1 month	9
	27	m	tech	yes	> 12 months	6
mean	23.833	2 f	5 tech	100%	> 6.5 months	6
median	24.5	4 m	1 non-tech		> 6.5 months	5.5
std	3.545				> 6.025 months	2.098

TABLE B.7: Description of All Participants of the Pilot Study.

B.2.2 Skill Test Results

	# of games	best	worst	average	best/min	worst/min	average/min
	3	20	18	18.67	13.16	11.92	12.36
	3	22	13	17.67	14.47	8.61	11.67
	3	25	12	19.67	16.56	7.95	13.06
	3	26	23	24.00	17.11	15.13	15.82
	5	22	15	18.60	14.47	9.87	12.22
	5	20	14	16.40	13.25	9.27	11.30
mean	3.667	22.50	15.83	19.17	14.84	10.46	12.74
median	3	22	14.5	18.63	14.47	9.57	12.29
std	1.03	2.51	4.07	2.61	1.66	2.66	1.63

TABLE B.8: Performance in the skill tests during the pilot study. One row represents one participant if not indicated otherwise.

B.3 Main Study

B.3.1 Population

	group	language	age	gender	occupation	mr score
NF1	C	German	26	f	other	13
F2	F	German	30	f	science	14
NF3	C	German	27	m	science	8
F4	F	German	25	m	other	16
NF5	C	German	25	m	science	11
F6	F	German	25	f	science	15
NF7	C	German	32	m	science	7
F8	F	German	26	m	science	10
NF9	C	English	30	f	humanities	12
F10	F	German	33	m	other	17
NF11	C	German	25	f	science	7
F12	F	German	22	m	science	11
NF13	C	German	25	m	science	22
F14	F	German	26	m	other	12
F15	F	German	14	f	other	3
NF16	C	English	24	f	science	7
NF17	C	English	26	m	science	20
F18	F	English	34	none	other	19
NF19	C	German	46	f	humanities	11
F20	F	German	30	m	science	9
NF21	C	German	24	f	other	13
F22	F	German	39	m	science	9
NF23	C	German	22	f	science	18
F24	F	English	28	f	science	12
NF25	C	English	25	m	science	14
F26	F	German	28	m	humanities	10
F27	F	German	27	f	science	12
NF28	C	German	25	m	science	21
NF29	C	German	27	m	science	9
F30	F	German	32	none	other	7
F31	F	German	19	m	science	14
NF32	C	German	24	f	science	16
NF33	C	German	25	m	science	14
F34	F	German	29	m	science	16
NF35	C	German	24	f	science	14
F36	F	German	24	m	science	16
NF37	C	English	30	f	social science	4
F38	F	German	26	m	science	12
NF39	C	German	34	m	science	13
F40	F	German	33	m	social science	10
NF41	C	German	33	m	social science	5
F42	F	German	30	f	social science	14
F43	F	German	29	f	humanities	7
mean	21 C	36 German	27.63	24 m	27 science	12.19
median	22 F	7 English	26	17 f	4 hum/soc sc.	12
std			5.26	2 none	8 other	4.45

TABLE B.9: Demographic Description of Participants of the Main Study. C refers to non framed participants, F refers to framed participants, mr stands for mental rotation.

	regularity dg	comp dg	knows t	when last played t	comp t
NF1	less than once a month	9	1	about 15 years ago	4
F2	less than once a month	4	1	about 10 years ago	4
NF3	more than once a month	4	1	about 4 months ago	7
F4	less than once a month	4	1	pre-study	5
NF5	more than once a week	4.5	1	about 2 years ago	8
F6	more than once a month	2	1	about 2 years ago	4
NF7	more than once a week	4	1	about 5 years ago	4
F8	less than once a month	7	1	pre-study	9
NF9	less than once a month	10	1	about 20 years ago	4
F10	less than once a month	7	1	about 3 years ago	7
NF11	less than once a month	10	1	never	0
F12	more than once a month	2	1	about 5 years ago	5
NF13	more than once a month	2	1	about 3 years ago	5
F14	more than once a week	2	1	about 10 years ago	8
F15	less than once a month	4	1	about 2 years ago	7
NF16	more than once a month	3	1	about 10 years ago	4
NF17	less than once a month	5	1	about 4 years ago	5
F18	more than once a week	4	1	pre-study	3
NF19	less than once a month	10	1	about 20 years ago	8
F20	more than once a week	2	1	about 4 months ago	4
NF21	more than once a month	4	1	pre-study	4
F22	more than once a week	1	1	about 6 months ago	4
NF23	more than once a month	4	1	about 3 weeks ago	4
F24	less than once a month	9	1	pre study	9
NF25	less than once a month	6	1	about 15 years ago	4
F26	more than once a month	4	1	about 5 years ago	4
F27	less than once a month	2	1	pre-study	3
NF28	daily or nearly daily	4	1	about 8 years ago	7
NF29	less than once a month	3	1	last month	5
F30	daily or nearly daily	7	1	about 6 months ago	6
F31	less than once a month	7	1	about 15 years ago	10
NF32	less than once a month	6	1	never	0
NF33	less than once a month	6	1	about 10 years ago	9
F34	less than once a month	3	1	about 4 years ago	6
NF35	more than once a month	5	1	about 1 year ago	6
F36	more than once a week	3	1	2 days ago	4
NF37	less than once a month	10	1	about 3 years ago	7
F38	more than once a week	3	1	about 1 year ago	5
NF39	less than once a month	0	1	about 15 years ago	9
F40	less than once a month	7	1	about 5 years ago	6
NF41	more than once a week	7	1	about 15 years ago	7
F42	more than once a week	3	1	about 2 months ago	4
F43	less than once a month	9	1	about 15 years ago	9
mean		4.94	1		5.51
median		4	1		5
std		2.59	0		2.29

TABLE B.10: Game Experience of Participants in Main Study. comp refers to competence, dg refers to digital games in general, t refers to TETRIS

B.3.2 Results According to Hypotheses

B.3.2.1 Performance Based Hypotheses

	Non-Framed	Framed
shapiro	$W = 0.980, p = 0.915$	$W = 0.928, p = 0.113$
mean	14.36	12.88
median	14	10.875
standard deviation	4.88	7.09
significance	$t = 0.872, df = 37.25, p = 0.389$	

TABLE B.11: Overview of Results for H-P1$_{SH}$(1) - Lines Made According to Framing Category. Significance has been established using Student's t-test.

