UNEXPECTED EVENTS AND USER EXPERIENCE

SURPRISE AS A DESIGN STRATEGY FOR INTERACTIVE PRODUCTS

vorgelegt von

Alice Marlene Gross, M.Sc.

von der Fakultät V– Verkehrs- und Maschinensysteme der Technischen Universität Berlin zur Erlangung des akademischen Grades

Doktorin der Naturwissenschaften
- Dr. rer. nat.-

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender:	Prof. Dr.-Ing. Marc Kraft
Gutachter:	Prof. Dr. phil. Manfred Thüring
Gutachter:	Prof. Dr. phil. Dietrich Manzey
Gutachter:	Prof. Dr. sc. nat. Hartmut Wandtke

Tag der wissenschaftlichen Aussprache: 24. Juni 2016

Berlin, 2016

"Our brightest blazes of gladness are commonly kindled by unexpected sparks."

Samuel Johnson

EIDESSTATTLICHE ERKLÄRUNG

Hiermit erkläre ich gemäß § 5 (5) der Promotionsordnung für die Technische Universität Berlin an Eides statt, dass ich die Dissertation selbstständig verfasst habe; die von mir verwendeten Hilfsmittel und Quellen sind aufgeführt und die Arbeit ist nicht in Zusammenarbeit mit anderen Wissenschaftlern oder Wissenschaftlerinnen erstellt worden.

Gemäß § 5 (7) der Promotionsordnung für die Technische Universität Berlin versichere ich außerdem, dass ich die Dissertation nur in diesem und keinem anderen Promotionsverfahren eingereicht habe und, dass diesem Promotionsverfahren keine endgültig gescheiterten Promotionsverfahren vorausgegangen sind.

Hiermit versichere ich, dass ich über die urheberrechtlichen Nutzungsrechte an allen Teilen des Werkes verfüge und Rechte Dritter, insbesondere Urheber- und Persönlichkeitsrechte Dritter, mit der Veröffentlichung nicht verletzt werden.

_____ _____

Datum Alice Marlene Gross

ABSTRACT

User Experience is a concept that extends traditional usability by providing a holistic perspective on the user's interaction with technology. It is seen as an important part of successful product design (ISO 9241- 210, 2010). Strong competition fosters innovative user experience design and the development of new design techniques. One of the pillars of user experience are emotional reactions to a product (Thüring & Mahlke, 2007). Surprise is an emotion that has proven to be beneficial for the user-product relationship (Ludden, 2008; Ludden Schifferstein & Hekkert, 2012). The goal of surprising users is to generate pleasant feelings and ultimately 'produce' satisfied users. Surprise offers designers the possibility to create pleasant user experience through positive, as well as negative experiences during interaction. Setting a product apart from others by creating strong experiences and emotions through surprise could separate success from failure. In summary, surprising users by using unexpected product features has proven to be beneficial for product user interaction and user experience. To investigate these effects for digital interactive products, three laboratory studies were conducted. The creation of surprise, the bipolar valence of surprise and its effects on users, as well as the use of different contexts of use were investigated. Several aspects of surprise were systematically varied and their effects on user experience ratings of the products were analyzed. Valence, as well as frequency of surprise was investigated in two different contexts of use. Results indicate that surprise is a promising design tool to enhance product evaluations and product success. Furthermore, the frequency of surprise has an influence on its effects. Nevertheless, it is a difficult emotion to design. Many factors complicate the use of surprise as a design feature and must be carefully studied. Designers need to know about the intricate interplay of these factors and should try to gain insights into the minds of their prospective users before and during the design process.

ZUSAMMENFASSUNG

User Experience erweitert das Konzept der Usability durch eine ganzheitliche Betrachtung der Mensch-Technik-Interaktion. Sie wird als wichtiger Bestandteil erfolgreicher Produktgestaltung betrachtet (ISO 9241- 210, 2010). Starker Wettbewerb fördert besonders innovatives User Experience Design sowie die Entwicklung von neuen Design-Techniken. Ein wichtiger Teil von User Experience sind die emotionalen Reaktionen auf ein Produkt (Thüring & Mahlke, 2007). Eine Emotion, die nachweislich von Vorteil für die Interaktion mit einem Produkt ist, ist Überraschung. Durch Überraschung entstehen positive Emotionen, und somit zufriedene Anwender (Ludden, 2008; Ludden Schifferstein & Hekkert, 2012). Überraschung bietet Designern die Möglichkeit, angenehme Benutzererfahrung sowohl durch positive, als auch negative Erfahrungen während der Interaktion zu schaffen. Ein Produkt kann sich dabei von anderen absetzen, indem es besondere Erfahrungen produziert und Gefühle hervorruft.

Benutzern mit Hilfe von Überraschung ein positives Nutzererleben zu ermöglichen kann also vorteilhaft sein. Um diese Effekte auch mit digitalen, interaktiven Produkten zu untersuchen, wurden drei Laborstudien durchgeführt. Der Erkenntnisgewinn dieser Arbeit liegt darin Aussagen über das Design und die Anwendung adäquater Überraschungen in verschiedenen Nutzungssituationen zu treffen. Hierbei stehen auch die besondere Valenz von Überraschung und die Rolle des Nutzungskontexts im Fokus. In insgesamt drei empirischen Experimenten wurden Valenz sowie Präsentationshäufigkeit in zwei verschiedenen Nutzungskontexten untersucht. Die Ergebnisse zeigen, dass Überraschung ein vielversprechendes Gestaltungswerkzeug ist, um Produktbewertungen und Produkterfolg zu verbessern. Des Weiteren hat die Präsentationsfrequenz einer Überraschung des Einfluss auf die Auswirkungen dieser. Insgesamt zeigt die Arbeit, dass es schwierig ist Überraschung zu designen, dass die positiven Auswirkungen einer gelungenen Überraschung auf den Nutzer jedoch sehr groß sein können. Einige Eigenschaften von Überraschung und deren Wirkung auf Nutzer erschweren die Verwendung von Überraschung als Gestaltungselement und müssen sorgfältig untersucht werden. Designer sollten sich über das komplizierte Zusammenspiel dieser Faktoren im Klaren sein und versuchen, einen Einblick in die Bedürfnisse potentieller Nutzer vor und während des Design-Prozesses zu gewinnen.

TABLE OF CONTENTS

1 INTRODUCTION 1

1.1 Research problems ... 2
1.2 Research goals ... 3
1.3 Research questions ... 4
1.4 Structure of the thesis ... 5

2 THEORETICAL BACKGROUND 7

2.1 User experience ... 7
2.2 Different UX models ... 10
2.2.1 UX model AttrakDiff ... 10
2.2.2 CUE model ... 12
2.2.3 A facet model of UX ... 14
2.2.4 Other approaches to UX – an overview ... 15
2.2.5 Summary ... 16
2.3 General emotions ... 17
2.3.1 What is an emotion? ... 17
2.3.2 Purpose of emotions ... 19
2.3.3 Affect, emotion, and mood ... 20
2.3.4 How to measure emotions ... 22
2.3.5 Conclusion ... 23
2.4 Emotional product design approaches ... 24
2.4.1 Pleasurable design approaches ... 24
2.4.2 Norman's model of emotional design ... 25
2.4.3 Desmet and Hekkert's basic model of product emotions ... 26
2.4.4 Other approaches to emotional product design ... 28
2.5 A special emotion: Surprise ... 30
2.5.1 The surprise process ... 31
2.5.2 Surprise and reaction times ... 33
2.5.3 Pleasant and unpleasant surprise ... 34
2.5.4 Different forms of surprise ... 36
2.5.5 Expectations and uncertainty ... 36
2.5.6 Facial expressions accompanying surprise ... 37
2.5.7 Emotions following surprise ... 38
2.6 Surprise as a design strategy ... 38
2.6.1 Surprising product design ... 38
2.6.2 Surprise and product evaluations ... 40
2.6.3 Conclusion ... 43
2.7 Summary ... 44

3 RESEARCH APPROACH 45

3.1 Fundamental restrictions when working with surprise ... 45
3.2 Research Paradigms ... 45

3.2.1 Gaming environment ... 45
3.2.2 Goal-directed environment ... 47

3.3 Questionnaires ... 48

4 STUDY I– THE INFLUENCE OF UNEXPECTED EVENTS ON UX IN A GAME I 51

4.1 Problem and research goals ... 51
4.2 Method ... 53

4.2.1 Participants ... 53
4.2.2 Materials ... 53
4.2.3 Procedure ... 54
4.2.4 Design ... 55
4.2.5 Hypotheses ... 58

4.3 Results ... 61

4.3.1 Analysis ... 61
4.3.2 Surprise ... 61
4.3.3 User Experience Questionnaires ... 64

4.4 Discussion ... 65

4.4.1 Surprise ... 65
4.4.2 User Experience ... 67
4.4.3 Summary and contribution ... 69
4.4.4 Critical evaluation ... 70

5 STUDY II - THE INFLUENCE OF UNEXPECTED EVENTS ON UX IN A GAME II 71

5.1 Problem and research goals ... 71
5.2 Method ... 71

5.2.1 Participants ... 71
5.2.2 Materials ... 72
5.2.3 Procedure ... 72
5.2.4 Design ... 73
5.2.5 Hypotheses ... 75

5.3 Results ... 76

5.3.1 Surprise ... 76
5.3.2 EMG Data ... 78
5.3.3 User Experience Questionnaires ... 79

5.4 Discussion ... 80

5.4.1 Reaction times ... 80
5.4.2 EMG ... 81

5.4.3 UX Questionnaires ... 82
5.4.4 Summary and contribution ... 84
5.4.5 Critical evaluation ... 85

6 STUDY III: THE INFLUENCE OF UNEXPECTED EVENTS IN A GOAL-ORIENTED TASK ENVIRONMENT 87

6.1 Problem and research goals ... 87
6.2 Method ... 88

6.2.1 Participants ... 88
6.2.2 Materials ... 88
6.2.3 Procedure ... 92
6.2.4 Design ... 92
6.2.5 Hypotheses ... 95

6.3 Results ... 96

6.3.1 Manipulation check ... 96
6.3.2 Reaction times ... 97
6.3.3 User Experience Questionnaires ... 98
6.3.4 Open questions ... 100

6.4 Discussion ... 102

6.4.1 Reaction times ... 102
6.4.2 User Experience Questionnaires ... 103
6.4.3 Open questions ... 104
6.4.4 Summary and contribution ... 105
6.4.5 Critical evaluation ... 105

7 GENERAL DISCUSSION 107

7.1 Digital products need various kinds of surprises. ... 108
7.2 Investigating the effects of pleasant, unpleasant, and neutral surprise on the UX of digital interactive products. ... 109
7.3 Investigate the influence of frequency of presentation of a surprising event on UX of a digital interactive product. ... 109
7.4 Returning to the theoretical basis ... 110

7.4.1 Difficulties when designing for surprise ... 110
7.4.2 Emotional design approaches ... 112
7.4.3 Mood and emotion ... 113

7.5 Critical evaluation of the empirical findings ... 114

7.5.1 Design of surprise ... 114
7.5.2 Context ... 115
7.5.3 Interpersonal differences ... 115
7.5.4 Method of measurement ... 115

8 OUTLOOK AND CONCLUSION 117

9 REFERENCES 119
APPENDIX A: STUDY I 131

A.1 Instructions positive group 131
A.2 Instructions negative group 132
A.3 Instructions control and neutral group 132
A.4 Consent Form and demographic questionnaire 133
A.5 UX questionnaires 135

A.5.1 SAM 135
A.5.2 Product Emotions MANO 136
A.5.3 AttrakDiff & Overall rating 137
A.6 Additional statistical data 139
A.6.1.RT data 139
A.6.2.Questionnaire data 139

APPENDIX B: STUDY II 143

B.1 Consent Form and demographic questionnaire 143
B.2 Instructions 144
B.3 UX questionnaires 145

B.3.2 meCUE questionnaire 145
B.3.3 AttrakDiff mini Questionnaire 147

B.4 Additional statistical data 148

B.4.1 Reaction time data 148
B.4.2 EMG data 150
B.4.3 Questionnaire data 150

APPENDIX C: STUDY III 155

C.1 Consent Form and demographic questionnaire 155
C.2 UX questionnaires 157

C.2.1 meCUE questionnaire 157

157

C.2.2 Surprise item and open questions 161

C.3 Instructions as seen on the tablet 162
C.4 Tasks descriptions 164
C.5 Additional statistical data 167

C.5.1 Reaction time data 167
C.5.2 SAM questionnaire data 168
C.5.3: MeCUE questionnaire data 170

LIST OF FIGURES

Figure 2.1: Key elements of the model of user experience from (a) a designer perspective and (b) a user perspective (Hassenzahl, 2005). 11

Figure 2.2: The CUE model of user experience by Mahlke and Thüring, 2007. 14

Figure 2.3: The circumplex model of affect by Russell (1980). 21

Figure 2.4. Hierarchy of consumer needs (Jordan, 1997; Hancock et al, 2005). 25

Figure 2.5. The basic model of product emotions (Desmet, 2003) 27

Figure 2.6. Classification of product emotions (Desmet. 2003) 28

Figure 2.7: The cognitive-evolutionary model of surprise (Meyer et al., 1997). 32

Figure 2.8: Stimulus events used in the study by Meyer et al., 1997. 34

Figure 2.9: Formation of pleasant and unpleasant surprise according to Reisenzein (2009). 35

Figure 2.10. Facial expressions of surprise (Ekman & Friesen, 1971, left), and surprise and fear (Ekman, 2007, right) 37

Figure 2.11: Surprising product designs. 39

Figure 2.12: Example of a product set for soft boxes looking as if made out of felt, ranging from soft and flexible to inflexible and rough. 43

Figure 3.1. The seven tetronimoes (left); Screenshot of the Tetris game (right). 46

Figure 3.2: Available connections as represented in *öffi* application and during Study III.47

Figure 4.1. Research computer (left); experimenter work space with EMG data screen, screen recorder screen, and copy of research screen (middle, from left to right); participant with EMG electrodes (right). 54

Figure 4.2: Unexpected events for negative group (right), and positive group (left). 56

Figure 4.3: Colour change as manifestation of the unexpected event in the neutral group.56

Figure 4.4. Original block colours (top); saturated block colours (middle); corresponding blocks (bottom.) 57

Figure 4.5. Electrode locations for measuring facial EMG activity (from van Boxtel, 2010). 58

Figure 4.6: Raw RTs per group averaged over all participants in each group. 62

Figure 4.7: Uncorrected RTs (left) and corrected RTs (right) with SE for factor event..... 62

Figure 4.8: Uncorrected RTs (left) and corrected RTs (right) with SE for factor group..... 63

Figure 5.1: Unexpected events for group neutral (left), group positive (middle), and group negative (right)..... 73

Figure 5.2: Raw RTs per group averaged over all participants in each group. 77

Figure 5.3: Uncorrected RTs (left), corrected RTs (right)..... 77

Figure 5.4: µV change over the lateral frontalis muscle for each event with SE. 78

Figure 5.5: Subjective surprise ratings with SE. 80

Figure 6.1: Experimental setup with tablet, camera, and folder with experimental tasks.. 89

Figure 6.2: UI of the application screens 1 to 4 (screen1: upper left, screen 2: upper right; screen 3: bottom left, screen 4: bottom right)..... 90

Figure 6.3: Example of a manipulation following a touch on the element "U6". 91

Figure 6.4: Experimental design with moment of appearance of unexpected events. 93

Figure 6.5: Screen three with information icon displayed for one station. 94

Figure 6.6: Example of sorting of paper representations of animations according to liking. 97

Figure 6.7: Baseline corrected RTs for both experimental groups per event. 98

Figure 6.8: Mean SAM valence ratings for group A and B after the baseline phase and the experimental phase with SE..... 99

Figure 6.9: Answers to open question 1 (absolute frequencies)..... 101

Figure 6.10: Answers to open question 2 (absolute frequencies)..... 101

LIST OF TABLES

Table 4.1: Mean ratings from UX questionnaires with standard deviations (SD)..... 64

Table 5.1: Mean UX ratings from questionnaires with SDs..... 80

Table 6.1: Original shapes and their corresponding visual manipulations and movement paths. 91

Table 6.2: Absolute differences between BL and EXP for group A and B with SD. 100

Table 6.3: X^2 table of rating of the animation per group. 102

ABBREVIATIONS

ANOVA	Analysis of Variances
BDTE	Belief Desire Theory of Emotion
BL	Baseline Phase
EEG	Electroencephalography
EMG	Electromyography
EXP	Experimental Phase
HCI	Human-Computer Interaction
HQI	Hedonic Quality Identification
HQS	Hedonic Quality Stimulation
MANOVA	Multivariate Analysis of Variances
MAYA	Most Advanced Yet Acceptable
MS	Milliseconds
PQ	Pragmatic Quality
ROI	Return on Investment
RT(s)	Reaction Time(s)
SAM	Self Assessment Manikin
SD	Standard Deviation
SE	Standard Error
SOA	Stimulus Onset Asymmetry
UCD	User-Centred Design
UI	User Interface
UX	User Experience

1 INTRODUCTION

Emotions are part of our everyday life, but are considered to only play a small role in our technological life. Nevertheless, in one way or another all our actions involve emotional interchanges, including interactions with the material world (Desmet, Overbeeke & Tax, 2001). Think about your first technological gadget. What was it? A Walkman? A personal computer? A mobile phone? Or an electrical musical instrument? Whatever it is that you are thinking about now, you have some memories of it, accompanied by feelings and emotions: maybe the walkman got sand in it on the beach and did not work anymore which made you sad. Or you spent hours and hours happily playing Snake[1] on your first mobile phone. Based on our experiences with those kinds of technologies, we form expectations about the interaction with other technologies, learn from mistakes, and attach feelings to those technological objects. From our first encounter with technology on, we acquire information and apply this new knowledge when interacting with other products. Though, we are interacting with lifeless objects, we experience emotions.

With the rapid development of technology and the fast entry of digital interfaces into everyday life, the need for understanding other aspects than mere functionality and utility of technology has become inevitable. In the early 1990s, a paradigm shift from functionality to usability was observed (Davis, 1993; Hassenzahl, Beu & Burmester, 2001; Hassenzahl & Tractinsky, 2006). When it became apparent that merely building technology to do the job was not enough, user centred approaches for interface design were advocated. These methods and the advertising of usability as a golden design rule improved the overall design of interfaces. In addition, the focus of Human-Computer Interaction (HCI) research changed again several years later, as "technology as a tool" gave way to "technology to play with" (Schrammel, Geven & Tscheligi, 2006).

A second paradigm shift could be observed around the turn of the millennium (Alben, 1996; Bargas-Avila & Hornbæk 2011; Logan, 1994; Logan, Augaitis, & Renk, 1994; Monk & Fröhlich, 1999; Schrammel et al., 2006). As Norman already mentioned in 2002, “usable designs are not necessarily pleasurable ones” (Norman, 2002; p. 38). The concept of mere

[1] A video game for mobile phones, popular in the 1990s.

usability was broadened and expanded by the introduction of non-instrumental, hedonic, aesthetic, or emotional factors. User experience (UX) was introduced as a construct to describe this holistic approach to user-product interaction. Since then, emotions have become a constant part of interface design approaches. Concepts like pleasure, fun, wow, and awe have found their way into every designer's work. At the same time, the importance of understanding aesthetic, emotional and fun-related aspects of human computer interaction has increasingly been recognised. McCarthy and Wright (2004) even state that the emotional quality of a product tends to summarize the whole experience. Especially consumer electronics strive for enjoyable interaction. One emotion that has proven to be especially beneficial for the user-product relationship is surprise (Ludden, 2008; Ludden Schifferstein & Hekkert, 2012). One goal of surprising users is to generate pleasant feelings and ultimately 'produce' satisfied users (Vanhamme & Snelders, 2003). However, it is difficult to manipulate surprise and be sure about its ultimate polarity (Vanhamme, 2000). Companies trying to elicit positive surprise have often discovered that they in fact triggered the opposite (Vanhamme, 2000). Surprising the user in order to create a positive and satisfying experience is not easy. In addition, some researchers argue for including unpleasant emotions in design to ultimately generate a pleasant experience (Fokkinga & Desmet, 2013; Silvia, 2009). Following their theory, negative surprise can yield positive effects if implemented correctly.