	Non-Framed	Framed
shapiro	$W = 0.951, p = 0.330$	$W = 0.921, p = 0.081$
mean	18.11	16.57
median	16.25	15.50
standard deviation	6.43	9.07
significance	$t = 0.650, df = 37.85, p = 0.520$	

TABLE B.12: Overview of Results for H-P1$_{SH}$(2) - Score According to Framing Category. Significance has been established using Student's t-test.

	Non-Framed	Framed
shapiro	$W = 0.935, p = 0.154$	$W = 0.908, p < 0.05$
mean	473.1ms	498.5ms
median	471.6ms	480.2ms
standard deviation	78.37ms	65.29ms
significance	$W = 202, p = 0.354$	

TABLE B.13: Overview of Results for H-P1$_{SH}$(3) - Speed at the End of a Game According to Framing Category. Significance has been established using the Wilcoxon-Mann-Whitney test. Note that a high number refers to a larger interval between refreshing steps of the tetromino and so indicates a slower game.

	Non-Framed	Framed
shapiro	$W = 0.932, p = 0.135$	$W = 0.935, p = 0.153$
mean	1.57	1.51
median	1.5	1.5
standard deviation	0.42	0.37
significance	$t = 0.503, df = 41.13, p = 0.617$	

TABLE B.14: Overview of Results for H-P1$_{SH}$(4) - Algorithm at the End of a Game According to Framing Category. Significance has been established using Student's t-test. Algorithms have been encoded from 1 = GRAB BAG to 5 = BUST HEAD according to previously established ranks of difficulty (see Figure 5.1).

	Conventional	Eye Movement Based
shapiro	$W = 0.866, p < 0.001$	$W = 0.937, p < 0.05$
mean	13.66	13.65
median	11	12
standard deviation	9.25	8.60
significance	$W = 904, p = 0.863$	

TABLE B.15: Overview of Results for H-P2$_{SH}$(1) - Lines Made According to Adaption. Significance has been established using the Wilcoxon-Mann-Whitney test.

	Conventonal	Eye Movement Based
shapiro	$W = 0.864, p < 0.001$	$W = 0.947, p < 0.05$
mean	17.69	16.95
median	14.5	15
standard deviation	12.81	10.31
significance	$W = 910, p = 0.904$	

TABLE B.16: Overview of Results for H-P2$_{SH}$(2) - Score According to Adaption. Significance has been established using the Wilcoxon-Mann-Whitney test.

	Conventional	Eye Movement Based
shapiro	$W = 0.4512, p < 0.001$	$W = 0.974, p = 0.434$
mean	385.7ms	586.5ms
median	372ms	582ms
standard deviation	55.73ms	132.90ms
significance	$W = 148.5, p < 0.001$	

TABLE B.17: Overview of Results for H-P2$_{SH}$(3) - Speed at the End of a Game According to Adaption. Significance has been established using the Wilcoxon-Mann-Whitney test. Note that a high number refers to a larger interval between refreshing steps of the tetromino and so indicates a slower game.

	Conventional	Eye Movement Based
shapiro	$W = 0.746, p < 0.001$	$W = 0.891, p < 0.001$
mean	1.37	1.71
median	1	1.5
standard deviation	0.50	0.59
significance	$W = 513.5, p < 0.01$	

TABLE B.18: Overview of Results for H-P2$_{SH}$(4) - Algorithm at the End of a Game According to Adaption. Significance has been established using the Wilcoxon-Mann-Whitney test. Algorithms have been encoded from 1 = GRAB BAG to 5 = BUST HEAD according to previously established ranks of difficulty (see Figure 5.1).

	Non-Framed	Framed
shapiro	$W = 0.828, p =< 0.01$	$W = 0.895, p < 0.05$
mean	15.88	11.55
median	13	10
standard deviation	10.46	7.56
significance	$W = 168.5, p = 0.132$	

TABLE B.19: Overview of Results for H-P3$_{SH}$(1) - Lines Made According to Framing Category for Conventionally Adapted Games. Significance has been established using the Wilcoxon-Mann-Whitney test.

	Non-Framed	Framed
shapiro	$W = 0.817, p < 0.01$	$W = 0.910, p < 0.05$
mean	20.43	15.07
median	16	13.25
standard deviation	14.69	10.40
significance	$W = 281, p = 0.229$	

TABLE B.20: Overview of Results for H-P3$_{SH}$(2) - Score According to Framing Category in Conventionally Adapted Games. Significance has been established using the Wilcoxon-Mann-Whitney test.

	Non-Framed	Framed
shapiro	$W = 0.491, p < 0.001$	$W = 0.496, p < 0.001$
mean	396.7ms	375.3ms
median	373ms	369ms
standard deviation	73.07ms	29.88ms
significance	$W = 303, p = 0.081$	

TABLE B.21: Overview of Results for H-P3$_{SH}$(3) - Speed at the End of a Game According to Framing Category for Conventionally Adapted Games. Significance has been established using the Wilcoxon-Mann-Whitney test. Note that a high number refers to a larger interval between refreshing steps of the tetromino and so indicates a slower game.

	Non-Framed	Framed
shapiro	$W = 0.747, p < 0.001$	$W = 0.739, p < 0.001$
mean	1.43	1.32
median	1	1
standard deviation	0.60	0.39
significance	$W = 241, p = 0.798$	

TABLE B.22: Overview of Results for H-P3$_{\mathrm{SH}}$(4) - Algorithm at the End of a Game According to Framing Category for Conventionally Adapted Games. Significance has been established using the Wilcoxon-Mann-whitney test . Algorithms have been encoded from 1 = GRAB BAG to 5 = BUST HEAD according to previously established ranks of difficulty (see Figure 5.1).

	Non-Framed	Framed
Lines Made		
shapiro	$W = 0.878, p < 0.05$	$W = 0.953, p = 0.364$
mean	13.07	14.20
median	12	12.25
standard deviation	7.98	9.31
significance	$W = 221, p = 0.817$	
Scores		
shapiro	$W = 0.898, p < 0.05$	$W = 0.961, p = 0.503$
mean	15.79	18.070
median	14.5	16.75
standard deviation	9.27	11.31
significance	$W = 209.5, p = 0.6098$	
Speed at the End		
shapiro	$W = 0.970, p = 0.73$	$W = 0.9166, p = 0.065$
mean	549.6ms	621.7ms
median	564ms	589ms
standard deviation	123.30	134.86
significance	$t = -1.83, df = 40.93, p = 0.074$	
Algorithm at the End		
shapiro	$W = 0.878, p < 0.05$	$W = 0.877, p < 0.05$
mean	1.71	1.71
median	1.5	1
standard deviation	0.54	0.64
significance	$W = 240.5, p = 0.821$	

TABLE B.23: Overview of Results for H-P3 for Games with Eye Movement Based Adaptation.