In today's market, products are judged and compared on the basis of their features, their usability, and their quality (Veryzer, 1995). This poses a challenge for companies striving to differentiate their products from those of their competitors. Especially with strong competition on the market, producers of mobile applications or websites, for example, could benefit from insights into possible beneficial and harmful effects of pleasant and unpleasant surprise. Setting a product apart from others by creating strong experiences and emotions through surprise could separate success from failure. Interaction experience will become the main success factor of a product, as we are approaching the 'loyalty decade' (Nielsen, 2008). Accordingly, close attention should be paid to its design.

1.1 RESEARCH PROBLEMS

Surprising the user of a digital product seems to be a promising design approach (Derbaix & Vanhamme, 2003; Ludden, 2008; Machlovi, 2015; Vanhamme, 2000). Surprise has even become part of brand claims for global players like the car manufacturer KIA ("KIA – The Power to Surprise"). Despite an abundance of work on emotional product design (e.g.

Desmet & Hekkert, 2002; Helander & Tham, 2003; Jordan, 1997; Monk, Hassenzahl, Blythe & Reed, 2002; Norman, 2002), surprise has barely been the focus of these efforts. Although surprise seems to have a natural appeal to designers, it is implemented, but not scrutinized scientifically. A stream of research by Ludden and colleagues (Ludden, 2008; Ludden & Schifferstein, 2007; Ludden, Schifferstein & Hekkert, 2008) about surprise as a design strategy represents a first attempt at a systematic investigation of the subject and has broadened the field of emotional design research. Unfortunately, these insights from systematic scientific investigations of surprise in design are restricted to non-digital products like vases or lamps.

The research conducted about classical product design and surprise also tends to implement surprise without accounting for the different forms this emotion can take. As an exception to the rule, surprise is neutral in valence and high in arousal (Russell, 1980), but leads to a valenced reaction after the unexpected event has been matched against a person's personal concerns, resulting in pleasant or unpleasant surprise (Reisenzein, 2009). Failure to distinguish these different forms of surprise when investigating user experiences is problematic. A systematic approach investigating neutral, pleasant and unpleasant surprise in digital interactive products therefore seems to be reasonable.

Recent approaches to emotional product design promote the beneficial effects of unusual and unpleasant emotions, including unpleasant surprise (Fokkinga & Desmet, 2012, 2013; Silvia, 2010). Consequences of unpleasant or negative surprises cannot be predicted definitely. In the right context with the suitable design frame, negative could lead to positive.

The creation of surprise in an artificial laboratory context is another challenge when investigating surprise as a design feature. In order to investigate the effects of surprise on user experience, it has to be ensured that all participants experience a comparably surprising event. Accordingly, a successful surprise creation needs to be validated in order to be able to attribute variations in product ratings to these surprise events.

1.2 RESEARCH GOALS

Three research goals are to be answered. First, the basic requirement for investigating surprising product design is to validly generate surprise in users as needed. Therefore, it is of particular importance to investigate the possibility of generating different valences of surprising events. Accordingly, the first research goal is to create different kinds of surprises for the users of a number of digital products.

Another open topic is the interplay between pleasant, unpleasant, and neutral surprises and the subsequent influence on UX. Building on findings from classical product design, where surprise as a design element had a positive effect on users[2], the aim is to transfer these findings to the product domain of digital interactive products. For this reason, the second research goal is to investigate the effects of pleasant, unpleasant and neutral surprise on the UX of digital interactive products.

Third, the frequency of presentation of unexpected events is investigated in order to find out when a surprising event loses its surprising character and becomes a simple feature of a product. The third research goal is therefore to investigate the influence of frequency of presentation of a surprising event on the UX of a digital interactive product.

1.3 RESEARCH QUESTIONS

The following research questions are going to be addressed:

- Is it possible to create surprise through the implementation of unexpected events during interaction with a digital product?
- Does the frustration[3] of expectations (i.e. surprise) influence the UX of the digital product?
- Do pleasant and unpleasant surprises have different effects on users?
- Does the frequency of unexpected events influence the subjective experience of surprise?
- Does this influence vary between different contexts of use?

In addition, it is expected to gain further insights into the following areas:

- the development of unexpected events to generate surprising experiences for users
- implications for designers about how, when and how often to use surprise as a design strategy

[2] As has been mentioned above, the experiments by Ludden and colleagues did not explicitly distinguish between different kinds of surprise, e.g. pleasant, unpleasant, or neutral surprise. It seems that the authors assume their unexpected events to elicit pleasant surprise.

[3] The terminology *to frustrate an expectation* will be used throughout the whole thesis to describe a situation in which one thing is expected and something else happens. In this context, *to frustrate* does not have a negative connotation, it merely describes the process of the failure to fulfill an expectation.

1.4 STRUCTURE OF THE THESIS

This dissertation attempts to answer to the research questions and to elaborate on the research goals through the several empirical studies.

Chapter 2 provides the reader with the necessary theoretical background to the approach of this dissertation. The construct of UX and associated concepts are explained. The connection from the definition of usability and UX over emotions (and surprise in particular) to emotional product design is established in order to illustrate the need for research on surprise as a design strategy. Difficulties are illustrated that result from the uniqueness of surprise, an emotion that can be neutral, pleasant, or unpleasant.

Chapter 3 gives an overview of the experimental paradigms used in the three studies, as well as a description of all employed questionnaires.

Chapter 4 presents Study I which is an investigation of unexpected events and corresponding consequences on UX ratings in a gaming context. Tetris, a puzzle like computer game, was used as experimental paradigm. Objective as well as subjective data were collected in order to elaborate on the success of the surprise manipulation and on the effects on UX ratings of the game.

Chapter 5 presents Study II which extended the findings from Study I by specifying the characteristics of the unexpected events that were necessary to reliably elicit a user's surprise reaction.

Chapter 6 presents Study III. The third study transferred findings from the first two studies from a gaming context into a goal-directed one. Additionally, frequency of events was introduced as an experimental factor.

Chapter 7 integrates and critically discusses the findings of the three studies and connects them to the theoretical background described in Chapter 2.

Chapter 8 gives an outlook for future empirical work as well as for practical implications.

2 THEORETICAL BACKGROUND[4]

Emotions were included relatively late in the design process for interactive products. For a long time, functionality, usefulness, and usability were considered important and sufficient for a successful HMI interface. The construct UX is built upon usability, as well as general emotions towards a product. Both aspects need to be considered when thinking about UX or other special emotions, like surprise.

2.1 USER EXPERIENCE

When it comes to the evaluation of interactive qualities of technical products, usability started its success in the late 1970s and early 1980s. It is defined in the ISO 9241-11 (ISO, 1998) as "...the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use." For a long time, usability, and functionality, were identified as the most important qualities of a product (Bias & Mayhew, 2005; Hassenzahl & Tractinsky, 2006; Norman, 2002). Usability primarily focuses on user cognition and performance in human-technology interactions. It leaves out all aspects of the context of use and consequences which define the non-utilitarian aspects of interaction like "user affect, sensation and the meanings as well as values of such interactions in everyday life" (Law, Roto, Hassenzahl, Vermeeren & Kort, 2009, p.719). With the increasing introduction of technological devices in areas other than the work place or the military domain, new requirements emerged apart from functionality or usability (Schrammel et al., 2006). To make up for the shortcomings of usability as a holistic concept to estimate quality of use and user satisfaction, the concept of user experience (UX) was developed in the scientific community. It is defined in the DIN EN ISO 9241-210:2010 standard, which states that UX

[4] Chapter 2 is partly based on: Gross, A, & Bürglen, J. (in press). The Impact of Expectations on User Experience: Surprising the User. In *Proceedings of HCI 2014: Sand, Sea & Sky: The 28th BCS Conference on Human-Computer Interaction.* Southport, UK, September 9-12, 2014. BCS.

consists of a "person's perceptions and responses that result from the use or anticipated use of a product, system or service". Already in 1986, Norman and Draper focused on the experience a user might have when interacting with a product (Norman & Draper, 1986). Nine years later, Norman, Miller and Henderson where among the first authors to use the term UX to describe all aspects a user may experience when interacting with a system, going beyond usability (Norman, Miller & Henderson, 1995). Norman describes his view that usability is a too narrow concept to grasp the holistic vision of human-computer interaction. Similarly, Alben focuses on an interaction that provides people with a satisfying experience (Alben, 1996). She focuses on this experience and subsumes all the aspects of how users interact with a product: the product's touch and feel, how well users understand how it works, their feelings during interaction, how well it matches their needs, and how well it fits the context of use. Alben subsumes all these experiences under the term 'quality experience'.

Likewise, Logan (1994) pointed out that the traditional notion of usability had not been challenged or expanded because it was fully adequate for military or process control, areas not necessarily linked to fun or enjoyment, but rather to functionality and accuracy. When more and more consumer and entertainment related products flooded the market, competing with classical work and productivity products for the users' attention, the definition of user needs changed. Logan, being one of the first interface designers to realize this shift, came up with the term emotional usability. Already in 1994, he summarized aspects of the user-product relation which encompass more than just functional characteristics, arguing at the same time for an expanded definition of usability (Logan, Augaitis, & Renk, 1994). The novelty of UX lies in its holistic perspective on HCI, including non-utilitarian concepts like joy (Malone, 1981), fun (Carroll and Thomas, 1988; Monk & Fröhlich, 1999), pleasure (Jordan, 2003), hedonic values (e.g. Hassenzahl, 2001, 2003), or ludic values (Gaver et al., 2004). Expanding the concept of pure usability by these and other non-instrumental[5] factors, some discuss UX in terms of an "extended and distinct perspective on the quality of interactive products" (Hassenzahl, 2008, p. 11).

Although the concept of UX has had this warm welcome into academia and the professional world, a consensus on a clear definition of the term and its associated concepts has not been reached (Gross & Bongartz, 2012; Law, Vermeeren, Hassenzahl & Blythe, 2007; Law, Roto, Hassenzahl, Vermeeren & Kort, 2009). Still, it is widely used and has become a buzz word for practitioners and researchers alike (Lallemand, Gronier &

[5] Instrumental and non-instrumental are terms describing the characteristics of a system. Instrumental refers to characteristics concerning utility and usability whereas non-instrumental refers to aspects like emotions or aesthetics (Mahlke & Thüring, 2007).

Koenig, 2014; Hassenzahl & Tractinsky, 2006). There are several reasons for the apparent difficulty involved in agreeing on a common definition. The term is associated with "a broad range of fuzzy and dynamic concepts" (Law et al., 2009, p. 719) which include a vast variety of psychological constructs like emotions, affect, aesthetics, and hedonism, as well as constructs adopted from classical usability like utility or consequences of use (Minge & Riedel, 2013). This diversity of UX definitions can be partly attributed to the fact that UX has been studied by researchers and practitioners from a range of disciplines, all taking different approaches to the topic. Another reason for the disagreement on a universal UX definition can be found in its highly context-dependent and subjective nature (Karapanos, Hassenzahl & Martens, 2008; Pohlmeyer, 2011).

In an attempt to reach agreement on the nature and scope of UX, Law et al. (2009) conducted a survey which gathered information of 275 researchers and practitioners working in the field of user research. Agreement could be found in the notion that UX is dynamic, context-dependent and subjective. Furthermore, participants considered UX as being something new that most definitely should be a part of HCI and User-Centred Design (UCD). Influential factors in the evaluation of UX originated from differences in duration of working professionally with UX and socio-cultural factors like country of origin. Participants agreed on factors influencing UX, such as the current internal state, earlier experiences, and current context.

In a recent attempt to replicate the findings of Law and colleagues (2009), a second large scale survey was conducted by Lallemand, Gronier & Koenig (2014), including the viewpoints of 758 practitioners about UX. The aim was to gain insights into the maturation process of the UX concept, as well as validating the previous findings which had been "...almost taken for granted by the HCI community" (p. 46). It was shown that about 6 years after the first survey, UX was no longer perceived as a new approach. On the contrary, participants understood UX as a concept that was now widely covered by existing engineering approaches, having emerged from UCD and usability. What had not changed was the view that UX is shaped by user-related factors, as well as contextual factors. Practitioners highlighted the user, as well as the complexity of the concept, the components of UX, and the consequences of the experience. New to this was the highlighting of the ambivalent affective feature of UX: practitioners agreed on the notion that UX can be positive as well as negative and researchers should therefore not be focused on the positive UX, exclusively.

2.2 DIFFERENT UX MODELS[6]

In the early 2000s, the growing interest in the UX concept led to the development of several models and frameworks on UX with the goal of better understanding the user's experience and to identifying and systemizing the factors influencing it. All of these approaches try to define UX and its associated concepts, trying to tie together the aforementioned variety of factors contributing to UX. The models emphasize different concepts to different extents, illustrating the heterogeneity prevailing in the area of UX research.

2.2.1 UX MODEL ATTRAKDIFF

In the 2005 book chapter *The Thing and I: Understanding the Relationship between User and Product*, Hassenzahl established a model of UX which subsequently led to the development of the AttrakDiff questionnaire for assessing UX (Hassenzahl, Burmester & Koller, 2003). The model was a more detailed version of a previous one (Hassenzahl, 2001), in that it addressed the subjective nature of experience, the perception of a product, emotional responses to a product, and different contextual settings, as well as two distinct perspectives on UX; the designer's perspective and the user's perspective. The most notable quality of Hassenzahl's UX model might be this distinction between the two perspectives. Two different kinds of interplay between product features and product attributes are therefore presented which can interact to evolve into utterly different product characters: the intended one, drafted and desired by the designer, and the apparent one, subjectively experienced by the user. The term "product character" refers to a higher level description of a product which summarizes its attributes, such as being novel, interesting, useful, or predictable. These attributes can be either pragmatic or hedonic. Upon contact with a product, a process is triggered that ultimately leads to consequences, namely a judgement of the product by the user (cf. Figure 2.1).

[6] This chapter does not represent an exhaustive overview of all existing UX models but rather describes the most influential ones. Other models exist but will not be discussed in this dissertation.

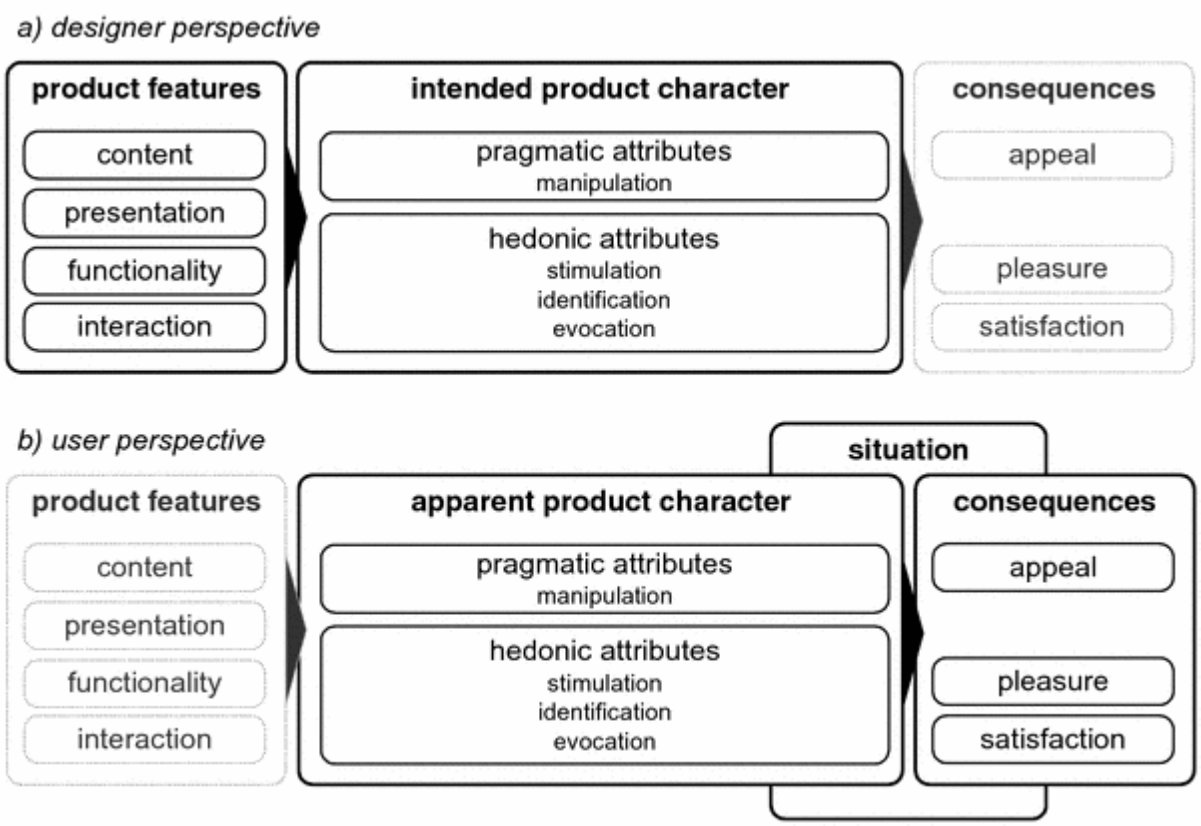


Figure 2.1: Key elements of the model of user experience from (a) a designer perspective and (b) a user perspective (Hassenzahl, 2005).

The intended product character is the result of a designer's selection and combination of product features in order to convey a certain strategy for handling the product and to reduce cognitive complexity. However, the intended product character is nothing more than an intention, since there is no guarantee that a user will in fact perceive and appreciate the product in the way the designer intended. In reality, the apparent product character is the result of a construction process, triggered by a user's contact with a product and the personal and subjective perception of the product features. The users construct their own version of the designer's original intended product character. Elements that influence the apparent product character are a user's personal standards and expectations, derived from comparisons with other similar products. The product attributes can be broadly divided into either pragmatic or hedonic classes of attributes. These define the product character as well as underlying human needs.

Hassenzahl (2005) pools all attributes that allow for manipulation of the environment, and thus require "relevant functionality (i.e. utility) and ways to access this functionality (i.e. usability)" (p.4), into the group of pragmatic attributes. As examples, he gives adjectives such as useful, clear, or supporting. All other product attributes are grouped under the

term hedonic attributes. He characterizes their most profound quality as emphasizing an individual's well-being and clearly connects all these attributes to the emotion of pleasure, expressing his belief that those attributes have a strong ability for fostering this emotion. He further divides the hedonic attributes into those providing stimulation, communicating identity, or provoking valued memories. All attributes that foster the increase of knowledge and the development of skills are described as stimulation. Those promoting social recognition and power are grouped under the label identification. Evocation assorts all attributes that represent past events, relationships or thoughts that are important to the user.

One product having several product characters for diverse users can be explained in terms of variable standards, comparisons and expectations. Time is identified as an influential factor, too. During interaction, users learn more about the product and comparable ones. The apparent product character leads to consequences, which is a judgement about the product's appeal, as well as emotional consequences and behavioural consequences. In other words, experiencing a product determines whether a user likes it, whether he experiences pleasure or displeasure, and ultimately whether he wants to go on using the product or abandon it. A moderating effect in this process can be attributed to the context in which the interaction takes place. Possible consequences of the interaction can be satisfaction, pleasure, and attraction. Satisfaction is associated with fulfilment of user expectations, pleasure with the occurrence of a desirable event, and attraction with the triggering of positive emotional responses. Special attention needs to be paid to the situation in which a product is used. To clarify personal conditions in certain situations, Hassenzahl defines two usage modes instead of defining situations: In the goal mode, users need to reach a certain goal during interaction, whereas in action mode, the journey of interaction is the reward. The momentary fit of the usage situation to the product character defines the emotional outcome of the interaction. Thus, "usage modes become the moderator between the product character and consequences" (Hassenzahl, 2005, p. 40). For Hassenzahl, UX is the combination of product attributes and the usage mode.

2.2.2 CUE MODEL

Another model that specifies the major components of UX and its interrelations is the Components of User Experience model (CUE model) by Mahlke and Thüring (2007). The model distinguishes between task-related and task-unrelated product qualities. Additionally, users' emotions are integrated into the model as a crucial mediating factor for the consequences of use like acceptance or overall rating. Mahlke and Thüring argue that UX develops through the interaction with a product that takes place in a particular

context, over a period of time, and is aimed at reaching a certain goal. Several characteristics of the user and features of the system influence the interaction and thereby influence the user's perception of the system qualities. Mahlke and Thüring argue for a broader perspective on HCI that defines UX as a compound of three elements: 1) the perception of instrumental qualities, 2) the perception of non-instrumental qualities, and 3) the emotions resulting from interaction with the product. They distinguish between the first two first qualities as follows:

> Instrumental qualities concern the experienced amount of support the system provides and the ease of its use. Features, such as the controllability of the system behavior and the effectiveness of its functionality, fall into this category. Non-instrumental qualities, on the other hand, concern the look and feel of the system. Features, such as visual aesthetics or haptic quality, belong to this class. (p. 916)

These two qualities are likely to affect the emotions a user experiences when interacting with a product. These emotions form the third element of UX. Mahlke and Thüring define emotions as "…episodes of subjective feelings accompanied by specific physiological reactions and expressive behavior" (p11). Together, these three components of UX determine the user's overall appraisal of the system and can influence future decisions and behavior (see Figure 2.2).

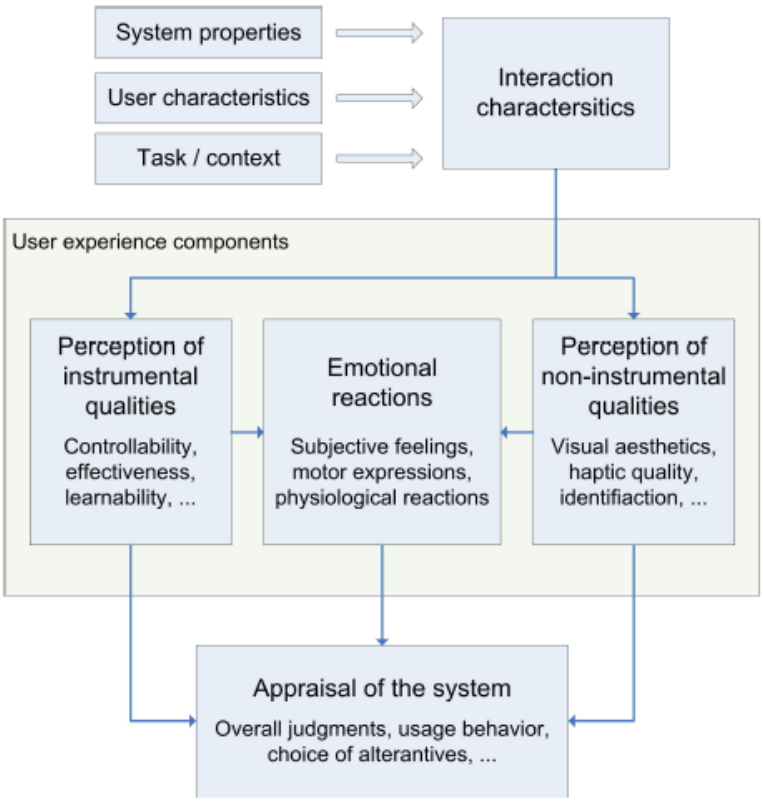


Figure 2.2: The CUE model of user experience by Mahlke and Thüring, 2007.

2.2.3 A FACET MODEL OF UX

Although there are numerous attempts to define UX which all stress different attributes, researchers' perspectives can be roughly tied together to produce a facet description of UX. Hassenzahl and Tractinsky (2006) give an overview of different perspectives on UX. They reviewed all influential publications on UX and developed a model that ties together three facets of UX which, taken separately, do not suffice to describe UX, but in combination capture the whole complexity of UX. They ascribe a human perspective to UX and stress the role of affect as an antecedent, a consequence and a mediator of technology use. Three facets are identified: non-instrumental needs should be better understood, better defined, and better operationalised. The experiential perspective on UX emphasizes two aspects of technology use: its situatedness and its temporality. Last, emotion and affect stress the importance of emotions as consequences of product use (e.g. Kim & Moon, 1998; Desmet & Hekkert, 2002; Hassenzahl, 2005; Tractinsky & Zmiri, in press), as well as their importance as antecedents of product use and evaluative judgments. Based on a summary of all facets used to describe UX in the scientific world, Hassenzahl and Tractinsky conclude that

> UX is a consequence of a user's internal state (predispositions, expectations, needs, motivation, mood, etc.) the characteristics of the designed system (e.g. complexity, purpose, usability, functionality, etc.) and the context (or the environment) within which the interaction occurs (e.g. organizational/social setting, meaningfulness of the activity, voluntariness of use, etc.). (Hassenzahl & Tractinsky, 2006, p. 95).

One key principle of Hassenzahl's and Tractinsky's view is the claim to design not merely for the absence of malfunction and adversity, but to rather actively and consciously design for pleasure and joy.

2.2.4 OTHER APPROACHES TO UX – AN OVERVIEW

As has been mentioned above, there are numerous approaches, models, and definitions of UX. The most influential ones will be described briefly.

In 2006, Schrammel et al. presented a model of UX which focuses primarily on identifying characteristics of current experiences and associated emotions. They divide UX into three parts: general experience, social experience, and personal experience.

Roto (2007) describes UX as the personal subjective feelings about a product and distinguishes between phases of UX, as well as the UX during interaction. For her, there is expected UX, which consist of information like brand image, advertisement, friends' accounts, or reports. Furthermore, there is the UX during interaction, and the overall/long-term UX. Special importance is ascribed to the interaction phase, which is seen as having the most influence on UX. Roto argues that in order to understand UX during the interaction phase, it is important to see the effect of the three components to UX as mentioned by Hassenzahl and Tractinsky (2006): user, context, and the system being used.

Karapanos, Zimmermann, Forlizzi and Maertens (2009) focus their model of UX on temporal aspects of interaction. Three phases make up the process of adoption of a new product: orientation, incorporation, and identification. The motor that powers the transition from one phase to the next is represented by three forces: familiarity, functional dependency, and emotional attachment. Depending on the phase they are in, users experience a product differently.

Forlizzi and Batterbee (2004) distinguish between three kinds of models and theories of UX: product-centred models, user-centred models, and interaction-centred models. They

focus on the latter and propose an interaction-centred framework of experience. Different user-product interactions, as well as different forms of experience are described to ultimately understand how social interaction and product use influence UX.

McCarthy and Wright (2004) propose a change in emotional design. They suggest abandoning the search for designing an emotion (and thus an experience) and advocate designing for an experience instead. Their framework for analyzing experience with technology describes four 'threads' of experience and six sense-making processes of UX.

2.2.5 SUMMARY

Since its introduction into the scientific world, there has been a debate about the scope and definition of UX. Several models and approaches have been proposed to put all related concepts in order. Although different foci are chosen, several commonalities can be highlighted. The efforts to come up with something that makes UX less fuzzy, more scalable and more comprehensible has one common motivator: the wish to understand UX is directly related to the wish to design meaningful and experiential products that are enjoyed by users.

The main focus of UX research is experience. Using various wordings (facets, threads, aspects, components etc.), shared insights about UX have been described. In summary, UX is a holistic concept that takes into account all aspects of the experience before, during, and after interaction.

- Usability is one pillar of UX. Product characteristics associated with usability are utility or functionality. It is a prerequisite for UX and is subsumed under terms like pragmatic qualities or instrumental qualities.
- Characteristics associated with an individual's well-being that are not associated with usability are the second pillar of UX. Concepts often related to personal taste like aesthetics or haptics are part of this category and are subsumed under non-instrumental or hedonic qualities of a product.
- Emotions are the third part of UX. All emotional reactions to a product can be subsumed under this category.
- Temporal aspects are important for UX: Both anticipated use based on expectations, as well as post-use reflections about the interaction are part of UX.
- The context of use also has an influence on UX.

2.3 GENERAL EMOTIONS

One pillar of UX is defined as emotions and the effect they can have on users. In order to understand emotional product design, it is important to comprehend the theoretical basis of emotions and their application in product research.

2.3.1 WHAT IS AN EMOTION?

James started a debate about the nature of emotion well over 100 years ago with an essay asking that exact question (1884). Since then, there has been extensive debate about the nature of emotion (Ekman & Davidson, 1994), its cognitive structure (Lazarus & Smith, 1988; Ortony, Clore & Collins, 1988), its psychophysical characteristics (Cacioppo, Berntson, Larsen, Poehlmann, & Ito, 2008), and most importantly, its definition (Izard, 1992; Kleinginna & Kleinginna, 1981). Defining emotions seems to be a notorious problem. One can broadly organize all existing theoretical approaches into three categories which are all based on different assumptions about the formation of emotion (cf. Cornelius, 2000; Leventhal & Tomarken, 1986). First, there are the *Darwinian* or *evolutionary* theories which focus on emotions as preconditions for survival. According to the Darwinian perspective, "…emotions are evolved phenomena with important survival functions that have been selected for because they have solved certain problems we have faced as a species" (Cornelius, 2000, para. 4). Put differently, emotions should be understood in terms of their survival value. In addition, because of the evolutionary history of humans, similar emotions should be observable in closely-related species. If we follow this train of thought, one proof for the correctness of this theoretical approach would be the existence of universal (or basic) facial expressions of emotions, both in humans with different cultural backgrounds, as well as in closely related animals. Several researchers have successfully dedicated their work to identifying these universal facial expressions and mention numbers from six (Ekman & Friesen, 1971) to ten (Izard, 1977). Emotions that are almost always included in the lists are happiness, sadness, fear, disgust, anger, and surprise (Cornelius, 2000, para. 5).

A second approach to emotion are the *Jamesian* or *body reaction* theories. Founded on William James' 1884 article, the main claim of these approaches is that it would be impossible to have emotions without accompanying physiological changes and that these changes always come first. Put differently, our bodily reactions to an event come first, and our experience of these changes constitutes and defines the emotion. Similar to Darwin, James claims that emotions are automatic responses to external stimuli that help an organism to survive. The perception of an exciting fact leads to a bodily reaction which

then leads to emotion. Thus, there is considerable overlap between these first two theoretical approaches.

Another point of view is represented by the *cognitive* or *cognitive appraisal* theorists, like Arnold (1945). Later these theories were further promoted by Schachter (1959). The main claim is the inseparability of thought and emotion. This means that emotions are the product of an appraisal process which judges events in our environment as being good or bad for us. Cognitive appraisal theories thus extend and define the process James called 'perception of the exciting fact', by looking further into what happens when 'perceiving' an exciting fact. Representatives of this approach can thus also be subsumed under the term *appraisal theorists*, who all make one common assumption: the occurrence of an emotion and its intensity towards an event solely depend on how a person interprets the event, especially in relation to his or her own goals and desires (see for example Roseman & Evdokas, 2004; Scherer, 2001; Smith & Ellworth, 1985). Nowadays, all current models of emotion argue for the use of an appraisal system (Stein, Hernandez & Trabasso, 2010).

There is broad agreement that these three perspectives on emotion formation are the most influential ones, with appraisal theories being the most current and most favoured. However, some researchers name other additional perspectives, for example the social constructivist approach (Cornelius, 2000), or the neural approach (Leventhal & Tomarken, 1986). A detailed discussion of all existing perspectives would go beyond the scope of the thesis, but it should be kept in mind that emotions and emotion research are by no means easy to grasp topics and allow for diversified contemplation, depending on the perspective one assumes.

Given the short review of the current state of emotion research above, it becomes apparent that agreeing on one definition of emotion is not an easy task. Many conflicting theories exist about emotion, and the main obstacle in the way of arriving at a mutual theory has been the disagreement over definitions. Kleinginna and Kleinginna (1981) list no less than 92 definitions of emotion which they divide into 11 categories (see the appendix of their 1981 article for a short description of each definition.). They attempt to do justice to every perspective on emotion in the following working model:

> Emotion is a complex set of interactions among subjective and objective factors, mediated by neural/hormonal systems, which can (a) give rise to affective experiences such as feelings of arousal, pleasure/displeasure; (b) generate cognitive processes such as emotionally relevant perceptual effects, appraisals, labelling processes; (c) activate

> widespread physiological adjustments to the arousing conditions; and (d) lead to behaviour that is often, but not always, expressive, goal-directed, and adaptive. (p. 355)

This working model aims to embrace all existing approaches, making it more or less unusable due to its broadness and sponginess. Another holistic definition of emotion was put forward by Matsumoto and Ekman (2009). They give a very broad definition of emotion, stating that emotion is simply one class of affective phenomena. Their more detailed definition is easy to grasp and apply. "We define emotion as transient, bio-psychological reactions designed to aid individuals in adapting to and coping with events that have implications for survival and well-being" (2009, p. 69). These two definitions nicely illustrate the struggle researchers experience when trying to incorporate all approaches to emotion in one definition.

With the rapid development in functional neural imaging technologies, some researchers argue for the predomination of the cognitive appraisal theories. Scanning the brain has led to the unification of evolutionary, Jamesian and neural theories in an approach that backs appraisal theorists in their views. In short, a distinction can be made between emotional approaches based on bodily changes and emotional approaches based on cognitive appraisals. Furthermore, when looking at the impressive body of work that has been produced about emotions, some rough agreements on the topic can be identified that will be discussed in the following paragraphs.

2.3.2 PURPOSE OF EMOTIONS

Similar to its theoretical fundaments, the practical purpose of emotions is not a unanimously agreed upon topic, either. However, it seems as if the primary purpose of emotion is to modulate or bias behaviour. When arguing in line with Darwin in his 1872 book *The Expression of the Emotions in Man and Animals*, this means emotions help us to think rationally, motivating us to engage in behaviour important for survival (as cited in Robbins, Judge, Millett & Boyle., 2013). Similarly, James (1884) argues that stating emotions are automatic responses to events, ensuring survival. Cognitive appraisal theorists have defined the purpose of emotions as bringing about a state of readiness to react to stimuli from the environment.

Concerning user-product relationship, emotions foster the attachment between product and user (Schultz, Kleine & Kergan, 1989). Product attachment is defined "as the strength of the emotional bond a consumer experiences with a specific product" (Mugge, Schoormans & Schifferstein, 2008, p. 426). Strengthening the product attachment is a way

of influencing the relationship users have with certain products, making these products more valuable to them.

2.3.3 AFFECT, EMOTION, AND MOOD

Besides the ongoing debate about the definition and underlying processes of emotion, a distinction must also be made to other similar affective constructs. Some constructs will be discussed in this section. Affect, emotion, and mood will be given special attention, since they are words that tend to be used interchangeably (Ekkekakis, 2012).

Affect

In terms of emotional constructs, *affect* or *core affect* is considered a generic term that covers a broad range of feelings. According to Robbins et al. (2013), it is an 'umbrella concept' that encompasses emotions and moods. One of the most influential works on affect is by Russell (1980, see Figure 2.3). In his Circumplex Model of Affect he defines affect around two bipolar dimensions, *arousal* and *valence*. Affect is a "neurophysiological state consciously accessible as a simple primitive non-reflective feeling most evident in mood and emotion but always available to consciousness" (Russell & Feldman Barrett, 2009, p. 104). It can remain constant or undergo rapid changes and is subject to many underlying factors, both conscious and unconscious (Russell, 2003). In Russell's model, affective concepts are arranged along a circle in the following order: pleasure (0°), excitement (45°), arousal (90°), distress (135°), displeasure (180°), depression (225°), sleepiness (270°), and relaxation (315°). The horizontal axis represents pleasure and displeasure, whereas the vertical axis represents arousal and sleepiness (see Figure 2.3).

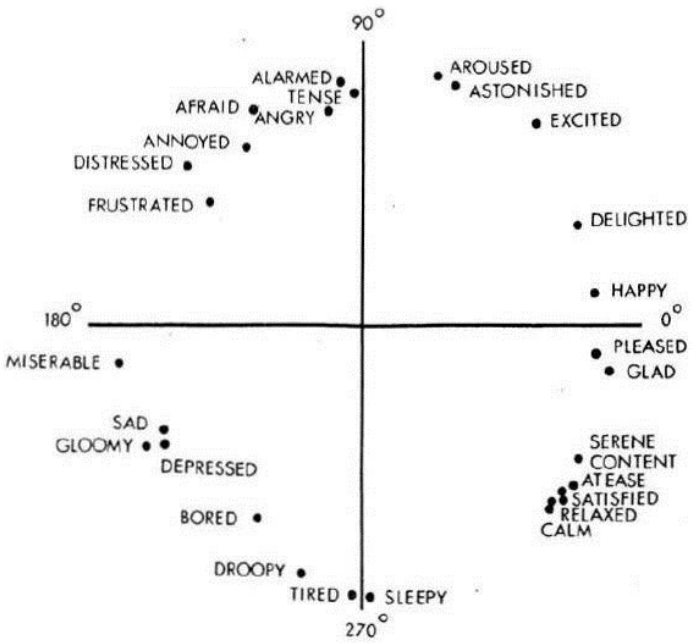


Figure 2.3: The circumplex model of affect by Russell (1980).

Feelings require little cognitive evaluation and play an important role in indicating the changes that occur either in the environment or in the body. Affective responses precede emotional responses. They do not require higher-order cognitive responses and do not encode a plan of action (Stein et al., 2010).

Examples of core affect include pleasure and displeasure, tension and relaxation, energy and tiredness. A person experiences core affect constantly, although the nature and intensity of it can vary over time. Core affect can be a component of emotions and moods (defined next), but it can also occur in pure, or isolated, form. For example, according to Russell (2003), "pride can be thought of as feeling good about oneself. The 'feeling good' is core affect and the 'about oneself' is an additional (cognitive) component" (p. 148).

Distinguishing mood and emotion

Mood is another affective construct which is sometimes confused with emotion by laypeople and researchers alike. Emotions and mood must be distinguished. Characteristics commonly used to distinguish emotions from moods are duration, facial expressions, and antecedent events. Emotions usually last shorter than moods (Ekman & Davidson, 1994; Beedie, Terry & Lane, 2005), have accompanying distinct facial expressions (Ekman & Friesen, 2003), and are usually preceded by recognizable antecedent events; i.e. they are triggered by something (Davidson, 1994). Although there has been debate about duration as a distinguishing characteristic of emotion and mood (Lazarus,

1994), most researchers agree on the specificity of their respective triggers. As Frijda put it in 1993, emotions are about something distinct, a special event or an object. Moods on the other hand are "...about nothing specific or about everything: They are about the world in general" (as cited in Frijda, 2009, p. 258). Similar to the struggle researchers experience when defining emotion, the distinction between emotion and mood is not the subject of unanimous agreement. What is important is the insight that moods are commonly the after-effect of pleasant or unpleasant emotional events (Frijda, 2009). Put differently, emotions can turn into moods when you lose the focus on the event or object that triggered the emotion. Moods may always be present, whereas emotions are brief bursts occurring on top of these ever-present affective states (Davidson, 1994).

Other affective constructs and conclusion

Besides emotion and mood, there are other affective constructs that are discussed in the literature. *Temperament* is considered a biologically based bias determining personality profiles which in turn influence emotional dispositions (Kagan, 2009). Emotional episodes are usually described as a sequence of behavior and feeling modes that all deal with the processing of one event. For example, apprehension of an event can grow into alarm, subsequent anger and finally, calm watchfulness (Frijda, 2010).

In summary, affect, emotion, mood, temperament, and emotional episodes have all been used to describe different aspects of affective phenomena (Davidson, 1994). They can be distinguished by criteria like duration, intensity, and distinctiveness of their triggers. Emotions are often the antecedents for moods, whereas affect can be seen as an umbrella term for all affective states we can find ourselves in. The objective of this dissertation is to identify possible effects of emotional design on UX of digital products. For this purpose, it is important to understand the distinction between emotion and mood, and their influence on each other.

2.3.4 HOW TO MEASURE EMOTIONS

According to Scherer (2005), there is no single gold-standard for measuring emotions. The best way to accomplish a holistic measurement approach would thus be the measurement of all aspects associated with emotion:

> In other words, in an ideal world of science, we would need to measure (1) the continuous changes in appraisal processes at all levels of central nervous system processing (i.e. the results of all of the appraisal checks, including their neural

> substrata), (2) the response patterns generated in the neuroendocrine, autonomic, and somatic nervous systems, (3) the motivational changes produced by the appraisal results, in particular action tendencies (including the neural signatures in the respective motor command circuits), (4) the patterns of facial and vocal expression as well as body movements, and (5) the nature of the subjectively experienced feeling state that reflects all of these component changes. (p.709)

Such an approach is far too elaborate to be completely implemented. However, there have been advances in measuring isolated aspects of emotions, thereby arriving at a valid explanatory power about the emotion at hand (see Scherer (2005) for a detailed description of different measurement approaches). The focus of this dissertation will be on the measurement of points 3, 4, and 5 of Scherer's holistic measurement framework by collecting data about reaction times (RT), facial expressions, and experienced emotions (questionnaire data).