	Non-Framed	Framed
shapiro	$W = 0.797, p =< 0.01$	$W = 0.838, p < 0.01$
mean	10.43	6.84
median	6	4.5
standard deviation	11.84	6.67
significance	$W = 208.5, p = 0.593$	

TABLE B.24: Overview of Results for H-P5$_{SH}$(1) - Differences in Lines Made According to Framing Category. Significance has been established using the Wilcoxon-Mann-Whitney test.

	Non-Framed	Framed
shapiro	$W = 0.798, p < 0.001$	$W = 0.808, p < 0.001$
mean	13.93	9.00
median	9.5	7.25
standard deviation	15.77	8.26
significance	$W = 204.5, p = 0.5271$	

TABLE B.25: Overview of Results for H-P5$_{SH}$(2) - Differences in Score According to Framing Category. Significance has been established using the Wilcoxon-Mann-Whitney test.

	Non-Framed	Framed
shapiro	$W = 0.954, p = 0.412$	$W = 0.924, p = 0.090$
mean	164.1ms	251.5ms
median	159ms	216ms
standard deviation	107.46ms	135.67ms
significance	$t = 2.35, df = 39.68, p - value < 0.05$	

TABLE B.26: Overview of Results for H-P5$_{SH}$(3) - Differences in Speed at the End of a Game According to Framing Category. Significance has been established using Student's t-test. Note that a high number refers to a larger interval between refreshing steps of the tetromino and so indicates a slower game.

	Non-Framed	Framed
shapiro	$W = 0.852, p < 0.01$	$W = 0.861, p < 0.01$
mean	0.57	0.61
median	0.5	0.5
standard deviation	0.53	0.62
significance	$W = 234.5, p = 0.940$	

TABLE B.27: Overview of Results for H-P5$_{SH}$(4) - Differences in Algorithm at the End of a Game According to Framing Category. Significance has been established using the Wilcoxon-Mann-whitney test . Algorithms have been encoded from 1 = GRAB BAG to 5 = BUST HEAD according to previously established ranks of difficulty (see Figure 5.1).

B.3.2.2 User Experience Based Hypotheses

Category	Aspect	shapiro p	mean	median	sd	sig N/F
Non-Framed	competence	$p = 0.105$	2.53	2.6	0.85	$p = 0.8274$ (t)
	flow	$p = 0.441$	2.95	3	0.82	$p = 0.410$ (t)
	challenge	$p = 0.029$	2.22	2	0.73	$p = 0.074$ (w)
	positive	$p = 0.493$	3.16	3.4	0.93	$p = 0.99$ (t)
	immersion	$p = 0.131$	2.18	2.2	0.87	$p = 0.59$ (t)
	post positive	$p < 0.01$	1.91	1.8	0.66	$p = 0.874$ (w)
	tension	$p < 0.001$	1.62	1.33	0.69	$p = 0.618$ (w)
	negative	$p < 0.01$	1.66	1.5	0.64	$p = 0.851$ (w)
	post negative	$p < 0.01$	1.28	1.2	0.30	$p = 0.891$ (w)
Framed	competence	$p = 0.814$	2.59	2.5	0.80	$p = 0.8274$ (t)
	flow	$p = 0.938$	3.17	3	0.91	$p = 0.410$ (t)
	challenge	$p = 0.551$	2.6	2.6	0.75	$p = 0.074$ (w)
	positive	$p = 0.377$	3.17	3.1	0.69	$p = 0.99$ (t)
	immersion	$p = 0.057$	2.05	1.8	0.71	$p = 0.59$ (t)
	post positive	$p = 0.089$	1.89	1.83	0.58	$p = 0.874$ (w)
	tension	$p < 0.001$	1.83	1.67	1.02	$p = 0.618$ (w)
	negative	$p < 0.01$	1.51	1.38	0.36	$p = 0.851$ (w)
	post negative	$p < 0.001$	1.28	1.17	0.29	$p = 0.891$ (w)

Table B.28: Overview of Results for H-U1$_{SH}$ - Reported Values for the GEQ Core and Post Game Module According to Framing Category. Significance has been established using Student's t-test (t) or the Wilcoxon-Mann-Whitney test (w).

Category	Aspect	shapiro p	mean	median	sd	sig N/F
Conventional	competence	$p = 0.093$	2.28	2.25	0.87	$p = 0.382$ (t)
	flow	$p = 0.381$	2.85	2.75	0.99	$p = 0.732$ (t)
	challenge	$p = 0.230$	3.08	2.5	0.89	$p = 0.589$ (t)
	positive	$p = 0.490$	2.76	2.75	0.90	$p = 0.837$ (t)
	tension	$p < 0.05$	2.15	2	0.85	$p = 0.316$ (w)
	negative	$p < 0.001$	1.51	1.25	0.55	$p = 0.715$ (w)
EM-Based	competence	$p = 0.147$	2.43	2.5	0.71	$p = 0.382$ (t)
	flow	$p = 0.074$	2.93	3.25	1.05	$p = 0.732$ (t)
	challenge	$p = 0.132$	2.98	3	0.90	$p = 0.589$ (t)
	positive	$p = 0.268$	2.80	3	0.93	$p = 0.837$ (t)
	tension	$p < 0.001$	1.99	1.75	0.84	$p = 0.316$ (w)
	negative	$p < 0.001$	1.47	1.25	0.50	$p = 0.715$ (w)

Table B.29: Overview of Results for H-U2$_{SH}$ - Reported Values for the In-Game Module of the GEQ According to Type of Adaptation. Significance has been established using Student's t-test (t) or the Wilcoxon-Mann-Whitney test (w).

Measure	Framing	shapiro p	mean	median	sd	sig N/F
Flow	F N	$p = 0.947$ $p = 0.522$	2.97 2.74	3 2.75	1.02 0.97	$p = 0.456$ (t)
Challenge	F N	$p = 0.232$ $p = 0.562$	3.38 2.77	3 2.75	0.76 0.93	$p < 0.05$ (t)
Competence	F N	$p = 0.097$ $p = 0.250$	2.06 2.51	2.13 2.5	0.79 0.91	$p = 0.088$ (t)
Positive	F N	$p = 0.643$ $p = 0.821$	2.56 2.96	2.5 2.75	0.81 0.95	$p = 0.140$ (t)
Tension	F N	$p = 0.223$ $p = 0.184$	2.26 2.04	2.13 2	0.91 0.78	$p = 0.387$ (t)
Negative	F N	$p < 0.01$ $p < 0.01$	1.48 1.55	1.25 1.25	0.52 0.59	$p = 0.794$ (w)

TABLE B.30: Overview of Results for H-U3$_{\mathrm{SH}}$ - Reported Values for the In-Game Module of the GEQ for Conventionally Adapted Games According to Framing. Significance has been established using Student's t-test (t) or the Wilcoxon-Mann-Whitney test (w).