2.3.5 CONCLUSION

It might seem that there is still no consensus about many topics of emotion research, although the debate has lasted for over 130 years. On closer inspection however, it seem as if a universal agreement on what emotions are, does exists, since "most contemporary emotion theorists view emotions as coherent, organized, and functional systems" (Desmet, Porcelijn & van Dijk, 2007, p. 141). They mostly agree on the Darwinian notion that emotions serve some sort of adaptive function, namely signalling the mind whether something is favourable or harmful to one's concerns. This appraisal of information is the common notion of contemporary emotion theorists. Accordingly, distinct emotions are associated with distinct appraisal patterns and can thus be predicted from the underlying concern of a person. As a consequence, emotions are interpreted as the result of a person's concerns in combination with that person's appraisal of new information. Put differently, a cognitive appraisal process is the prerequisite for evaluating a stimulus and experiencing an emotion.

What has to be kept in mind is that emotions are not the same as moods. Emotions are mostly shorter than moods, elicited by distinct events, accompanied by facial expressions, and serving as probable antecedents for longer lasting moods. In the domain of interaction design, moods and emotions are regarded as closely related: emotions can influence

moods, in turn moods can influence the probability of a certain emotion being triggered (Davidson, 1994). Users like products that elicit positive emotions. Moods, as well as emotions, should therefore be considered in the overall design process (Spillers, 2010).

2.4 EMOTIONAL PRODUCT DESIGN APPROACHES

Because of their impact on users, emotions are an important part of product design. Accordingly, emotional design is the analysis of products in a holistic way to include their behaviour, their attractiveness, and the image they present to the user in their design (Norman, 2004). One of the main tendencies of this approach is to design for positive emotions. Fun and enjoyment are the core emotions many designers strive to evoke through their product design. Monk et al. (2002) coined the term *funology* to describe this phenomenon. The investment in emotional design has been proven to be a promising strategy: when using complex interactive technology, positive emotions have been shown to decrease usage anxiety (Helander & Tham, 2003) while contributing to overall usability (Tractinsky, Katz & Ikar, 2005), perceived ease-of-use, and perceived usefulness (Mun & Hwang, 2003). In other words, "products that evoke positive emotions are bought more often, used more often, and are more pleasurable to use." (Desmet, 2012, p.1). Good emotional design can increase return-on-investment, and productivity (Bias & Mayhew, 2005), as it can improve social relationships and overall well-being (Klein, Moon & Picard, 2002).

To expand the importance of positive emotions in relation to product design, HCI contributors have promoted several design approaches for designing positive and pleasant experience.

2.4.1 PLEASURABLE DESIGN APPROACHES

Jordan (1997) proposes that a positive affect, namely pleasure, is a source of a pleasant user-product relationship. To describe the evolving users' demands for technology that goes beyond usability, Jordan defines a hierarchy of consumer needs (Jordan, 1997, 2003), based on Maslow's hierarchy of needs (Maslow, Frager, Fadiman, McReynolds & Cox, 1970). In his view, functionality is the basis of consumer needs, followed by usability, and pleasure. Following his argumentation, consumers expect products to work well and function flawlessly, which in turn should bring designers to think about designing additional pleasurable experiences (Jordan, 1997). Based on the general framework of pleasure by the anthropologist Lionel Tiger (1992), he identifies four main pleasures which

should be considered when creating pleasurable products: *physio-pleasures* are experienced through the senses, i.e. they are associated with the sensual experience of a product. *Socio-pleasures* are associated with the pleasure derived from the social company of others and social relationships through product use. *Psycho-pleasures* are pleasures generated by reaching one's goals or accomplishing a task with the help of a product, which in turn reduces cognitive demands. Finally, *ideo-pleasures* pertain to people's values, which can be associated with aesthetics or the value a product signifies for a person (Jordan, 1997).

Similarly, Hancock, Pepe and Murphy (2005), propose the term *hedonomics* to define "...that branch science and design devoted to the promotion of pleasurable human-technology interaction" (p. 8). Besides a more basic level of safety, they add a level of individuation to the hierarchical model of consumer needs. This level concentrates on the possibility to individually customize a product to optimize pleasure and efficiency of an interaction. Figure 2.4 shows a combination of both hierarchical models.

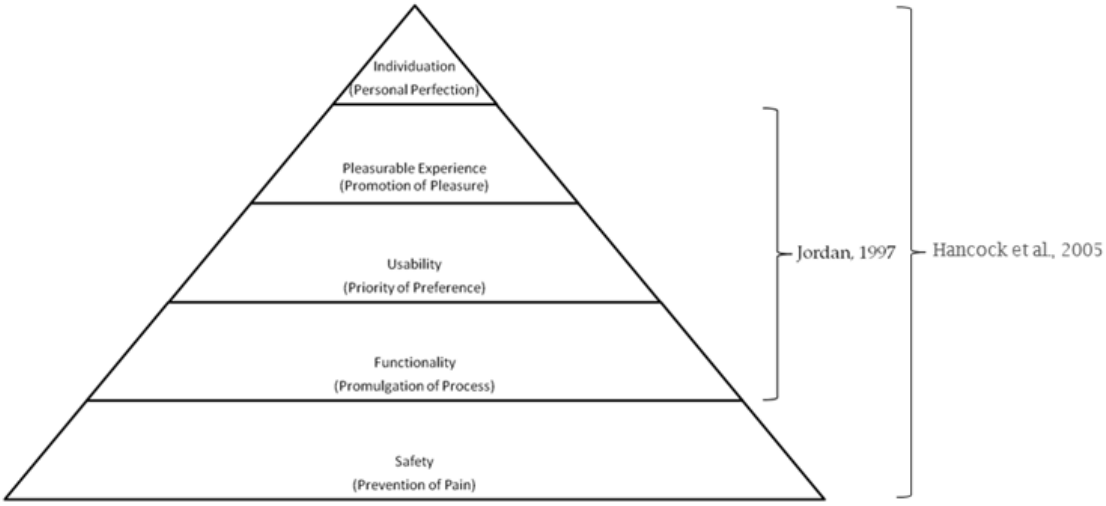


Figure 2.4. Hierarchy of consumer needs (Jordan, 1997; Hancock et al, 2005).

2.4.2 NORMAN'S MODEL OF EMOTIONAL DESIGN

In 2002, Donald Norman called on the scientific and professional communities to stop being 'usability bigots' and to start taking affect and pleasure into account when designing products (Norman, 2002). Progressing from his original request for usable design (Norman, 1988), he stressed the need for pleasurable design, pointing out that "usable designs are not necessary pleasurable ones" (Norman, 2002, p. 38). By explaining why usability cannot be the only component of good design, he argues that pleasing products have manifold advantages: according to Norman, they "work better, are easier to learn,

and produce a more harmonious result" p. 38). The main contributor to these positive phenomena is affect. Norman explains that affective signals always take the fast lane in our cognitive system, changing the way we judge things independently from conscious thoughts. Norman's simple claim is that affective things work better; as long as an adequate level of usability is still included in the product.

Another important factor in successful emotional design is the context. Which design is preferable in a given situation depends on the use-context. Norman argues that in neutral situations, a poor interface design could be more easily drowned out by pleasant and pleasurable design than in a stressful situation. In stressful situations, joyful design cannot be able to make up for basic interface flaws. In other words, to be truly beautiful, wondrous and pleasurable, the product has to fulfil a useful function, work well, and be usable and understandable, all in the right context.

Norman holds the view that there are three levels of product experience, and that accordingly, there should be three levels of design. The *visceral level* is the most basic, pre-conscious one, taking only product appearance into account to form first impressions. Visceral design should therefore aim to leave a positive first mark. The *behavioural level* is all about using the product. Behavioural design should therefore concentrate on performance. Norman identifies four aspects of good performance: usability, function, understandability, and physical feel. The last and cognitively most demanding level is the *reflective level*. This includes self-image, personal satisfaction, and memories. Reflective design is much more subtle than design at the first two levels, but is equally important, since it determines the lasting value of a product for its owner.

2.4.3 DESMET AND HEKKERT'S BASIC MODEL OF PRODUCT EMOTIONS

Desmet (2002) and Desmet and Hekkert (2002) establish a basic model of product emotions that explains underlying processes of developing a response to a product (see Figure 2.5). By identifying the underlying cognitive processes in product experience, the model can be used to further explain the character of product emotions. They base the approach on the cognitive model of Ortony et al. (1988) which names three aspects of the world people are focused on: events, agents, and objects. They define emotions as being merely valenced reactions to one of these aspects. Based on this approach, the model of Desmet and Hekkert (2002) consists of four parameters, *appraisal*, *concern*, *product*, and *emotion*. The interplay of the first three components determines whether a product evokes an emotional response. It further determines what kind of emotion is elicited. Not the event itself but

rather its personal significance to a person determines what emotion will be the result of the appraisal. Concerns are reference points in the appraisal process, the preference for a certain state of the world (Frijda, 1986). We constantly appraise our concerns and find matches or mismatches which in turn determine whether our wellbeing is promoted by a product: if our concerns are matched, the product is appraised as being beneficial. Products that do not match our concerns are appraised as being harmful. Emotional meaning develops in the context of a person's concerns, only.

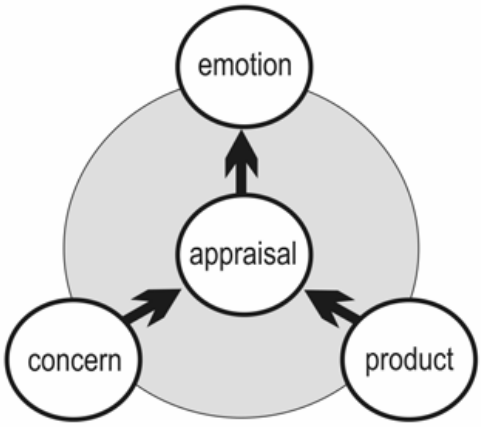


Figure 2.5. The basic model of product emotions (Desmet, 2003)

Desmet (2003) distinguishes five product emotions (see Figure 2.6). *Instrumental emotions* represent the user's wish to fulfil a goal or motive by using a product. *Aesthetic emotions* involve aspects of liking or disliking a product based on personal predispositions and attitudes. *Social emotions* correspond to values and standards every person holds. Because we cannot separate our values and norms from how we appraise products, we appraise them in terms of legitimacy. *Surprise emotions* are special emotions as they are preceded by an appreciation of novelty, i.e. sudden unexpected concern mismatch. *Interest emotions* are emotions like fascination, or its opposite boredom and they are all elicited by an appraisal combination of challenge and promise.

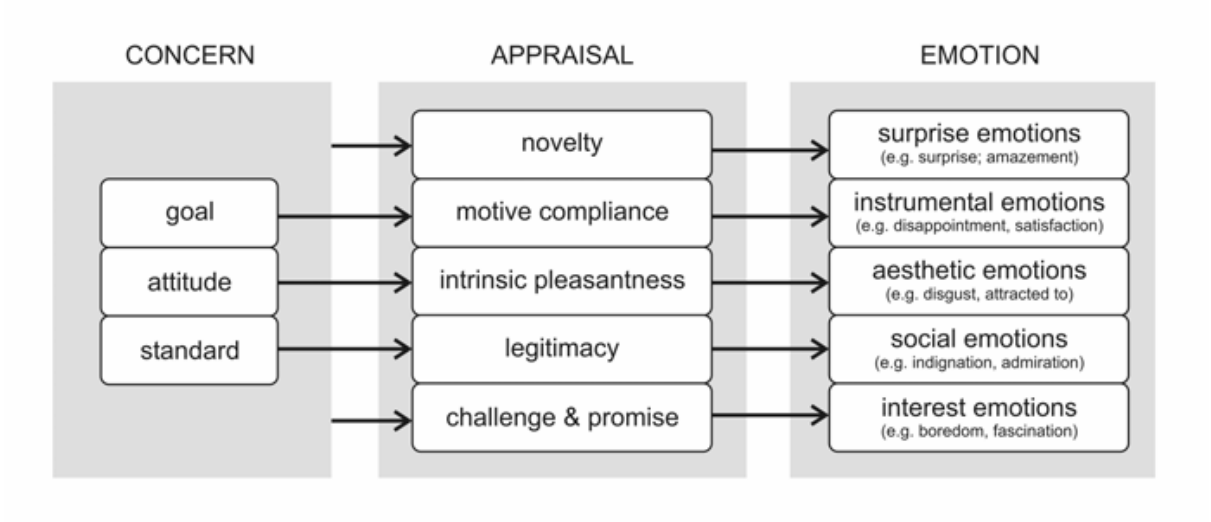


Figure 2.6. Classification of product emotions (Desmet. 2003)

In summary, Desmet and Hekkert developed a model for product experience and identified five main product emotions that are the result of the interplay of concerns and appraisals. They paint a more detailed picture of product emotions than other authors (Jordan, 1997, 2003; Norman, 2002), including pleasant and unpleasant emotions in their model. Following Frijda in his suggestion that paradoxical emotions can be beneficial to experiencing art work (1986), they propose to investigate the possibilities of designing paradoxical emotions (elicit positive and negative emotions simultaneously) instead of purely positive and pleasant ones. They hope products will result which are "unique, innovative, rich, challenging – and therefore, desirable" (Desmet, 2003, p. 11).

2.4.4 OTHER APPROACHES TO EMOTIONAL PRODUCT DESIGN

Emotions mentioned in the design approaches named above mostly highlight in some way the importance of creating a positive affect when the user interacts with a product. Besides the generation of positive emotions, like pleasure or enjoyment, some researchers argue for the inclusion of other emotions, like confusion, surprise, anger, or disgust, to achieve an interesting and outstanding product design (Demir, 2008; Desmet, 2012; Fokkinga & Desmet, 2012, 2013; Ludden, 2008; Silvia, 2009).

Demir (2008) describes emotions as one integral element of design in the process of convincing a user to accept products in their lives and to increase their quality. He states that single emotional moments can accumulate into one long-lasting emotional relationship between a user and a product. This process is comparable to the reciprocal influence of moods and emotions (see section 2.3.3).

In his later work, Desmet (2012) reiterates that a main limitation of most approaches and frameworks to product experiences is the ignorance of all emotions that are not purely positive and pleasant. He claims that products can elicit far more complex positive emotions than generalized pleasure, for example pride, contentment, admiration, desire, relief, or hope (Desmet, 2002; Desmet & Schifferstein, 2008). In his 2012 article, he proposes a set of positive emotions and discusses how these can be experienced during human-product interaction. Based on existing emotion typologies, a componential analysis of 150 positive emotion words produced a set of 25 positive emotions, clustered in nine general categories (empathy, affection, aspiration, enjoyment, optimism, animation, assurance, interest, and gratification). In a questionnaire study, 729 example cases were collected that illustrated when each of the 25 emotions are experienced during human-product interaction. Thus, he identifies six main sources of product emotions (the object, the meaning of the object, the interaction with the object, the activity that is facilitated by this interaction, the user, and others involved in the interaction). Accordingly, Desmet (2012) discloses the possibility for designers to compose products with elevated and intricate emotional significance for users. He showed that it is not only possible, but even beneficial, to design product emotions other than pleasure and joy, accommodating users' complex emotional states and needs when interacting with products.

Desmet's proposition is an example of contemporary approaches to emotional product design which open up perspectives beyond the ordinary pursuit of pleasure and joy as a design goal. Still, they only focus on the creation of positive emotions. Advancing from its beginnings in pleasurable product design, other researchers have found ways to include initially unpleasant, negative emotions in their design frameworks, while still generating pleasant interactions and positive emotions as the outcome of human-product interaction.

Fokkinga and Desmet (2012, 2013) explicitly argue for the inclusion of negative emotions to give users a richer experience. They argue that emotionally rich experiences "are notable and memorable experiences that involve a mix of positive and negative emotions" (p. 20). They further argue that in order to have such a memorable and beneficial rich experience, negative emotions can foster mental and bodily effects which in turn can lead to a positive overall experience. They propose a framework which helps to generate rich experiences from negative emotions. Three prerequisites are necessary: 1) a negative stimulus, 2) subjective transformation, and 3) a protective frame. The negative stimulus evokes the negative emotion, whereas the subjective transformation must take place in order to change what a person attends to. In a final step, the protective frame is essential for the transformation of a negative emotion to an experience that can be enjoyed by the user. It gives the user security and removes unpleasant aspects of the negativity. After

testing their assumption within six design prototypes, the authors conclude that negative emotions are "a viable and interesting starting point for creating emotionally rich product experience" (p. 19).

Silvia (2009) discusses aesthetic experience, which to him is a "simple feeling of liking, preference, and pleasure" (Silvia, 2009, p. 48). He challenges the usual ways of measuring these feelings that focus on how much someone subjectively likes a product, using questionnaires that exclusively address on joy, liking, or pleasantness. He discusses the possibility of including unusual aesthetic emotions into the current focus of research on product liking and user experience. In order to do so, Silvia views appraisal theories of emotion as offering new and interesting angles to product aesthetics because they open up focus to a wide range of emotions that can be studied. This opens the way for studying unusual aesthetic emotions, apart from basic pleasure/displeasure or liking/disliking.

Silvia distinguishes between three groups of emotions. Knowledge emotions comprise all emotions that are connected to comprehending and thinking. Members of this class are interest, confusion, and surprise. Their appraisals are rooted in people's appraisals of what they know, what they expect, and what they might learn and understand. Additionally, these emotions promote the growth of knowledge by motivating learning, thinking, and exploring. A second group of emotions are hostile emotions like anger, disgust, and contempt. They are characterized by an appraisal process of goal incongruence in conjunction with the belief that the emotion has been deliberately caused. A third group of emotions are the self-conscious emotions, e.g. pride, shame, guilt, regret, and embarrassment. In contrast to hostile emotions, they are the result of goal incongruence that has been caused by oneself, rather than by others. According to Silvia, the inclusion of unusual aesthetic emotions is essential in order to encompass the diversity and complexity of everyday user experience. To him, complex emotions are a fertile territory for enterprise research.

2.5 A SPECIAL EMOTION: SURPRISE

Similarly to emotion itself, there has been an extensive debate about one distinct emotion, namely surprise, its underlying processes (Lorini & Castelfranchi, 2007) and definition and theories (Macedo, Cardoso, Reisenzein, Lorini & Castelfranchi, 2009). See Russell (1980) and Kleinginna and Kleinginna (1981) for a detailed discussion of the problem. The first mention of surprise, as well as first theoretical approaches, can be dated back as far as

Aristotle (ca. 350 B.C.). Since then, it has been a topic of interest, especially after the 'cognitive revolution' of the 1960s (Macedo et al., 2009).

Surprise constitutes a special kind of emotion. Following Russell's two dimensional organization, it is high in arousal and neutral in valence (Russell, 1980). Since the feeling of being surprised itself is not linked to a specific valence, some researchers take this lack of affective characteristic as proof of the fact that surprise cannot be identified as one (Ortony et al., 1988). Other authors regard surprise as an emotion, since it is almost always followed by a valenced reaction, resulting in a pleasant or unpleasant surprise (Russell, 1980; Silvia, 2005). Ekman describes surprise in his book *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life*:

> Surprise is the briefest of all the emotions, lasting only a few seconds at most. In a moment surprise passes as we figure out what is happening, and then surprise merges into fear, amusement, relief, anger, disgust, and so forth, depending upon what it was that surprised us, or it may be followed by no emotion at all if we determine that the surprising event was of no consequence. (2007, p.148)

This description illustrates the three possible forms of surprise: neutral, pleasant, and unpleasant. To distinguish between different emotions, theorists agree on taking one of two approaches: the categorical or the appraisal one. Categorical approaches separate emotions by defining 'basic emotions', which are cross-cultural, universally recognizable and fundamentally different from each other, e.g. fear, happiness, anger, or shame. They furthermore identify emotions that do not possess these universal characteristics and may be a mix of two or more basic emotions, like hope or pride (Ekman & Friesen, 1971; Izard, 1992; Smith & Ellsworth, 1985). Appraisal theorists on the other hand, see emotions as the result of an individual's evaluation and interpretation (e.g. appraisal) of events in an environment. In the case of surprise, these appraisals have been associated with unexpectedness, novelty, unfamiliarity, amazement, or pleasantness (Reisenzein, 2009; Roseman & Evdokas, 2006; Scherer, 2005).

2.5.1 THE SURPRISE PROCESS

A schematic appraisal model that describes the surprise process in terms of beliefs, expectations, attention, and schemas is the *cognitive-psychoevolutionary model of surprise*, proposed by Meyer, Reisenzein and Schützwohl (1997), see Figure 2.7.

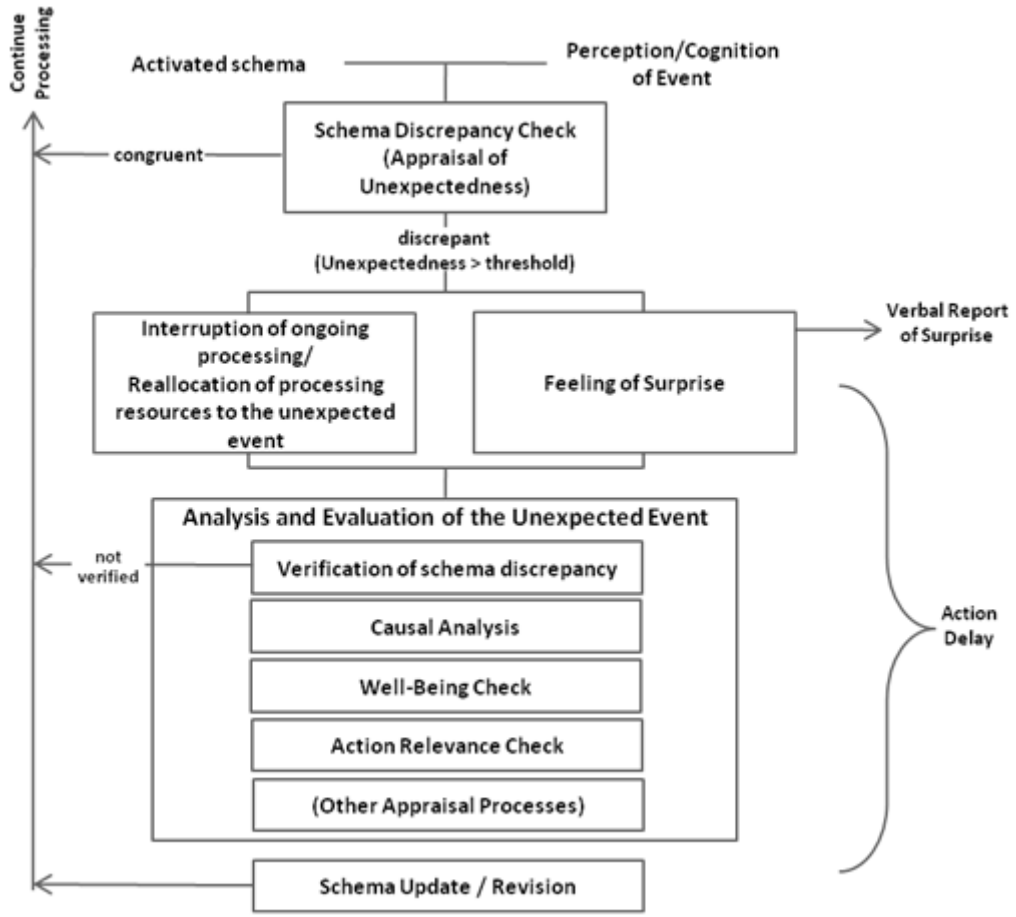


Figure 2.7: The cognitive-evolutionary model of surprise (Meyer et al., 1997).

The model takes schemas as the starting point of a four step-surprise process. Schemas are defined as "informal, unarticulated theories [...] about objects, event sequences (including actions and their consequences) and situations. Schemas serve the interpretation of present and past, and the prediction of future events, and thereby the adaptive guidance of action" (Macedo et al., 2009, p. 270). As schemas cannot always be entirely correct, they need to be continuously monitored and if necessary, updated. This is accomplished by the surprise mechanism, consisting of an interruption of ongoing processes and the subjective feeling of surprise (middle part of Figure 2.7; Meyer et al., 1997). The function of the surprise process is to detect schema discrepancies (i.e. incongruence between schemas and newly acquired information). In the event of such incongruence, the processes of event analysis and schema revision are enabled, which in turn lead to the interruption of all ongoing information processing ('action delay'). As a last step, the new information is integrated into the original schema, leading to a schema update.

The analysis and evaluation of surprising events can be broken down further into several consequential steps. First, a schema discrepancy is verified. Second, the causes of the unexpected event are identified. Third, its significance for one's own well-being is evaluated, and fourth, its significance for the ongoing action is estimated. All these processes classically lead to an observable interruption in the original behaviour, which directly translates to prolonged reaction times (Meyer, Niepel, Rudolph & Schützwohl, 1991; Meyer et al., 1997; see section 2.5.2 for more details on this phenomenon).

2.5.2 SURPRISE AND REACTION TIMES

In a sequence of four experiments, Meyer et al. (1991) investigated the effects of unexpected and thus surprising events on reaction times (RTs) and recollection of stimuli. The experiment involved a forced choice task: participants had to press one of two buttons depending on the location of the stimulus presented. Stimuli consisted of pairs of words. Simultaneously to the words, a dot appeared above or beneath the words, determining which button should be pressed. RTs, position of the dot, recollection of the words from trial 30 (appearance of an unexpected event), as well as subjective surprise ratings (on a scale of 1=very weak to 9=very strong) were analyzed as dependent variables.

In the experimental group, the inverted display of words led to an increase of RT; in the control group, a decrease of RT was observed which was attributed to a practice effect. The recall of the words from trial 30 was significantly better in the experimental group.

One model that accounts for delays in RTs, focusing of attention, and the subjective experience of surprise when being confronted with an unexpected event, is the cognitive psycho-evolutionary model of surprise by Meyer et al. (1997, see section 2.5.1). Because new and schema-discrepant information is integrated into the model, an identical stimulus that is presented several times should most likely fail to elicit surprise once again. Besides this schema-update of newly acquired information, an action-relevance check is also part of the surprise process. If an unexpected event is relevant to the activity one was engaged in, this action-relevance check will be more extensive, resulting in longer RTs.

Meyer et al. (1997) tested these two assumptions with a simple RT experiment. Participants had to press a button depending on the position of a target stimulus (a circle) on a display with three frames (top frame=left button, bottom=right button). The unexpected event was implemented through a reversal of figure and background colour (Figure 2.8). When the colour change concerned the target stimulus, longer RTs were expected than when it concerned the distractor stimulus, due to the action-relevance check mentioned above.

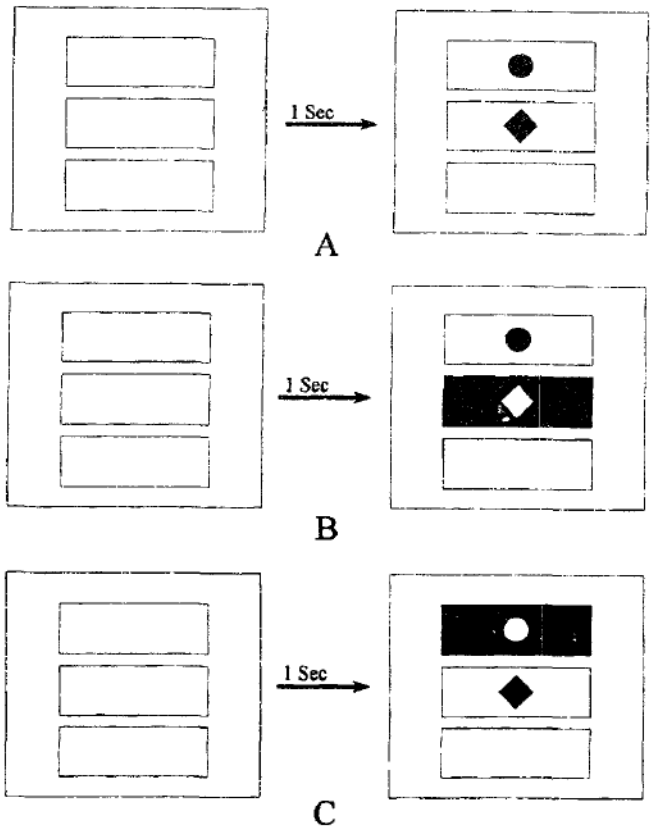


Figure 2.8: Stimulus events used in the study by Meyer et al., 1997.

Participants who experienced two different events indicated significantly higher subjective surprise ratings than participants with two identical events (4.3 vs. 3.6). This result is consistent with the notion of a schema updating process as the final step of the surprise process: participants who experienced the same event twice are less surprised by the second event, because they have stored the new information in their schemata for later reuse ('schematization', Meyer et al. 1997, p. 272). Concerning RT data, a significantly longer reaction to the first unexpected event compared to a baseline could be confirmed. These results support the notion of an action-relevance check, as well as the schematization of new information. According to the findings of Meyer et al. (1991), RTs can be used as a reliable tool to objectively measure surprise reactions in users.

2.5.3 PLEASANT AND UNPLEASANT SURPRISE

As mentioned above, surprise can be seen as a neutral emotion, or as an emotion that can convey either a pleasant and unpleasant value. The cognitive psychoevolutionary model mentioned earlier also serves as the basis for Reisenzein's Belief Desire Theory of Emotion

(BDTE), which defines the emergence of surprise in terms of expectations and desires about the occurrence of an event, taking personal wishes into account, and thereby accounting for the formation of pleasant and unpleasant surprise (Reisenzein, 2009). When an unexpected event is congruent with what a person has wished, the result is a pleasant surprise. If an event mismatches a person's wishes, the result is an unpleasant surprise.

Consider the new smartphone you just bought. You expect it will be easy to synchronize all your contacts. Your friends told you about how smoothly it went for them and all test reports in magazines praised the outstanding ease of this process. So you open your account and as you press synchronize you are confronted with an error message, saying all your contacts have been lost. You are surprised and then you will feel a negative emotion like frustration or anger. You are experiencing unpleasant surprise because your expectations were not met and your desires were not satisfied. Now imagine the opposite situation. Your smartphone broke and a friend gives you his old one. You remember him telling you how annoying the synchronization was and how he had to enter contacts manually. So you open your account and you press synchronize and all your contacts appear in your phone, nicely sorted and complete. You are surprised and then you will feel a positive emotion like joy or amazement. Now you are experiencing pleasant surprise because something unexpected happened that you did desire. Figure 2.9 depicts the schematic processes of the formation of pleasant and unpleasant surprise according to Reisenzein (2009).

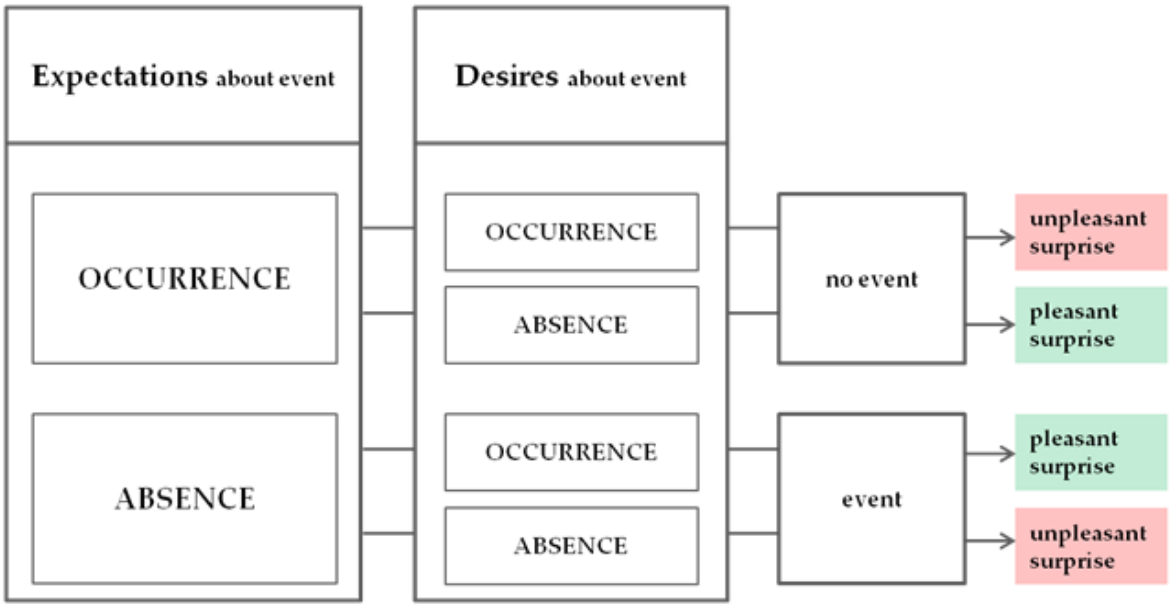


Figure 2.9: Formation of pleasant and unpleasant surprise according to Reisenzein (2009).

2.5.4 DIFFERENT FORMS OF SURPRISE

There is an on-going debate about the different sorts of surprise and their sources. Lorini and Castelfranchi (2007) distinguish between *mismatch-based surprise* and *inference-based astonishment*. These two forms of surprise can be discriminated based on the different underlying expectations.

Mismatch-based surprise

Mismatch-based surprise is the outcome of being surprised due to the realization of a conflict between what is experienced and what is expected. The active search for the appearance of a certain event is crucial to this form of surprise. If an event is anticipated, incoming data is matched against this expectation. If a mismatch is detected, surprise is the result. The intensity of the surprise reaction can be deducted from the probability assigned to the expectation that has been frustrated: the stronger the expectation was, the stronger the surprise reaction will be (Lorini & Castelfranchi, 2007). The underlying expectations are active: they occupy consciousness and require attention. Lorini and Castelfranchi call these *expectations and beliefs under scrutiny*.

Inference-based astonishment

Inference-based astonishment is the result of perceiving an event and recognizing its implausibility. It is important to note that there were no prior expectations about the perceived fact - nothing was expected, no prior knowledge was activated (Lorini & Castelfranchi, 2007). In this case, the underlying expectations are passive: they are available at an automatic level, not requiring any conscious attention. Lorini and Castelfranchi call these expectations background expectations and beliefs. They draw on Kahneman and Tversky's distinction between active and passive expectations (see Kahneman & Tversky, 1982), describing active expectations as occupying consciousness, and passive ones as automatic and effortless.

2.5.5 EXPECTATIONS AND UNCERTAINTY

Expectations are mental constructs "generated from past and current experience that serve to disambiguate the future" (Geers & Wellman, 2009, p. 164). Oliver and Winer (1987) mention three different sources of expectations as described by Tolman: 1) memories of actual experiences, 2) perceptions of current stimuli, and 3) inferences drawn from related experiences. Wilson and Klaren (1992) also mention three different sources: 1) direct experience from the environment, 2) communication with others, and 3) combination and

extrapolation of one's knowledge (as cited in Sander & Scherer, 2009, p. 164). Theorists mention different sources for expectation formation. However, on closer inspection they mean the same thing: expectations are formed based upon current and past experiences, and inferences drawn from that information.

An expectation always involves uncertainty, since it is not a proven fact but merely a conscious or unconscious assumption. When expectations are based on memories, uncertainty is usually lower than when they are based on inferences drawn from similar products or situations (Oliver & Winer, 1987). Furthermore, expectations differ in terms of how explicitly they are formed and therefore can be expressed, as well as in their specificity (Geers & Wellman, 2009). Surprise is classically associated with high levels of uncertainty, in contrast to anger or happiness. People are uncertain when feeling surprised about what is happening in this situation (Smith & Ellsworth, 1985).

2.5.6 FACIAL EXPRESSIONS ACCOMPANYING SURPRISE

According to categorical classifications, surprise is one of the six basic emotions (Ekman 1992; Turner, 2000). An accompanying distinct facial expression is one chore characteristic of basic emotions (Ekman & Friesen, 2003). These facial expressions communicate the emotional status of a person to the environment. Or as Frijda (1986) puts it: emotions "...communicate requests and intentions. They are there to influence the behaviour of others, and not to promote understanding. Or, to put it differently, they do not communicate anything, they *are* forms of communication" (p.60). Surprise is characterized by raised eyebrows, widely opened eyes, a dropped jaw and parted lips (Ekman & Friesen, 1971, see Figure 2.10, left side). Although surprise is expressed through these distinct facial changes, it can easily be confused with fear (see Figure 2.10, right side). However, fear is also characterized by tensioned lips, wrinkles exclusively in the centre of the forehead, and exposing of the upper sclera.

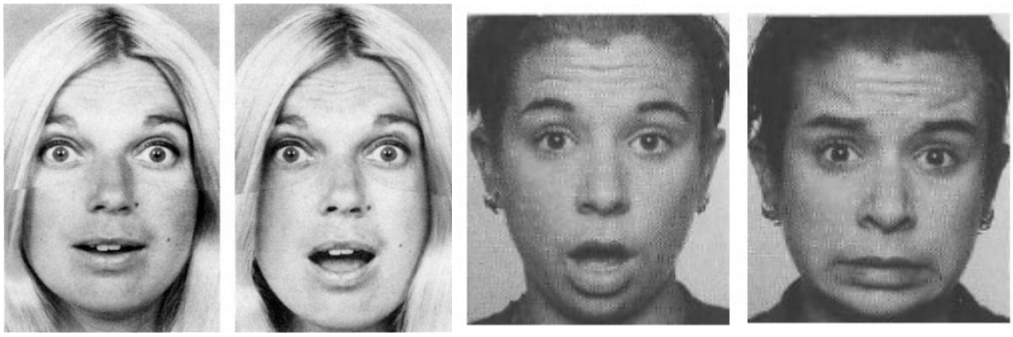


Figure 2.10. Facial expressions of surprise (Ekman & Friesen, 1971, left), and surprise and fear (Ekman, 2007, right)

2.5.7 EMOTIONS FOLLOWING SURPRISE

There is a reason why some researchers reject surprise as an emotion (Ortony et al., 1988). Surprise is per se hedonically unspecific. This claim rests on the observation that surprise can be pleasant, unpleasant, and neutral, meaning that these valenced feelings are not necessarily part of surprise but result from other processes, following surprise. It is not until the significance of an event is appraised as being positive or negative for one's well-being that surprise gets a valence. Scherer (2001) states that surprise has one core appraisal as something novel and unexpected. Other appraisals can shift the subjective feeling of surprise to another emotion like joy, fear, anger, amusement, fascination, disappointment, or irritation (Ekman, 2007; Ludden, Hekkert & Schifferstein, 2006; Ludden, Schifferstein, Hekkert, 2009; Scherer, 1987). Silvia (2009) follows this argumentation and specifies it for the case of user-product interaction by stating that during the interaction with a product, surprise can shift to other feelings like interest, amusement, or disappointment. Surprise is also prominently linked to consumer satisfaction (Vanhamme, 2000; Vanhamme & Snelders, 2003). Ludden tested the assumption that "surprise can be seen as the first stage in a sequential process of appraisals that is followed by the experience of other emotions" (Ludden et al., 2012b, p. 8). They confirmed a positive relationship between surprise and six emotions (interest, fascination, amusement, confusion, indignation, and irritation) confirmed when presenting participants with products.

In summary, surprise can be considered a catalyst for other emotions that will follow an event and other consequential appraisal processes. After an initial appraisal of unexpectedness, a surprising event is further evaluated; its causes, significance, and relevance are appraised, possibly resulting in another emotion. This evaluation of the unexpected event determines which emotion exactly will follow the surprise.

2.6 SURPRISE AS A DESIGN STRATEGY

In line with Norman (2004), it is argued that especially in an age of fast technological development and strong competing products, the emotional components of the design of a product may be more important to the user than its practical elements.