Measure	Adaptation	shapiro	mean	median	sd	sig N/F
Flow	C E	$p = 0.947$ $p = 0.691$	2.97 3.0	3 3.25	1.02 1.00	$p = 0.911$ (t)
Challenge	C E	$p = 0.232$ $p = 0.254$	3.38 3.17	3.25 2.13	0.76 0.89	$p = 0.151$ (t)
Competence	C E	$p = 0.097$ $p = 0.141$	2.06 2.55	2.13 2.5	0.79 0.67	$p < 0.05$ (t)
Positive	C E	$p = 0.643$ $p = 0.177$	2.56 2.85	2.5 3	0.81 0.77	$p = 0.221$ (t)
Tension	C E	$p = 0.223$ $p < 0.01$	2.26 2.07	2.13 1.875	0.91 0.87	$p = 0.422$ (w)
Negative	C E	$p < 0.01$ $p < 0.01$	1.48 1.43	1.25 1.38	0.52 0.41	$p = 0.952$ (w)

TABLE B.31: Overview of Results for H-U4$_{\mathrm{SH}}$ - Reported Values for the In-Game Module of the GEQ for Framed Players According to Type of Game. Significance has been established using Student's t-test (t) or the Wilcoxon-Mann-Whitney test (w).

Measure	Framing	shapiro	mean	median	sd	sig N/F
Flow	F	$p < 0.001$	0.33	0.25	0.41	$p = 0.262$ (w)
	N	$p < 0.01$	0.43	0.25	0.40	
Challenge	F	$p < 0.01$	0.43	0.5	0.25	$p = 0.336$ (w)
	N	$p < 0.01$	0.40	0.25	0.38	
Competence	F	$p = 0.076$	0.72	0.75	0.61	$p = 0.650$ (t)
	N	$p = 0.081$	0.80	0.75	0.56	
Positive	F	$p < 0.05$	0.75	0.63	0.5	$p = 0.588$ (w)
	N	$p = 0.055$	0.65	0.75	0.53	
Tension	F	$p < 0.01$	0.49	0.25	0.42	$p = 0.114$ (w)
	N	$p = 0.066$	0.74	0.75	0.53	
Negative	F	$p < 0.05$	0.36	0.38	0.26	$p = 0.296$ (w)
	N	$p < 0.01$	0.29	0.25	0.31	

TABLE B.32: Overview of Results for H-U5$_{\mathrm{SH}}$ - Differences of In-Game GEQ Values between Type of Adaptation According to Framing. Significance has been established using Student's t-test (t) or the Wilcoxon-Mann-Whitney test (w).

B.3.2.3 Expertise Based Hypotheses

Measure	Framing	shapiro	mean	median	sd	sig N/F
Fixations	N	$p = 0.226$	1378	1376	307.97	$p = 0.163$ (t)
	F	$p = 0.855$	1227	1260	386.74	
Fix. Length	N	$p = 0.261$	299.6ms	290.9ms	36.25ms	$p = 0.717$ (t)
	F	$p = 0.469$	303.6ms	273.2ms	37.20ms	
Horizontal EM	N	$p < 0.01$	0.153	0.123	0.073	$p = 0.420$ (w)
	F	$p = 0.237$	0.125	0.122	0.047	
Saccadic Ampl.	N	$p < 0.01$	285.4	278.4	98.44	$p = 0.838$ (w)
	F	$p < 0.01$	281.9	270.4	71.50	
Epistemic Actions	N	$p < 0.05$	0.142	0.120	0.095	$p = 0.508$ (w)
	F	$p < 0.05$	0.159	0.136	0.095	

Table B.33: Overview of Results for H-E1$_{SH}$ - Differences of Expertise Establishing Measures According to Framing. Significance has been established using Student's t-test (t) or the Wilcoxon-Mann-Whitney test (w).

Measure	Game	shapiro	mean	median	sd	sig N/F
Fixations	C	$p = 0.396$	1294	1310	497.51	$p = 0.966$ (w)
	E	$p < 0.05$	1308	1224	473.58	
Fix. Length	C	$p = 0.069$	302.2ms	292.8ms	40.04ms	$p = 0.886$ (t)
	E	$p = 0.283$	301.0ms	297.3ms	34.69ms	
Horizontal EM	C	$p < 0.001$	0.253	0.127	0.668	$p = 0.559$ (w)
	E	$p = 0.237$	0.151	0.124	0.085	
Saccadic Ampl.	C	$p < 0.001$	288	271.7	121.93	$p = 0.931$ (w)
	E	$p < 0.001$	279.2	258.3	90.53	
Epistemic Actions	C	$p < 0.001$	0.155	0.147	0.084	$p = 0.718$ (w)
	E	$p < 0.001$	0.194	0.137	0.197	

Table B.34: Overview of Results for H-E2$_{SH}$ - Differences of Expertise Establishing Measures According to Type of Adaptation. Significance has been established using the Wilcoxon-Mann-Whitney test (w).

Measure	Framing	shapiro	mean	median	sd	sig N/F
Fixations	N F	$p = 0.720$ $p = 0.840$	1362 1229	1332 1255	573.33 415.87	$p = 0.392$ (t)
Fix. Length	N F	$p = 0.162$ $p = 0.537$	300.2ms 304.1ms	287.9ms 296.1ms	42.16ms 38.80ms	$p = 0.749$ (t)
OOI-Fixations	N F	$p = 0.867$ $p = 0.932$	738.5 535.8	712.5 527	292.13 222.63	$p < 0.05$ (t)
Horizontal EM	N F	$p < 0.001$ $p = 0.657$	0.360 0.115	0.147 0.109	0.950 0.051	$p = 0.107$ (w)
Saccadic Ampl.	N F	$p < 0.001$ $p < 0.05$	304.8 272	274.5 269	156.35 76.83	$p = 0.914$ (w)
Epistemic Actions	N F	$p = 0.674$ $p = 0.116$	0.136 0.174	0.134 0.158	0.075 0.091	$p = 0.142$ (t)

Table B.35: Overview of Results for H-E3$_{SH}$ - Differences of Expertise Establishing Measures According to Framing for Conventionally Adapted Games. Significance has been established using Student's t-test (t) or the Wilcoxon-Mann-Whitney test (w).

Measure	Game	shapiro	mean	median	sd	sig N/F
Fixations	C E	$p = 0.840$ $p < 0.01$	1229 1225	1255 1215	415.87 530.88	$p = 0.844$ (w)
Fix. Length	C E	$p = 0.537$ $p = 0.402$	304.1ms 303.1ms	296.1ms 299.0ms	38.80ms 37.44ms	$p = 0.927$ (t)
Horizontal EM	C E	$p = 0.657$ $p < 0.05$	0.116 0.143	0.109 0.117	0.051 0.072	$p = 0.311$ (w)
Saccadic Ampl.	C E	$p < 0.05$ $p < 0.001$	272 291.8	269 257.8	76.83 108.45	$p = 0.807$ (w)
Epistemic Actions	C E	$p = 0.116$ $p < 0.001$	0.174 0.182	0.158 0.141	0.091 0.165	$p = 0.522$ (w)

Table B.36: Overview of Results for H-E4$_{SH}$ - Differences of Expertise Establishing Measures According to Type of Adaptation for Framed Players. Significance has been established using the Wilcoxon-Mann-Whitney test (w).

Measure	Framing	shapiro	mean	median	sd	sig N/F
Fixations	F N	$p < 0.001$ $p < 0.05$	408.9 563.5	350 413.5	368.92 514.82	$p = 0.528$ (w)
Fix. Length	N F	$p < 0.001$ $p =< 0.001$	13.33ms 11.65ms	6.95ms 7.72ms	14.38ms 11.73ms	$p = 0.952$ (w)
Horizontal EM	N F	$p < 0.001$ $p < 0.001$	0.052 0.304	0.032 0.119	0.065 0.928	$p = 0.157$ (w)
Saccadic Ampl.	N F	$p < 0.001$ $p < 0.001$	73.09 82.13	49.01 43.5	98.47 116.79	$p = 0.990$ (w)
Epistemic Actions	N F	$p < 0.001$ $p < 0.001$	0.101 0.122	0.070 0.068	0.107 0.176	$p = 0.782$ (w)

Table B.37: Overview of Results for H-E5$_{SH}$ - Differences of Expertise Establishing Measures of Deltas Between Type of Adaption for Framed Players. Significance has been established using the Wilcoxon-Mann-Whitney test (w).

B.3.2.4 Intersectional Hypotheses

Measure	shapiro	mean	median	sd	correlation to lines
Lines Made	$p = 0.292$	13.66	12.25	6.15	-
Flow	$p = 0.259$	3.07	3	0.87	$r = 0.230$
Challenge	$p = 0.064$	2.41	2.4	0.76	$r = -0.062$
Competence	$p = 0.077$	2.56	2.6	0.82	$r = 0.171$
Immersion	$p < 0.05$	2.11	2	0.79	$\rho = -0.077$
Positive	$p = 0.324$	3.16	3.2	0.80	$r = -0.041$
Tension	$p < 0.001$	1.73	1.33	0.87	$\rho = 0.006$
Negative	$p < 0.001$	1.58	1.5	0.52	$\rho = 0.208$
Post Positive	$p < 0.01$	1.90	1.83	0.61	$\rho = -0.065$
Post Negative	$p < 0.01$	1.28	1.17	0.29	$\rho = 0.106$

Table B.38: Overview of Results for H-I1$_{\mathrm{SH}}$ - Correlations Between Lines Made and GEQ Values. Correlation has been established using Pearson's product correlation (when r) or Spearman's ρ.

Measure	shapiro	mean	median	sd	correlation to lines
Fixations	$p = 0.954$	1301	1311	354.66	$r = 0.200$
Fix. Length	$p = 0.126$	301.6ms	294.2ms	36.36ms	$r = 0.038$
OOI-fixations	$p = 0.489$	634.8	613.5	275.56	$r = 0.258$
Horizontal	$p < 0.001$	0.193	0.130	0.337	$\rho = 0.232$
Saccadic Ampl.	$p < 0.001$	283.6	270.9	84.70	$\rho = 0.010$
Epistemic Actions	$p < 0.001$	0.175	0.145	0.122	$\rho = 0.306$

Table B.39: Overview of Results for H-I2$_{\mathrm{SH}}$ - Correlations Between Lines Made and Measured Eye Movements/Epistemic Actions. Correlation has been established using Pearson's product correlation (when r) or Spearman's ρ. Note that OOI-fixations are only looked at for conventionally adapted games.

GEQ	Fixations	Fix. Lengths	OOI	Horizontal	Saccades	Episteme
flow	$r = 0.202$	$r = -0.269$	$r = 0.211$	$\rho = -0.170$	$\rho = -0.075$	$\rho = 0.325$
challenge	$r = 0.114$	$r = 0.028$	$r = 0.086$	$\rho = -0.019$	$\rho = 0.062$	$\rho = 0.165$
competence	$r = -0.066$	$r = -0.014$	$r = -0.013$	$\rho = -0.005$	$\rho = -0.095$	$\rho = 0.044$
immersion	$\rho = -0.117$	$\rho = -0.156$	$\rho = 0.089$	$\rho = -0.193$	$\rho = -0.159$	$\rho = 0.028$
positive	$r = 0.168$	$r = 0.135$	$r = 0.098$	$\rho = -0.317$	$\rho = 0.013$	$\rho = 0.183$
tension	$\rho = -0.013$	$\rho = 0.110$	$\rho = 0.112$	$\rho = -0.121$	$\rho = -0.088$	$\rho = -0.014$
negative	$\rho = -0.157$	$\rho = 0.180$	$r = -0.136$	$\rho = 0.279$	$\rho = 0.090$	$\rho = -0.170$
post positive	$\rho = -0.118$	$\rho = -0.044$	$\rho = -0.015$	$\rho = -0.228$	$\rho = 0.017$	$\rho = 0.118$
post negative	$\rho = -0.144$	$\rho = -0.016$	$\rho = -0.029$	$\rho = 0.213$	$\rho = 0.044$	$\rho = 0.066$

TABLE B.40: Overview of Results for H-I3$_{SH}$ - Correlations Between GEQ Values and Measured Eye Movements/Epistemic Actions. Correlation has been established using Pearson's product correlation (when r) or Spearman's ρ. Note that OOI-fixations are only looked at for conventionally adapted games.