2.6.1 SURPRISING PRODUCT DESIGN

A surprise reaction to a product has been proven to yield beneficial effects for users. In his 2008 book *The New Strategic Brand Management: Creating and Sustaining Brand Equity Long*

Term, Kapferer states "the brands that ultimately last are those that are able to surprise their customers and the customers of tomorrow in particular" (p. 272). Surprise makes a product more interesting to interact with and it increases person-to-person recommendation, as well as product recall and recognition (Derbaix & Vanhamme, 2003). According to Dixon and Hudson (2004), a surprise reaction elevates a product beyond the ordinary (as cited in Ludden, 2008). In marketing research, surprise has been linked to satisfaction (Vanhamme & Snelders, 2003).

Contemporary product designers already use various strategies to create surprises in products. Employing different materials, making use of contrast and new shapes, utilizing humour, or inventing innovative functionality are all ways designers create surprising products. Figure 2.11 shows some examples: porcelain cups that mimic the looks of ordinary plastic cups, a bracelet that receives its form from a mathematical function, tableware made out of edible biscuit dough, a book shelf that plays with shapes and laws of physics, a keyboard that has a very special layout, or an 'utterly useless' wine glass.

Figure 2.11: Surprising product designs.

All these products create surprise, puzzlement, and amusement. Besides being deployed as a product property, surprise is also used in product marketing. The South Korean car manufacturer KIA uses surprise in its brand claim: "KIA – The Power to Surprise". In January 2015, the topic of surprise as a design strategy came up in a blog entry which focused on increasing the user engagement through unexpected, thus surprising interaction design (Machlovi, 2015). Examples are funny and entertaining messages to bridge loading times, and excuse error messages (slack.com), small animations to celebrate

a successful deployment of campaign e-mails (MailChimp) or reaching the bottom of homepages (photojojo.com). Adding humour to their website is also employed by GrubHub (grubhub.com), an online food ordering app; celebrating a complete order with a cartoon or making ironic remarks adds funny surprises to the ordering process. All such examples represent designs that adhere to two design recommendations made by the author of the blog entry mentioned above. 1) Surprises should respond to a user's interaction, and at the same time excite the user and motivate him to look for more surprises. 2) Surprises should complement the user's journey non-intrusively, so that they do not disturb the work flow.

2.6.2 SURPRISE AND PRODUCT EVALUATIONS

Ludden examines sensual incongruities and resulting surprise in product design. She focused her work on classical products, including items like lamps, vases, tablecloths, juicers, and rubber ducks (Ludden, 2008). Working with users' expectations, she created a mismatch between a product's look and its corresponding smell, sound, or haptics. In summary, Ludden and her colleagues sought to answer the question "But what happens when our expectations about the sensory characteristics of a product are somehow disconfirmed? (…) Will the product be more appreciated because it offers a new experience?" (Ludden, 2008, p.11). Surprises most likely affect interaction and thus users' evaluation of a product (Ludden et al., 2009). The result is a whole body of work on surprising products and their effects on users.

Surprise creation through visual-tactual incongruities

In their first two experimental studies, Ludden et al. (2009) create surprise through the implementation of visual-tactual incongruities. Those incongruities are the result of a mismatch between information received about a product's vision and about its touch. Vision and touch can be used to perceive the same properties of an object, like its shape or surface material. Expectations can be formed about how a product will feel based on its appearance. When we experience something different when touching it, the original expectations are disconfirmed. Ludden et al. (2009) anticipates users to feel surprised when experiencing the visual-tactual incongruities, which should subsequently change their behaviour and their reactions toward the product. Because the expectations that have been formed earlier are subject to uncertainty (see section 2.5.4), the familiarity of the product influences how reliable expectations are. Accordingly, Ludden et al. (2009) distinguish between *Visible Novelty* (VN) and *Hidden Novelty* (HN). VN products are unfamiliar to the user, which leads to the formation of uncertain expectations. HN

products are familiar to the user, which leads to the formation of strong expectations. Because of the visual-tactual incongruities, people are surprised when touching both products.

Accordingly, nine products were selected for presentation to users in the two studies. Three versions of three products (a VN, HN and a no novelty (NN) version) first in a 'see'-condition (manipulation check) and afterwards in a 'see and feel'-condition. An Analysis of Variances (ANOVA) of the certainty and surprise scales was carried out to investigate possible differences between VN, HN, and NN products regarding certainty and surprise ratings. The analysis showed that products with visual-tactual incongruities are able to evoke higher surprise reaction ratings. Furthermore, the two different types of products (VN and HN) are distinguishable and evoke different affective and behavioural reactions. Significantly more exploratory behaviour was observed when interacting with VN products than with HN and NN products. Vocal expressions were more frequent when interacting with VN and HN products than when interacting with NN products. In summary, it is possible to design surprising products influencing users' behaviour.

In study 2, facial expressions of surprise were investigated, using the same set of stimuli as in study 1. A total of 119 videos were scored by two independent observers who coded four components of a facial expression of surprise as described by Ekman and Friesen (2003), and used by Reisenzein (2000) and Vanhamme (2000) (see section 2.5.6). Analogous to study 1, surprise ratings for HN and VN products were significantly higher than for NN products. With regard to the manifestation of facial expressions of surprise, raters "hardly ever observed a complete facial expression of surprise" (Ludden et al., 2009, p. 85). Only 37.5 percent of participants exhibited one of the three components of surprise mentioned by Ekman and Friesen (2003). Also, correlations between subjective ratings of surprise and facial expressions of surprise were low, if significant.

Surprise creation through visual-auditory incongruities

In a second set of experiments, Ludden and Schifferstein (2007) investigated the effect of visual-auditory incongruity with a similar setup as in study 1. Instead of pairing conventional products with unexpected materials, conventional products like a dust-buster or a juicer were paired with unexpected tone. The researchers distinguished three kinds of sounds: the product's actual sound, a fitting sound and a non-fitting sound, the last two being a digital manipulation of the original sound. As expected, products with non-fitting sounds were rated significantly more surprising than products with a fitting sound. Apparently, expectations about a product are not restricted to the material and shape, but are also formed about the sound of a product. Interestingly, these results were

not always as consistent as the results obtained with products with visual-tactual incongruities. Some products were rated as being surprising, others weren't. Forming an expectation of how complex products will sound during usage seems to be more complicated than determining how a product feels.

Surprise creation through visual-olfactory incongruities

In a third set of experiments, Ludden and Schifferstein (2009) investigated the influence of visual-olfactory incongruity on product experience. Participants were presented with a variety of everyday products (e.g. a wooden bowl, boots, dice, and a kitchen paper holder). These products either had an added odour or not. Results show that products with an unexpected odour are evaluated moderately positively. In order to gain more insight into this effect, additional commentaries were analysed, giving important insight into the appropriateness of surprise. In the case of odour and products, participants only liked those surprising odours that were related to the product. The paper towel holder that smelt of coffee (a smell that can be found in a kitchen) was more likeable than the odour of the dice (lavender smell seemed too fresh for the participants). Additionally, for products that normally have a smell (lemons), no positive effect of surprising odour was found, whereas for products which normally do not smell (wooden bowl), the addition of an incongruent odour leads to a more positive evaluation of that product. In summary, the more incongruent an odour is, the smaller is the positive effect on product evaluation. This effect is strongest for products that normally do not have their own smell.

Longitudinal effects of surprise

Besides studying the immediate effect of surprise on product evaluation, long-term effects were also investigated. In a longitudinal study about items with visual-tactual incongruities, 57 participants were presented with products over the course of four to five weeks. Stimuli consisted of a set of three identical-looking items which differed in their tactual characteristics (see Figure 2.12 for an example). The visual appearance of these stimuli indicated how they might feel. The three versions of the product consisted of a match between the visual and tactual properties, a moderate mismatch and a strong mismatch. Three variables were investigated: levels of surprise, overall liking, and tactual liking. Over the long term, surprising products were evaluated positively, although no greater interest was expressed to experience them again in comparison to non-surprising products.

Figure 2.12: Example of a product set for soft boxes looking as if made out of felt, ranging from soft and flexible to inflexible and rough.

2.6.3 CONCLUSION

Geke Ludden's work is one of the first methodical attempts to systematically investigate surprise as a design strategy. The creation of surprise is achieved by eliciting incongruities between product properties (Ludden, Kudrowitz, Schifferstein & Hekkert, 2012; Ludden, Schifferstein & Hekkert, 2012). Although surprise has been mentioned by several researchers due to its possible positive effects on users (Filipowicz, 2006; Fokkinga & Desmet, 2012, 2013), little attention has been paid to this emotion and its potential for interactive systems and software. The advantages of gaining further insight into the deliberate implementation of surprise as a design strategy are manifold. Understanding the mechanisms of surprise and the impact it can have on users of digital interactive products, will help designers applying surprise is a helpful design tool. By creating unexpected moments during usage, a product's unique selling point can be established, the commitment to a product enhanced, and recognition, recall and word-of-mouth can be increased. On the other hand, designers could learn when it is not useful to implement surprise because the desired effect might not be achieved under certain circumstances. When reflecting on the potential of surprise as a design strategy, possible negative consequences of unexpected events must also be considered. Users could be disappointed by an unexpected event, feeling they have been misled and even fooled on experiencing a surprise (Ludden, Schifferstein & Hekkert, 2008). Long-term effects of surprise as a design strategy should be investigated, too, since the initial beneficial effects could vanish over time or even turn into an unpleasant experience.

2.7 SUMMARY

There are many affective phenomena like core affect, emotions, moods, or temperament. Emotions are inseparable from our behavior; they accompany us every day, for example through facial expressions (Ekman, 1993). They are intricate constructs that are difficult to define but seem to be best described as appraisal processes (Scherer, 2001). We appraise what we experience and the outcomes of these appraisals are our emotions. Depending on what we wish for and what we need, these emotions can be either positive, like joy or pleasure, or negative, like anger or disgust. One exception to this rule is surprise. While all other emotions take on positive or negative valence, surprise is high in arousal and neutral in valence (Russell, 1980). Nevertheless there are pleasant and unpleasant surprises. As other emotions, surprise is the result of an appraisal process (Meyer et al., 1991). The valence of surprise is the result of a match or mismatch with our wishes and concerns (Reisenzein, 2009). Accordingly, surprise is often followed by other emotions like amusement or interest (Ludden et al., 2006). One way to measure surprise reactions is through RT data. If a prolongation of RTs is observed during the encounter of something unexpected, this shows the typical interruption of ongoing processes and attention reallocation observed during the surprise process (Meyer et al., 1991, 1997). Because of the ability to elicit pleasant emotions, surprise is used as design tool (Machlovi, 2015). But this characteristic of surprise has been the subject of little systematic research. In classical product design, surprise seems to have beneficial effects on product ratings when implemented as a product characteristic (see Ludden, 2008 for an overview of the research). Several authors highlight the importance of emotions in the product design process (Jordan, 1997; Norman, 2002; Desmet & Hekkert, 2002). They focus on pleasant emotions and their effects on the product-user relationship, attachment, or satisfaction. Other authors argue for the inclusion of unpleasant emotions, as well (Fokkinga & Desmet, 2012, 2013; Silvia, 2009). Finally, the concept of UX was discussed as a measure of these emotions and the effects they can have on the user. Different models exist about UX, all highlighting different aspects, but agreeing on several points: UX is a holistic view of the user-product relationship; it takes into account usability related, non-instrumental aspects, as well as emotional aspects, and even considers factors like context and temporality of experience. In conclusion, emotions are an important part of interacting with products, surprise being especially interesting to designers because of its bipolarity. It should thus be investigated systematically in order to support designers applying it, to create enjoyable experiences for the users

3 RESEARCH APPROACH

To investigate the possible influence of surprise as a design feature in digital interactive products, three consecutive studies were carried out. A first approach to the research questions was realized in the context of digital computer games. Tetris was chosen as an experimental paradigm for the first two experiments. In a third experiment, a task-directed context was used instead, using a non-gaming mobile application as the experimental paradigm. Experimental data were analysed using IBM SPSS Statistics version 22.

3.1 FUNDAMENTAL RESTRICTIONS WHEN WORKING WITH SURPRISE

As has been argued, surprise is based on the mismatch between expectations and schemata (see section 2.5.1). After the encounter of an unexpected event, corresponding schemata are updated. This leads to the delay in RT (see section 2.5.2). Because the new information is incorporated in existing schemata, it is not possible to investigate surprise on the basis of a within-subjects variable. Participants will have certain expectations about experimental setups if they have already encountered 'strange' unexpected events during an earlier study. Possible surprise reactions would be at best less pronounced. The variable *group* was thus always implemented as a between-subjects variable. Furthermore, no participant took part in more than one of the three studies.

3.2 RESEARCH PARADIGMS

3.2.1 GAMING ENVIRONMENT

The goal of the first two experiments was to shed light on the effect of unexpected events, i.e. surprising behaviour, on the UX ratings of a digital, interactive product. Tetris was chosen as a suitable paradigm to investigate the research questions for several reasons:

First, to create a surprise, one has to create expectations which can later be frustrated, resulting in a possible surprise reaction. Because Tetris is a widely known game, there is no need to artificially create expectations, as would otherwise have been necessary. Second, Tetris is an open source game which allowed for the creation of suitable manipulations within the game by changing the code of an existing Java version. Third, the Tetris version could be played on a desktop computer, making it easier to log data and to send markers from the PC to a recorder, storing data from an electromyography[7] (EMG).

Tetris is a classical puzzle video game which became popular in the 1980s through the Nintendo Gameboy©. It was invented by the Russian mathematician Alexei Paschitnow who never took out a patent on it, making it an open source game until today. When playing Tetris, a player is confronted with one of seven so called tetronimoes: geometric shapes, each composed of four squares which abstractly form the letters I, J, L, O, S, T, and Z (see Figure 3.1). To simplify terminology, I will call these geometric shapes *blocks*. The game has relatively simple rules (see Lindstedt & Gray, 2013). The player consecutively manipulates descending blocks on a board, to fit them into the accumulation at the bottom of the game space. By moving them sideways and rotating them, the player can create horizontal rows without any gaps. Those complete rows disappear, and all blocks on top of them drop down a row. The accumulation gradually builds up because some rows remain incomplete. When one block of the accumulation reaches the top of the game space, the game ends. Classically, players enter a higher level when a certain number of gapless rows have been cleared. Some minor changes were made to the version of the game to make it suitable for the purpose of the study.

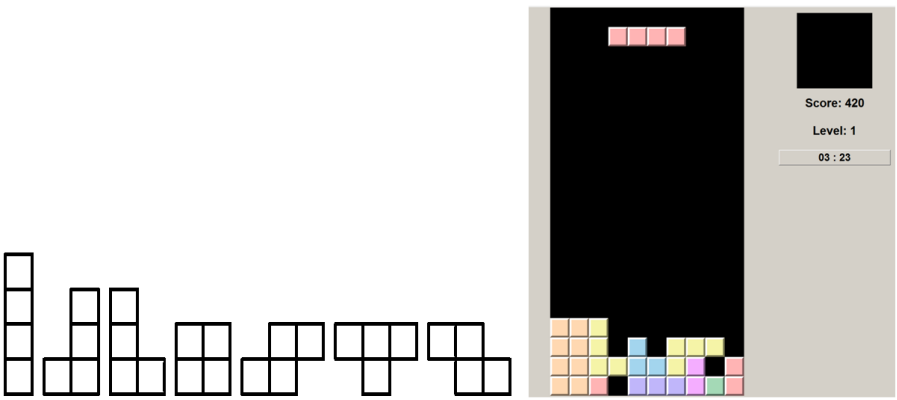


Figure 3.1. The seven tetronimoes (left); Screenshot of the Tetris game (right).

[7]EMG records the activity produced by muscles. Different emotions can be identified through EMG by distinct patterns of facial muscle activity (Schwartz, Fair, Salt, Mandel & Klerman, 1976).

3.2.2 GOAL-DIRECTED ENVIRONMENT

The third experiment, investigated the transfer of findings from a game to another context. The new experimental environment required the creation of a goal-oriented experimental setting. Because RTs had to be measured and averaged, it was necessary to let participants execute the same task many times, while allowing for the introduction of unexpected events. It was decided to investigate neutral surprise in the application, since it proved impossible to generate valenced unexpected events that did not interfere with the task-flow: the overlay of the interface with messages would have prolonged time on task, directly contradicting usability guidelines. Eliciting pleasant and unpleasant surprises without making use of messages to convey valence can result in an individual interpretation of the event, making valence uncontrollable.

As mentioned before, it is necessary to create expectations in order to frustrate them and thus possibly create surprise. As a result, an application was developed that was similar to existing and widely used applications for public transport (e.g. *öffi*, for Android devices). While the application used in Study III was not identical to these apps, the representation of available connections was similar (see Figure 3.2). Note that the app was used on a tablet computer in landscape mode during Study III, rather than in portrait mode, as for mobile phone use.

This form of presentation allowed for visual manipulation of distinct elements of the application, i.e. the rectangular shapes representing one means of transport.

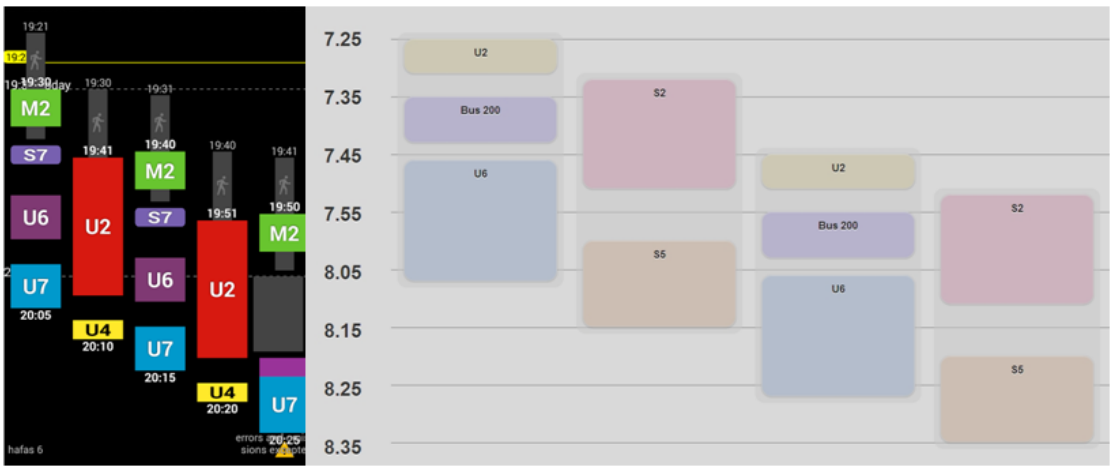


Figure 3.2: Available connections as represented in *öffi* application and during Study III.

3.3 QUESTIONNAIRES

To gain insight into subjective UX ratings, participants had to answer several questionnaires after the completion of every study. Not every study implemented the same set of questionnaires (see sections *dependent variables*).

Self-Assessment Manikin

The self-assessment manikin (SAM) is a 2-item 9-point non-verbal instrument for the evaluation of emotions measuring the dimensions valence, arousal, and dominance (Bradley & Lang, 1994). Participants indicate their feelings about a product on a 9-item scale ranging from 1=active to 9=calm (arousal), and 1=happy to 9=sad (valence). Because dominance is a difficult concept to grasp and a potential source of confusion, it is often omitted. Because no hypotheses concerning dominance were formulated, it was not included in the experiments. See Appendix A.5.1 for the version of SAM used in the experiments.

Product emotions

A product emotions questionnaire was developed and validated by Mano (1996). It covers eight subscales of pleasant and unpleasant product emotions: Arousal, elation, quietness, boredom, pleasantness, calmness, unhappiness, and distress. Each subscale is covered by three items. Items are presented as one word ('happy') with a 7-point scale (1=not at all; 7=very much). The subscale arousal is of particular interest since it includes the sub-item 'surprised'. Participants are asked to indicate their feelings during product use on these item scales (see Appendix A.5.2).