B.3.3 Side Effects

B.3.3.1 Gender

Measure	Gender	shapiro p	mean	median	sd	sig m/w
Lines Made	male female	$p < 0.01$ $p = 0.278$	13.51 13.51	10.75 15	6.71 5.23	$p = 0.508$ (w)
Scores	male female	$p < 0.05$ $p = 0.222$	17.1 17.09	15.5 18	8.79 6.29	$p = 0.404$ (w)
Speed	male female	$p = 0.332$ $p = 0.487$	485.7ms 468.2ms	471.8ms 474.5ms	71.85ms 67.12ms	$p = 0.428$ (t)
Algorithm	male female	$p = 0.11$ $p < 0.05$	1.57 1.44	1.5 1.5	0.43 0.31	$p = 0.324$ (w)

TABLE B.41: Overview over Relationships between Gender and Performance Measures

Measure	Gender	shapiro p	mean	median	sd	sig m/w
Flow	male female	$p = 0.462$ $p = 0.060$	2.96 3.25	3 3	0.90 0.85	$p = 0.302$ (t)
Challenge	male female	$p = 0.215$ $p = 0.546$	2.45 2.38	2.4 2.4	0.67 0.87	$p = 0.772$ (t)
Competence	male female	$p = 0.078$ $p = 0.272$	2.46 2.66	2.6 2.8	0.70 0.94	$p = 0.456$ (t)
Immersion	male female	$p < 0.05$ $p < 0.01$	2.10 2.08	1.9 2	0.78 0.80	$p = 0.968$ (w)
Positive	male female	$p = 0.458$ $p = 0.343$	3.04 3.34	3 3.6	0.77 0.89	$p = 0.270$ (t)
Tension	male female	$p < 0.001$ $p < 0.01$	1.76 1.73	1.67 1.33	0.94 0.84	$p = 1$ (w)
Negative	male female	$p < 0.001$ $p < 0.01$	1.51 1.60	1.25 1.5	0.46 0.57	$p = 0.691$ (w)
Post Positive	male female	$p = 0.053$ $p < 0.05$	1.81 2.03	1.67 2	0.49 0.70	$p = 0.319$ (w)
Post Negative	male female	$p < 0.01$ $p < 0.01$	1.34 1.20	1.26 1.17	0.33 0.22	$p = 0.166$ (w)

TABLE B.42: Overview over Relationships between Gender and GEQ Measures

Measure	Gender	shapiro p	mean	median	sd	sig m/w
Fixations	male	$p = 0.462$	1276	1223	297.16	$p = 0.601$ (t)
	female	$p = 0.999$	1340	1317	434.32	
OOI	male	$p = 0.280$	614.4	592.2	244.88	$p = 0.918$ (t)
	female	$p = 0.256$	604.5	587.5	331.82	
Horizontal	male	$p < 0.001$	0.136	0.121	0.049	$p = 0.525$ (w)
	female	$p = 0.527$	0.149	0.144	0.065	
Saccades	male	$p < 0.001$	286.8	264.8	82.06	$p = 0.979$ (w)
	female	$p < 0.05$	283.5	278.4	93.37	
Epistemic	male	$p < 0.001$	0.186	0.144	0.126	$p = 0.560$ (w)
	female	$p = 0.119$	0.150	0.156	0.088	

Table B.43: Overview over Relationships between Gender and Expert Behaviour

B.3.3.2 Age

The age variable was not normally distributed ($W = 0.920, p < 0.01$).

Measure	shapiro	correlation
Lines Made	$p = 0.292$	$\rho = 0.281$
Score	$p = 0.103$	$\rho = 0.258$
Speed	$p < 0.05$	$\rho = 0.104$
Algorithm	$p < 0.05$	$\rho = 0.172$

Table B.44: Overview over Relationships between Age and Performance Measures. Correlation has been established with Spearman's ρ

Measure	shapiro	correlation	significance
Flow	$p = 0.259$	$\rho = -0.365$	
Challenge	$p = 0.064$	$\rho = -0.318$	
Competence	$p = 0.077$	$\rho = -0.013$	
Immersion	$p < 0.05$	$\rho = -0.274$	
Positive	$p = 0.324$	$\rho = -0.207$	
Tension	$p < 0.001$	$\rho = -0.340$	
Negative	$p < 0.001$	$\rho = 0.205$	
Post Positive	$p < 0.01$	$\rho = -0.266$	
Post Negative	$p < 0.001$	$\rho = 0.192$	

Table B.45: Overview over Relationships between Age and GEQ Results. Correlation has been established with Spearman's ρ

Measure	shapiro	correlation	significance
Fixations	$p = 0.954$	$\rho = -0.018$	
OOI	$p = 0.480$	$\rho = -0.100$	
Horizontal	$p < 0.01$	$\rho = 0.390$	
Saccades	$p < 0.001$	$\rho = 0.144$	
Episteme	$p < 0.001$	$\rho = -0.217$	

Table B.46: Overview over Relationships between Age and Expert Behaviour. Correlation has been established with Spearman's ρ

B.3.3.3 Language

Measure	Language	shapiro p	mean	median	sd	sig g/e
Lines Made	German English	$p = 0.417$ $p = 0.552$	14.31 10.32	13.88 10.25	6.099 5.64	$p = 0.126$ (t)
Scores	German English	$p = 0.127$ $p = 0.715$	18.1 13.29	17 11.75	7.87 7.27	$p = 0.148$ (t)
Speed	German English	$p = 0.257$ $p = 0.159$	480ms 517.6ms	471.8ms 491ms	68.86ms 92.31	$p = 0.338$ (t)
Algorithm	German English	$p < 0.05$ $p = 0.609$	1.57 1.39	1.5 1.5	0.42 0.24	$p = 0.340$ (w)

Table B.47: Overview over Relationships between Language and Performance Measures

Measure	Language	shapiro p	mean	median	sd	sig m/w
Flow	German English	$p = 0.564$ $p = 0.062$	3.02 3.31	3 2.6	0.80 1.20	$p = 0.5502$ (t)
Challenge	German English	$p < 0.01$ $p = 0.331$	2.54 2.49	2.6 2.6	0.80 0.90	$p = 0.631$ (w)
Competence	German English	$p = 0.103$ $p = 0.959$	2.46 2.66	2.6 2.6	0.70 0.97	$p = 0.761$ (t)
Immersion	German English	$p < 0.05$ $p = 0.604$	2.03 2.54	1.8 2.2	0.69 1.12	$p = 0.222$ (w)
Positive	German English	$p = 0.184$ $p = 0.826$	3.09 3.54	3 3.6	0.74 1.07	$p = 0.319$ (t)
Tension	German English	$p < 0.001$ $p < 0.05$	1.78 1.48	1.5 1.33	0.93 0.50	$p = 0.566$ (w)
Negative	German English	$p < 0.001$ $p = 0.054$	1.58 1.57	1.5 1.25	0.48 0.72	$p = 0.553$ (w)
Post Positive	German English	$p < 0.05$ $p = 0.397$	1.85 2.14	1.78 1.83	0.49 1.07	$p = 0.655$ (w)
Post Negative	German English	$p < 0.001$ $p < 0.05$	1.28 1.17	1.17 1.26	0.29 0.30	$p = 0.787$ (w)