AttrakDiff 2[8]

The AttrakDiff is a questionnaire based on the model of UX by Hassenzahl (2005). It is an instrument to assess how users personally rate the usability and UX of an interactive product. It consists of 28 items with the subscales pragmatic quality (PQ), hedonic quality identification (HQI), hedonic quality stimulation (HQS), and attractiveness (Hassenzahl, Burmester & Koller, 2004). Every subscale is made up of 7 items that consist of semantic differentials. These are made up of two words representing the extreme values of one characteristic of the product, e.g. 'good' and 'bad'. On a 7-point scale, users can indicate how they rate the product. They are asked to indicate their evaluation without giving each

[8] The AttrakDiff 2 is the advancement of the AttrakDiff 1 questionnaire developed by Hassenzahl, (2001). It was developed to divide hedonic qualities of a product into *Stimulation* and *Identification*. It will be called AttrakDiff throughout the thesis.

item too much thought, ensuring spontaneity of ratings. The PQ subscale corresponds essentially to usability. Hedonic qualities refer to user needs that are the basis for the interaction with a product. The HQI subscale describes the extent to which a user can identify with a product. HQS on the other hand describes the personal striving after enhancement of skills and knowledge. The attractiveness subscale is consistent with an overall judgment. The AttrakDiff questionnaire can be found in Appendix A.5.3.

AttrakDiff mini

The AttrakDiff mini is a shorter version of the AttrakDiff questionnaire, consisting of ten items covering the same subscales as the AttrakDiff (Hassenzahl & Monk, 2010; see Appendix B.3.3).

meCUE

The meCUE was been developed to give researchers a modular tool when investigating the UX of a given product. There are three modules in the meCUE:

Product Perception incorporates the subcategories utility and usability, as well as visual aesthetics, status, and bonding.

User Emotions covers positive, as well as negative emotions.

Consequences of Use covers product loyalty, intention of use, and a global rating. All items are presented as statements to which the user can indicate agreement on a 7-point scale ranging from 1=no agreement at all, to 7= total agreement. See Minge, Riedel & Thüring (2013) for a more detailed description of the development and validation of the questionnaire. The questionnaire can be found in Appendix B.3.2.

Overall rating

An overall rating of the game was implemented through a 6-point non-verbal scale showing a thumb down at the left end and a thumb up at the right end, see Appendix A.5.3.

4 STUDY I – THE INFLUENCE OF UNEXPECTED EVENTS ON UX IN A GAME I

4.1 PROBLEM AND RESEARCH GOALS

In today's market it is difficult to set one product apart from others. Besides flawless functionality and quality, it is important to consider users' feelings and emotions when designing products (Jordan, 1997; Norman, 2002). Although traditionally the focus was on designing for pleasant emotions, newer research suggests a beneficial effect of negative emotions on the user product relationship, as well (Fokkinga & Desmet, 2012, 2013; Silvia, 2009). By combining traditional and new approaches of emotional product design, the attachment between user and product could be strengthened, product evaluations could be elevated, and product value as a whole. One emotion that can trigger pleasant as well as unpleasant feelings is surprise. Some researchers have investigated the influence of surprise on product ratings in classical product design. It has been shown that (pleasant) surprise can be beneficial for product ratings (see section 2.6.2). Research on the effects of unpleasant surprise on product ratings and UX has not been done, yet.

There are two main research goals for the first study. First, the producibility of artificial surprise[10] is tested. While it is not clear whether it is possible to generate unexpected events that lead to a surprise reaction in users during the interaction with a digital interactive product, surprise has been successfully evoked by products like vases or lamps. Because surprise is the result of a mismatch between expectations and perceived information, it has to be ensured that all users rely on similar expectations about the product that is tested. In addition, the conflicting unexpected information needs to be strong enough to elicit a surprise reaction, in an artificial laboratory setting. Because the

Chapter 4 is partly based on: Gross, A., & Thüring, M. (2013). Encountering the Unexpected: Influencing User Experience through Surprise. In *Proc. of UMAP Annual Conference*, Rome, Italy: Springer, 3-9.

[10] *Artificial surprise* is defined and used in this thesis as a surprising event that is deliberately evoked by the introduction of information that intentionally contradicts a person's expectations. In practice, it describes the act of systematically introducing unexpected events to participants during the course of a study in order to evoke a surprise reaction.

goal of this study is to investigate the effects of both pleasant and unpleasant surprise, an experimental manipulation has to be designed that validly and strongly elicits pleasant und unpleasant feelings. It is of general interest how a repeated presentation of unexpected information is perceived by users. Based on earlier works on surprising stimuli, one way to validate a successful surprise evocation is through the measurement of RTs (Meyer, et al. 1991, 1997). Another objective indicator of surprise is facial expression (Ekman, 1993, Ekman & Friesen, 2003). Measuring muscle activity indicative of the typical surprised facial expression can be a verification of a successful surprise evocation. The first research goal of this study is thus:

- to create surprise through the implementation of unexpected events during interaction with a digital interactive game.

Successful surprise generation should be measurable indirectly through RT prolongations and facial expressions, detected through EMG data.

The effects of pleasant, unpleasant, and neutral surprises on the UX ratings of users of a digital computer game are also to be investigated. Neutral surprises are included to ensure that effects on UX can be attributed to the valence of surprise, not to surprise itself. The second research goal of this study is therefore:

- to investigate the effects of pleasant, unpleasant, and neutral surprise on the UX of digital interactive game.

Surprise will be elicited by the introduction of unexpected events in the form of point subtraction or addition from the score a participant has achieved during the game. A neutral event is assumed to elicit neutral surprise. UX ratings are compared with the ratings of a control group and the neutral group. Based on earlier findings by Ludden (see Ludden, 2008) it is assumed that positive unexpected events create pleasant surprise and lead to an overall better rating of the game compared to the control and the neutral group. UX ratings of the group that experiences unpleasant surprise are expected to be worse than those of the neutral and the control groups. Although it has been mentioned in the literature that negative emotions can elicit pleasant product experience, it is not assumed that the general conditions that are required to do so are met in this experimental setup (see Fokkinga & Desmet, 2012, 2013). Neutral surprise is assumed to have no effect on UX ratings.

4.2 METHOD

4.2.1 PARTICIPANTS

Participants were recruited from the Technische Universität Berlin via bulletins distributed in several university buildings. Furthermore, participants could sign up for the study online via the prometei online platform for study volunteers. At least a fair knowledge of the game Tetris was stated as a prerequisite for participation in the study. A sufficient mastery of the German language was also required. 77 participants initially took part in the study. Eleven participants had to be eliminated from further analysis due to technical problems or insufficient knowledge and mastery of the game. A total of 66 participants were analysed, of which 43 were male and 23 were female. The average age of the participants was 24.9 years (aged 18-40, SD = 4.2). Roughly one third of the participants (36.4%) had a university degree and about two thirds (60.6%) had a secondary school qualification ('Abitur') as their highest educational qualification at the time of the study. Prior knowledge of Tetris was assessed through a 6-point item in the demographic questionnaire with the anchors 'no experience at all' (1) and 'a lot of experience' (6). Participants were explicitly advised to take earlier experience from their childhood and youth into account when assessing their experience with Tetris. All of them said they were familiar with the game and estimated their mastery as being at least fair (3 or higher on the experience scale). There were no self-reported differences for the factor experience with Tetris between the groups ($F(65,3)=2.578$, $p \geq .05$). Participants received ten euros as gratification for their participation (see section *independent variables* for a more detailed description of the monetary reward structure).

4.2.2 MATERIALS

Three computers were connected over a local area network to generate the Java Tetris game and to allow for simultaneous EMG, RT, and video recording. The research computer consisted of a fixed monitor (EIZO Flex Scan L568 17 inch screen) and keyboard (Lenovo, model SK-8825) and a Dell Precision 390 computer with Windows SP Professional, 1GB RAM and an Intel Core 2 DUO Processor. Screens were recorded with a Pentium Core 2 Quad, 2.4 Gigahertz processor, 2GB RAM computer with Windows XP. The screen recorder software was Screen recorder V3.19 by biobserve GmbH. The video camera used was a Sony digital camera, type handycam DCR-TRV20E. EMG data was recorded with the Varioport system by Becker Meditec. The system ran on a computer with a Pentium 4, 2.6 gigahertz processor with 512 MB RAM and Windows XP. Electrodes

used for recording EMG data were disposable electrocardiography electrodes with a 24mm diameter by Kendall™ (see Figure 4.1 for the whole experimental setup).

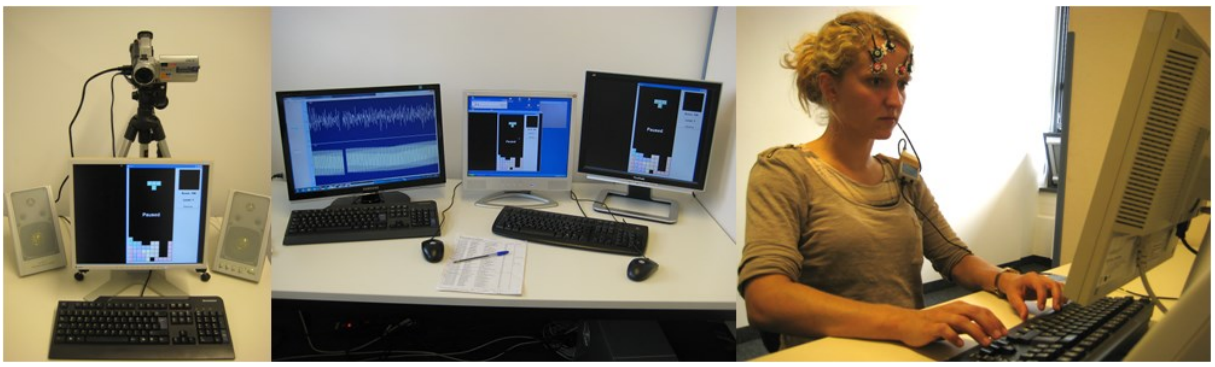


Figure 4.1. Research computer (left); experimenter work space with EMG data screen, screen recorder screen, and copy of research screen (middle, from left to right); participant with EMG electrodes (right).

4.2.3 PROCEDURE

The study was conducted in a laboratory of the Chair of Cognitive Psychology and Cognitive Ergonomics at Technische Universität Berlin. Each session lasted for about 50-60 minutes on average. Upon arrival, participants were seated in front of the experimental computer and were asked to fill out a demographic questionnaire. Electrodes were placed on the face of participants after cleaning the skin. The signal strength was tested on the experimenter computer and the fit of the electrodes was adjusted, if necessary. Participants were then made familiar with Tetris by playing a training game. The training version stopped after 300 blocks and participants were not told to reach a specific number of points. If no further questions arose, the experimental Tetris trial was started. Participants were given the task to complete a Tetris game within a time frame of five minutes, reaching a certain number of points, depending on the experimental group they were assigned to (see section 4.2.2, *independent variables*). RTs were defined as the time between the first appearance of a block with the accompanying message (both appearing simultaneously), and the first key press. To introduce surprise, a message was flashed up on the computer screen at three different points of time during the game. UX questionnaires were filled out immediately after completion of the game.

4.2.4 DESIGN

Independent variables

Two independent variables were manipulated in the study. The first was a between-subjects variable called *group*. All participants were told that the goal of the game was to reach a certain amount of points in five minutes. Participants were told that they would be reimbursed with seven euros for participation, and would get an additional three euros if they reached their points in the time, without producing a game over. Actually, all participants were reimbursed with 10 Euros, but only participants who reached the objective were included in data analysis.

Negative group (17 participants)

In the negative group, participants were told to reach 500 points within the five minute time frame. Simultaneously with the appearance of the 38th stone, a message was displayed on the screen, reading: "!!!Abzug -50 Punkte!!!" (!!!Deduction, -50 points!!!), as shown in Figure 4.2. The same message was displayed for the 47th and the 51st stone. Because every time the message was displayed, 50 points were indeed subtracted from the player's account, the score was 350 after the third reduction. The player then had to reach a score of 500 points, handling 66 blocks in total.

Positive group (17 participants)

In the positive group, participants were told to reach 800 points within the five minute time frame. The surprising message in this group read: "!!!Bonus +50 Punkte!!!" (!!!Bonus, +50 points!!!), see Figure 4.2. The same message was displayed for the 47th and the 51st stone. Because every time the message was displayed, 50 points were added to the player's account, the score was 650 after the third bonus. The player then had to reach a score of 800 points, handling 66 blocks in total.

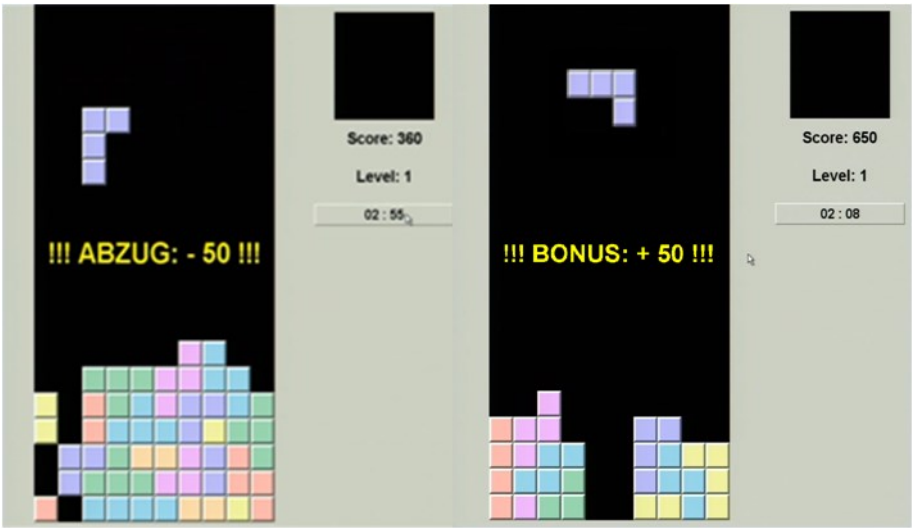


Figure 4.2: Unexpected events for negative group (right), and positive group (left).

Control group (16 participants)

In the control group, participants were told to reach 650 points within the five minute time frame, handling 66 blocks in total. No unexpected events occurred during the game.

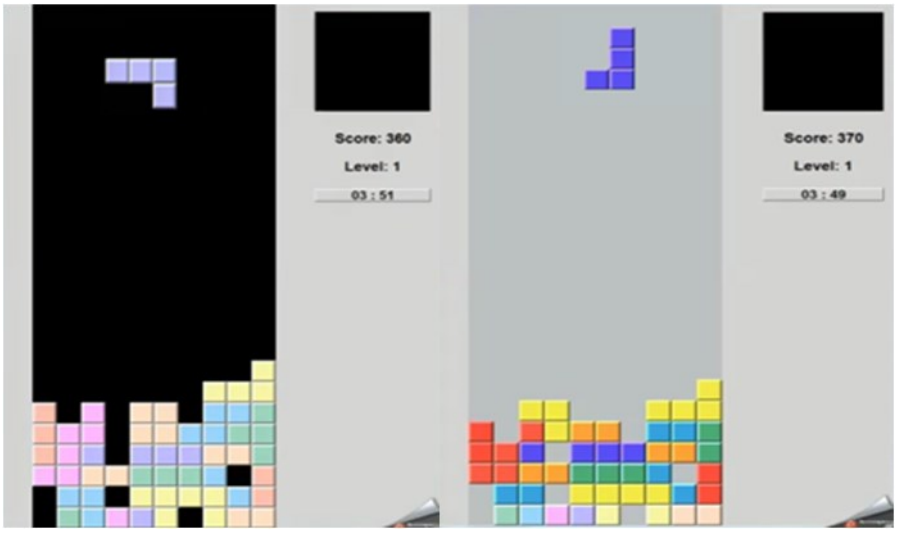


Figure 4.3: Colour change as manifestation of the unexpected event in the neutral group.

Neutral group (16 participants)

The neutral group played the same Tetris game as the other groups, this time without messages about the current score. Instead, all blocks, as well as the background changed colour when block 38, 47, and 51 appeared. The original colours of the blocks were pastel variations of yellow, orange, red, pink, purple, blue, and green. Background colour was

black. For colour change, saturation of all colours was augmented, making them brighter. Background colour was changed to a light grey. All colours were reversed to their original when the following block appeared. (See Figure 4.4 for all colours, their corresponding hexadecimal colour codes, and their corresponding blocks used in the Tetris game.)

Event

The within-subjects factor *event* served as second independent variable. It consisted of three treatments, i.e. the first, second and third time an unexpected event occurred (event1, event2, and event3). Events were coupled with blocks 38, 47, and 51.

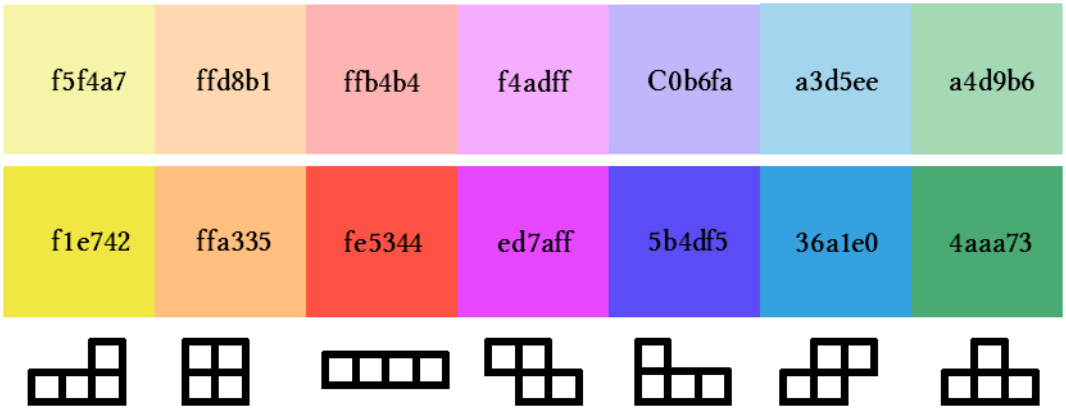


Figure 4.4. Original block colours (top); saturated block colours (middle); corresponding blocks (bottom.)

Dependent Variables

Two objective measures for surprise were employed, as well as four subjective measures for assessing changes in UX. RTs were recorded for processing a Tetris block that was accompanied by a surprising event. Data from facial EMGs was included as a measure of surprise reactions. Activity of the lateral frontalis muscle was measured (see Figure 4.5) which has been associated with the facial manifestation of surprise reactions (Schwartz et al., 1976; van Boxtel, 2010).

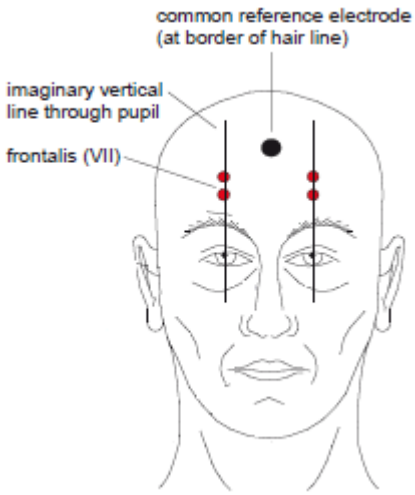


Figure 4.5. Electrode locations for measuring facial EMG activity (from van Boxtel, 2010).

UX was assessed using four questionnaires: the SAM questionnaire, the AttrakDiff, the product emotions questionnaire by Mano (1996), and a self-developed single-item questionnaire for judging the overall UX of the game (see section 3.3).[11]

Study Design

The study was based on a 3x4 mixed design. "Event" was treated as the within–subjects factor with repeated measurements and three levels. "Group" was treated as the between-subjects factor with four levels to which participants were randomly assigned ensuring equal group size and balancing of gender.

4.2.5 HYPOTHESES

Surprise

Following Meyer et al., (1991, 1997), RT data serve as indicator for a successful surprise manipulation:

H1: An interaction effect is predicted between the factors *group* and *event*.

For the positive group, the negative group, and the neutral group, mean RTs will decrease from event 1 over event 2 to event 3 back to baseline, similar to RTs in group control. The control group will not show this pattern.

This hypothesis is based on the assumption that the extent of surprise diminishes when an unexpected event is encountered more than once (Meyer et al., 1991, 1997): prolonged reaction indicate an interruption of ongoing processes. This prolongation

[11] Hypothesis concerning EMG data are not shown due to the data not being analysable.

should vanish over a repeated presentation of unexpected events due to the assimilation of the unexpected new information into the mental model of the situation.

H2: A main effect on RTs is predicted for the factor *group*.