Table B.48: Overview over Relationships between Language and GEQ Measures

Measure	Language	shapiro p	mean	median	sd	sig m/w
Fixations	German English	$p = 0.389$ $p = 0.071$	1281 1401	1328 1279	348.22 398.89	$p = 0.479$ (t)
OOI	German E nglish	$p = 0.579$ $p = 0.726$	648.8 563	619.5 587.5	290.07 183.13	$p = 0.3286$ (t)
Horizontal	German English	$p < 0.01$ $p = 0.409$	0.133 0.200	0.123 0.186	0.049 0.064	$p < 0.05$ (w)
Saccades	German English	$p < 0.001$ $p = 0.580$	283.3 285.4	264.8 282.6	87.15 76.84	$p = 0.664$ (w)
Epistemic	German English	$p < 0.001$ $p = 0.114$	0.175 0.171	0.143 0.199	0.128 0.096	$p = 0.834$ (w)

TABLE B.49: Overview over Relationships between Language and Expert Behaviour

B.3.3.4 Mental Rotation

The results of the mental rotation tests were normally distributed ($W = 0.986, p = 0.860$).

	Lines Made	Score	Speed	Algorithm
MR	$r = 0.107$	$r = -0.088$	$\rho = 0.180$	$\rho = -0.024$

Table B.50: Overview over Relationships between Mental Rotation and Performance Measures

	Flow	Challenge	Competence	Immersion	Positive
MR	$r = 0.052$	$r = 0.219$	$r = -0.005$	$\rho = 0.049$	$\rho = 0.104$
	Tension	Negative	Post Positive	Post Negative	
MR	$\rho = -0.013$	$\rho = -0.106$	$\rho = -0.020$	$\rho = -0.240$	

Table B.51: Overview over Relationships between Mental Rotation and GEQ Results

	Fixations	OOI	Horizontal	Saccades	Episteme
MR	$r = 0.106$	$r = 0.218$	$\rho = -0.110$	$\rho = 0.208$	$\rho = 0.059$

Table B.52: Overview over Relationships between Mental Rotation and Expert Behaviour

B.3.3.5 Self-Reported Expertise

The data for the self-reported expertise was not normally distributed ($W = 0.93, p < 0.05$).

	Lines Made	Score	Speed	Algorithm
SR Expertise	$\rho = -0.161$	$\rho = -0.190$	$\rho = 0.015$	$\rho = -0.218$

Table B.53: Overview over Relationships between Self Reported Expertise and Performance Measures

	Flow	Challenge	Competence	Immersion	Positive
SR	$\rho = 0.170$	$\rho = 0.041$	$\rho = -0.263$	$\rho = -0.177$	$\rho = -0.124$
	Tension	Negative	Post Positive	Post Negative	
SR	$\rho = 0.105$	$\rho = -0.101$	$\rho = -0.028$	$\rho = 0.237$	

TABLE B.54: Overview over Relationships between Self Reported Expertise and GEQ Results

	Fixations	OOI	Horizontal	Saccades	Episteme
SR	$\rho = -0.025$	$\rho = 0.123$	$\rho = -0.176$	$\rho = -0.176$	$\rho = 0.072$

TABLE B.55: Overview over Relationships between Self Reported Expertise and Expert Behaviour

B.3.3.6 Pytris Established Expertise

The values for expertise as established by PYTRIS were normally distributed ($W = 0.967, p = 0.243$).

	Lines Made	Score	Speed	Algorithm
PE	$r = 0.435$	$r = 0.435$	$\rho = 0.047$	$\rho = 0.347$

TABLE B.56: Overview over Relationships between PYTRIS Established Expertise and Performance Measures

	Flow	Challenge	Competence	Immersion	Positive
PE	$r = -0.292$	$r = 0.012$	$r = -0.090$	$\rho = -0.054$	$\rho = -0.103$
	Tension	Negative	Post Positive	Post Negative	
PE	$\rho = 0.117$	$\rho = 0.256$	$\rho = -0.063$	$\rho = 0.116$	

TABLE B.57: Overview over Relationships between PYTRIS Established Expertise and GEQ Results

	Fixations	OOI	Horizontal	Saccades	Episteme
PE	$r = -0.089$	$r = -0.322$	$\rho = 0.011$	$\rho = 0.042$	$\rho = 0.186$

TABLE B.58: Overview over Relationships between PYTRIS Established Expertise and Expert Behaviour

Appendix C

German Translations of the Material

C.1 Consent Form

Vielen Dank für Deine Teilnahme an diesem Test. Du kannst ein Spiel sowie auch den ganzen Test jederzeit beenden. (Es macht allerdings Sinn, danach dennoch die Fragebögen auszufüllen, auch wenn Du nicht alle Spiele beendet hast.) Wenn Du irgendwelche Fragen während des Tests haben solltest, kannst Du dich jederzeit an Deine Betreuerin wenden; manchmal kann sie Dir allerdings vielleicht keine konkrete Antwort geben. Während Du den Eyetracker trägst, wäre es allerdings hilfreich, wenn Du die Fragen nach den Spielen stellst. Falls irgendwelche Probleme auftauchen, sag bitte sofort Bescheid.

Mit Deinen Daten wird vertraulich umgegangen. Dein Name wird in keinerlei Form gespeichert; wir benutzen einen codierten Identifizierer für die Fragebögen. Die aufgenommenen Daten werden nur in Forschungszusammenhängen genutzt.

Neben den Fragebögen speichern wir auch die Test Sessions in Form von Logs der Interaktionen und der Aufnahme von Augenbewegungen für die Analyse. Solltest Du irgendwelchen dieser Maßnahmen auch zu einem späteren Punkt nicht zustimmen, informiere bitte sofort die Betreuerin.

Wir hoffen, dass der Test Dir Spaß machen wird!

Bitte unterschreibe hier, um Deine Zustimmung zu den oben beschriebenen Testbedingungen zu erklären.

C.2 Framing Texts

C.2.1 Framed Participants

Du wirst zwei TETRIS Spiele á maximal fünf Minuten spielen und dann, nach einer Pause, noch einmal zwei. Die Version von TETRIS, die du spielen wirst, ist besonders, weil sie die Schwierigkeit daran anpasst was ein*e Spieler*in macht und daran, wie dabei die Augen bewegt werden. Wenn du Genaueres wissen willst, frag deine Betreuerin nach dem Test.

C.2.2 Non-Framed Participants

Du wirst zwei TETRIS Spiele á maximal fünf Minuten spielen und dann, nach einer Pause, noch einmal zwei. Da Augendaten uns etwas über die Performance von Spieler*innen aussagt, zeichen wir diese auf, um zu überprüfen, ob die Mechanismen, die anhand der Spielbewegungen das Spiel adaptieren gut funktionieren und deine Spielerfahrung verbessern. Wenn du Genaueres wissen willst, frag deine Betreuerin nach dem Test.

C.3 Questionnaires

C.3.1 Statistical Data

- Alter ______________________
- Geschlecht ______________________
- Beruf/Ausbildungsrichtung
  technisch○ sozialwissenschaftlich○ geisteswissenschaftlich○ andere○
- Sind bei Dir Sehstörungen bekannt? ja○ nein○ k.A.○
  Falls ja, welche? ______________________

  Falls ja, trägst du gerade Kontaktlinsen? ja○ nein○ k.A.○
  Falls ja, welche? hart○ weich○
- Digitale Spielerfahrung
  - Genre ______________________
    Wie oft?

    weniger als einmal im Monat○ mehr als einmal im Monat○
    mehr als einmal die Woche○ täglich oder fast-täglich○

    Kompetenz

    sehr kompetent○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ohne Kompetenz —— keine Antwort○

- Genre ____________________

  Wie oft?

  weniger als einmal im Monat ○ mehr als einmal im Monat ○
  mehr als einmal die Woche ○ täglich oder fast-täglich ○

  Kompetenz

  sehr kompetent ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ohne Kompetenz —— keine Antwort ○

- Kennst Du TETRIS ja ○ nein ○
- Wenn ja, wann hast Du zuletzt TETRIS gespielt?

  ____

- Wie kompetent glaubst du, bist du in TETRIS?

  sehr kompetent ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ohne Kompetenz —— keine Antwort ○

C.3.2 After every Game

- Ich fühlte mich erfolgreich

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich war gelangweilt

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich vergaß alles um mich herum

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich war frustriert

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich fand es ermüdend

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich war gereizt

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich fühlte mich geschickt

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich war vollständig absorbiert

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich war zufrieden

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich fühlte mich herausgefordert

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich musste viel Mühe aufwenden

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich fühlte mich gut

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

C.3.3 Game Experience Fragebogen

- Ich war zufrieden

  gar nicht ○ ein bisschen ○ moderat ○ ziemlich ○ extrem ○ —— k. A. ○

- Ich fühlte mich geschickt

  gar nicht ○ ein bisschen ○ moderat ○ ziemlich ○ extrem ○ —— k. A. ○

- Ich hatte Spaß

  gar nicht ○ ein bisschen ○ moderat ○ ziemlich ○ extrem ○ —— k. A. ○

- Ich war vollständig mit dem Spiel beschäftigt

  gar nicht ○ ein bisschen ○ moderat ○ ziemlich ○ extrem ○ —— k. A. ○

- Ich war glücklich

  gar nicht ○ ein bisschen ○ moderat ○ ziemlich ○ extrem ○ —— k. A. ○

- Es hat mich in schlechte Laune gebracht

  gar nicht ○ ein bisschen ○ moderat ○ ziemlich ○ extrem ○ —— k. A. ○

- Ich dachte über andere Dinge nach

  gar nicht ○ ein bisschen ○ moderat ○ ziemlich ○ extrem ○ —— k. A. ○

- Ich fand es ermüdend

  gar nicht ○ ein bisschen ○ moderat ○ ziemlich ○ extrem ○ —— k. A. ○

- Ich fühlte mich kompetent

  gar nicht ○ ein bisschen ○ moderat ○ ziemlich ○ extrem ○ —— k. A. ○

- Ich fand es schwierig

  gar nicht ○ ein bisschen ○ moderat ○ ziemlich ○ extrem ○ —— k. A. ○

- Es war ästhetisch ansprechend

  gar nicht ○ ein bisschen ○ moderat ○ ziemlich ○ extrem ○ —— k. A. ○

- Ich vergaß alles um mich herum

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich fühlte mich gut

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich war gut darin

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich war gelangweilt

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich fühlte mich erfolgreich

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich fühlte mich einfallsreich

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich hatte das Gefühl, dass ich Dinge erforschen konnte

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich fand Gefallen daran

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich war schnell dabei, die Ziele des Spiels zu erreichen

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich fühlte mich genervt

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich fühlte mich unter Druck gesetzt

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich fühlte mich gereizt

gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich verlor mein Zeitgefühl

gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich fühlte mich herausgefordert

gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich fand es beeindruckend

gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich war tief konzentriert im Spiel

gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich war frustriert

gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Es fühlte sich nach einer reichhaltigen Erfahrung an

gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich habe die Verbindung zur Außenwelt verloren

gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich fühlte mich unter Zeitdruck

gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich musste viel Mühe aufwenden

gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

C.3.4 Post-Game Fragebogen

- Ich fühlte mich wie neu belebt

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich fühlte mich schlecht

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich fand es schwierig, in die Realität zurückzukommen

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich fühlte mich schuldig

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Es fühlte sich wie ein Sieg an

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich fand es war Zeitverschwendung

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich fühlte mich angeregt

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich war zufrieden

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich war desorientiert

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich fühlte mich erschöpft

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich hatte das Gefühl, dass ich sinnvollere Dinge hätte machen können

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich fühlte mich mächtig

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich fühlte mich ausgelaugt

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich fühlte Reue

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich habe mich geschämt

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich war stolz

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

- Ich hatte das Gefühl, als ob ich von einer langen Reise zurückgekommen wäre

  gar nicht○ ein bisschen○ moderat○ ziemlich○ extrem○ —— k. A.○

Vielen Dank für Deine Mithilfe.

Declaration of Authorship

I, Katharina Spiel, declare that this thesis titled, 'Frames and Lenses

Framing Gameplay Experience in Games with Eye Movement Based Adaptation' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at the Bauhaus Universität Weimar.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this university or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: ____

Date: ____