H.2.1 Mean RTs for the positive group will be longer than for the control group for the first unexpected event (event1).

H.2.2 Mean RTs of the negative group will be longer than those of the control group for the first unexpected event (event1).

H.2.3 Mean RTs of the neutral group will be longer than those of the control group for the first unexpected event (event1).

These hypotheses are derived from findings by Meyer et al. (1991, 1997). RT prolongation after the occurrence of an unexpected event can be explained in terms of the cognitive psycho-evolutionary model of surprise. Due to the formation of a schema about the Tetris game, the unexpected event will be appraised as schema-discrepant and will require allocation of attention resources to be analysed, resulting in prolonged RT for event1.

User Experience

(Hypotheses are formulated for subscales of questionnaires for which an effect is anticipated, only.)

SAM ratings (measuring emotional reactions to a product):

H3: A main effect for the factor *group* is predicted.

H.3.1 SAM valence ratings made by the positive group will be more positive than those of the control group and the neutral group.

H.3.2 SAM valence ratings made by the negative group will be more negative than those of the control group and the neutral group.

These hypotheses are based on findings by Ludden (2008) that show a beneficial effect of (pleasant) surprise on product evaluations. Although it has been mentioned that negative surprise can have a beneficial effects on users (Fokkinga & Desmet, 2012, 2013), the framework requirements for this are not met (see section 2.4.4)

H.3.3 SAM arousal ratings made by the positive group, the negative group, and the neutral group will be higher than those of the control group.

Surprise is an emotion high in arousal (Russell, 1980), which should be measurable by this subscale.

Mano ratings:

H.4: A main effect for the factor *group* is predicted.

H.4.1 Ratings for the subscale *arousal* given by the positive group, the negative group, and the neutral group will be higher than those of the control group.

The reasons for hypotheses formulation are identical to hypotheses about SAM arousal ratings.

H.4.2 Ratings for the subscales *elation* and *pleasantness* given by the positive group will be more positive than those of the control group and the neutral group.

H.4.3 Ratings for the subscales *elation* and *pleasantness* given by the negative group will be more negative than those of the control group and the neutral group.

H.4.4 Ratings for the subscale *unhappiness* given by the positive group will be lower than those of the control group and the neutral group

H.4.5 Ratings for the subscale *unhappiness* given by the negative group will be higher than those of the control group and the neutral group

These hypotheses are based on findings by Ludden (2008) that show a beneficial effect of (pleasant) surprise on product evaluations.

AttrakDiff ratings:

H.5: A main effect for the factor *group* is predicted.

H.5.1 AttrakDiff ratings on the HQI subscale will be higher in the positive group than for the control and neutral group.

H.5.2 AttrakDiff ratings on the HQI subscale will be lower for the negative group than for the control group and the neutral group.

H.5.3 AttrakDiff ratings on the attractiveness subscale will be higher for the positive group than for the control group and the neutral group.

H.5.4 AttrakDiff ratings on the attractiveness subscale will be lower for the negative group than for the control group and the neutral group.

These hypotheses are based on findings by Ludden (2008) that show a beneficial effect of (pleasant) surprise on product evaluations. Participants are expected to identify more with a pleasant product and rate it better overall.

Single item overall rating:

H.6: A main effect for group is expected.

H.6.1 Ratings will be higher for the positive group than for the control and the neutral group.

H.6.2 Ratings will be lower for the negative group than for the control and the neutral group.

4.3 RESULTS

4.3.1 ANALYSIS

RT data was analysed using repeated measures ANOVA procedure in SPSS. Polynomial contrasts were calculated to confirm a decrease in RTs from event1 to event2. Simple and repeated contrasts were calculated to assess significant differences in RTs between the four groups. EMG data could not be analysed due to inaccuracy of markers which were set manually and did not prove to be usable. Questionnaire data was analysed employing multivariate analysis of variances (MANOVA), with group as the between subjects factor and all questionnaire subscales as the dependent variables. All questionnaire data was z-transformed before analysis to compensate differences in item scaling. Tables and graphs report untransformed questionnaire data for better understanding.

4.3.2 SURPRISE

Raw and corrected RT

Visual inspection of raw RT data for all participants divided by group already indicated elevated RTs for at least the first two unexpected events in the neutral group, negative group, and the positive group (see Figure 4.6). No such RT patterns are found for participants in the control group.

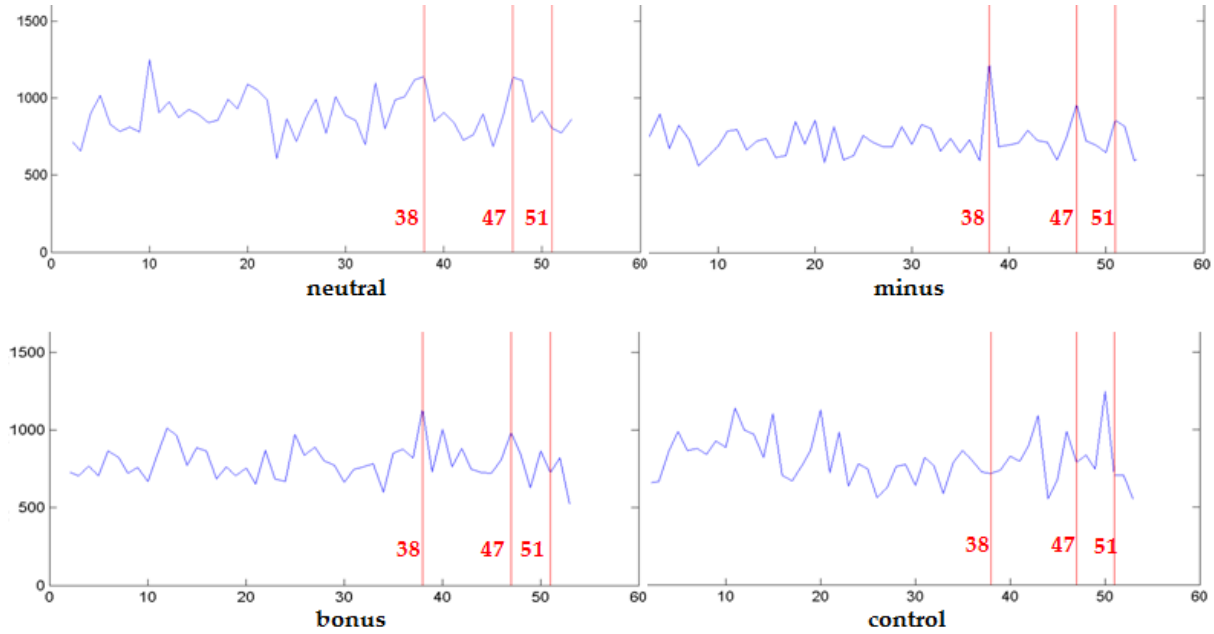


Figure 4.6: Raw RTs per group averaged over all participants in each group.

To take into account inter-individual differences in response speed, all RTs were corrected for mean prior RTs. Because every participant was in a different gaming situation concerning number of stacked lines, holes in stacked rows, and thus stress level, an individual baseline was conducted for every participant. This baseline was the mean of the three RTs prior to the block coupled with an event. Thus, individual speed differences, as well as differences in the gaming situation were compensated. Figure 4.7 shows uncorrected (left) and corrected (right) mean RTs for all groups over the three events including standard errors (SE).

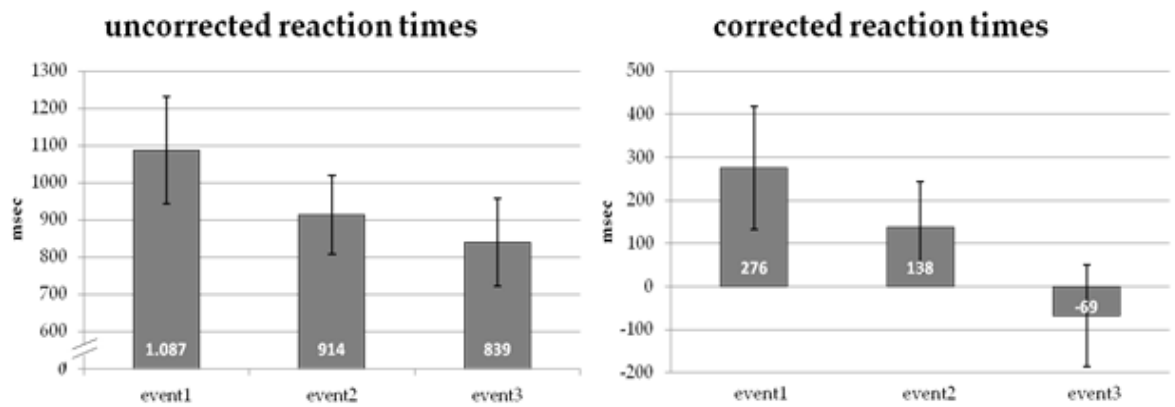


Figure 4.7: Uncorrected RTs (left) and corrected RTs (right) with SE for factor event.

A 3x4 repeated measures ANOVA was carried out with group as the between subjects factor. Mauchley's Test of Sphericity indicated no assumption violation ($\chi^2(2) = 1.936$, $p = .380$). The test revealed no significant interaction effect between *group* and *event* ($F(6,124)=1.060$, $p=.390$, partial $\eta 2 = .049$). However, the main effect for the factor *group* proved to be significant ($F(3,62)=9.009$, $p=.000$, partial $\eta 2 = .304$). As can be seen in Figure 4.8, the control group was faster than all other groups. The negative group was overall slowest in responding, followed by the positive group, and the neutral group. Simple contrasts revealed that participants in the negative group were significantly slower than participants in both the positive group (p=.034), and in the control group (p=.000), and in the neutral group (p=.001). The difference between the positive group and the control group was also statistically significant (p=.007). There was also a main effect for factor *event* ($F(2,124)=7.813$, p=.001, partial $\eta 2 = .112$). Contrasts revealed a significant linear trend, ($F(1,62) = 16.749$, p=.000, partial $\eta 2 = .213$), indicating that as an event appears repeatedly, RTs decrease and stabilize at a baseline level for the third event.

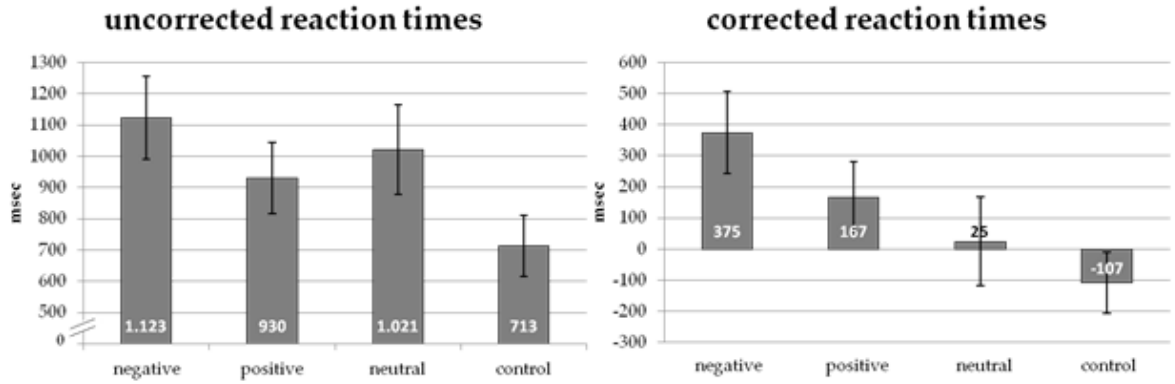


Figure 4.8: Uncorrected RTs (left) and corrected RTs (right) with SE for factor group.

Because the interaction effect did not prove to be significant for factors group and event, a one-factorial ANOVA for event1 with group as the between factor was conducted. The goal of this analysis was to evaluate possibly different reactions to the first surprising event between user groups. Surprise should be strongest for event1, due to its novelty and unexpectedness. The ANOVA revealed a significant effect ($F(3,65)=6.984$, $p=.000$). Post-hoc Bonferroni comparisons revealed that the negative group showed significantly longer RTs to event1 than the control and the neutral group. Other comparisons did not prove to be statistically significant.

4.3.3 USER EXPERIENCE QUESTIONNAIRES

To investigate surprise effects on UX ratings, a one-factorial MANOVA was carried out with *group* as between-subjects factor (all values z-transformed). Event was no longer a factor in the analysis because UX questionnaires were completed after finishing of the game. They were thus not coupled with one of the three events. There was a statistically significant difference between the groups based on questionnaire ratings ($F(45, 143.376) = 1,750$, $p=.007$; Wilk's $\Lambda = 0.272$, partial $\eta 2 =.352$). Significant effects were found for the SAM subscale Valence ($F(3,62)=3.008$, $p=.037$, partial $\eta 2 =.127$), the AttrakDiff subscale PQ ($F(3,62)=3.064$, $p=.035$, partial $\eta 2 =.129$), the AttrakDiff subscale HQI ($F(3,62)=3.260$, $p=.0272$, partial $\eta 2 =.136$), and the Mano subscale boredom ($F(3,62)=3.902$, $p=.013$, partial $\eta 2 =.159$).

Table 4.1: Mean ratings from UX questionnaires with standard deviations (SD).

Questionnaire subscales	*negative (SD)*	*positive (SD)*	*neutral (SD)*	*control (SD)*
SAM Valence	3.47 (1.73)	2.06 (1.34)	2.69 (1.53)	2.31 (1.14)
SAM Arousal	4.64 (2.02)	4.76 (2.54)	4.81 (1.68)	4.75 (1.69)
AttrakDiff PQ	5.03 (0.43)	5.19 (0.70)	5.55 (0.45)	5.21 (0.36)
AttrakDiff HQI	3.51 (0.80)	4.21 (0.75)	3.68 (0.78)	4.04 (0.72)
AttrakDiff ATT	4.96 (0.79)	5.47 (0.78)	5.24 (0.93)	5.33 (0.62)
Mano Boredom	2.49 (1.30)	1.94 (1.04)	1.38 (0.48)	2.21 (0.85)
Mano Arousal	3.19 (0.89)	3.31 (0.89)	3.15 (1.03)	3.65 (0.86)
Mano Elation	4.35 (0.98)	4.73 (0.72)	4.98 (0.89)	4.71 (1.17)
Mano Pleasantness	4.78 (0.80)	5.02 (0.77)	5.06 (0.76)	5.44 (0.75)
Mano Boredom	2.49 (1.30)	1.94 (1.04)	1.38 (0.48)	2.21 (0.85)

Contrasts showed that for SAM Valence and AttrakDiff HQI subscales, participants in the negative group gave significantly worse ratings than participants in the positive group and the control group on both of these scales. Furthermore, contrasts revealed that for AttrakDiff PQ subscale, participants in the neutral group rated the game significantly better than participants in the positive group and the negative group. For the subscale boredom of the Mano questionnaire, participants indicated significantly lower levels of boredom in the neutral group when compared to the negative and the positive group.

Mean ratings for the dimensions of the UX-questionnaire subscales with significant differences between groups are shown in Table 4.1.

A main effect for *group* was expected for the non-verbal overall rating of the game (H.6). This could not be confirmed ($F(3,62)=1.477$, $p=.229$, partial $\eta 2 =.067$)).

4.4 DISCUSSION

The focus of the first experiment was the investigation of the influence of pleasant and unpleasant unexpected events on the UX ratings of a digital Tetris game, when compared to ratings of a game with or without a neutral unexpected event. The occurrence of an unexpected event was expected to elicit a surprise reaction in participants, with consequences on RTs and UX ratings of the game.

4.4.1 SURPRISE

RT prolongation was expected as a result of the cognitive appraisal processes following the perception of an unexpected event, pointing towards a successful surprise manipulation.

It was assumed that participants would show prolonged RTs when encountering an unexpected event, regardless of its valence. A prolongation of RTs for a first unexpected event could be interpreted as the direct consequence of a successful elicitation of surprise (H1). This assumption could not fully be confirmed. For the first unexpected event, prolongations in RTs for all three experimental groups were observed, but no prolongation whatsoever for the control group (see Figure 4.7). Although the RT prolongation of the positive group and the neutral group is clearly observable in the data, it is not pronounced enough to become statistically significant (H.2.1 & H.2.3). However, significant differences could be confirmed between the negative group and the control group (H.2.2). Curiously, the difference in RTs for event1 between the negative group and the neutral group also proved to be significant. The manipulation of colour as the implementation of unexpectedness does not seem to have as strong an effect as the manipulation of the game score. Furthermore, it seems that the introduction of unexpected events in the form of a manipulation of the score had a stronger effect for the negative group, than for the positive group. These findings do not contradict earlier findings on the topic of RT patterns and surprise. When Meyer et al. (1991, 1997) found significant RT prolongations following unexpected events, no distinction was made between different kinds of surprise. Although the same neutral unexpected event (a change of colour, see section 2.5.2) was used to

create surprise, the task that had to be mastered was cognitively unchallenging, when compared to a complex puzzle game like Tetris. Therefore, one issue has to be discussed in light of the inconclusive results regarding RTs for event1. Why do only RTs for the unpleasant event differ significantly from the control group?

Although similar surprise manipulation were used as in earlier studies (Meyer et al., 1991, 1997), it was not possible to significantly create surprise through the colour change in the neutral group. One explanation is the cognitive demand of the Tetris game. Whereas the task in earlier studies was a forced choice RT task that consisted of pushing one of two buttons depending on whether a circle appeared above or below a rectangular box, playing Tetris is a cognitively highly challenging task (Lindstedt & Grey, 2013): participants could have been just too concentrated on the game to pay attention to colour change, being of no importance to the outcome of the game. Furthermore, the colour change did not interfere with the goal of 'winning the game', thus it was not relevant for the task. Accordingly, the action-relevance check as described by Meyer et al. (1997) in the cognitive psycho-evolutionary model of surprise did not turn out to be as excessive as for an unexpected manipulation of the game score, which has a direct influence on one's probability of winning the game.

Still the question remains why RTs for the positive group did not differ significantly from the control group. The answer can be found when investigating the nature of the human mind. According to the prospect theory, people estimate losses as more severe than gains (Kahneman & Tversky, 1979). Although Kahneman and Tversky base their theory on studies about the subjectively assessed chances to gain or lose money, the task at hand is comparable: the game score has been earned and is directly related to the prospect of gaining an extra monetary reward. Accordingly, in order to achieve a surprise reaction in the positive group that is as strong as the one in the negative group, the amount of points added to the score should have been higher than the amount deducted. This could have resulted in an equally strong reaction, observable in RT data of the first unexpected event. Or as Tversky and Kahneman (1991) put it: "losses loom larger than corresponding gains" (p. 1047), and profits should thus be adjusted accordingly.

Another explanation could be found in the context of use. While playing a computer game, the unexpected addition of points could have been understood by the participants as some kind of bonus that was not introduced randomly, but rather to reward them for an aspect of their performance.

H2 concerned the change of RTs over time and predicted an interaction effect of the two dependent variables *group* and *event* due to the attenuation of the surprising character of

the events with repeated occurrence. Although visual inspection of the data suggests such an effect, it did not prove to be significant. However, there was a significant linear trend for the variable event, indicating the typical pattern of RT data for repeated presentation of unexpected events. Meyer et al. (1991, 1997) showed that already on the second presentation of an unexpected event RTs are not as elevated as they were for the first presentation. The data shows that for the third occurrence of an event, RTs are back to baseline level. Furthermore, there was a significant main effect for event, indicating that although group control does not experience any changes in RTs, RTs for all other groups decreased from the first to the third occurrence of the unexpected event.

In summary, it seems that the manipulation of surprise was only partially successful. Apparently, the unexpected bonus was not as surprising as the unexpected reduction of points. This interpretation is supported by both RTs and UX ratings. The unexpected loss of points did have the predicted effect. Trials with unpleasant surprises took longer to process and the ratings of the respective group indicate a less positive UX.

4.4.2 User Experience

Research on surprise as a design strategy has shown beneficial effects on the appraisal of a variety of non-digital artefacts (see Ludden, 2008, for an overview). The second goal of the first study was thus to investigate the effects of surprise on product ratings and UX.

To measure the impact of surprise on UX, a number of rating scales were used. Because RT data results were not fully conclusive regarding the success of the surprise manipulation for all groups, reliable interpretation of UX questionnaire data is not fully possible. Nevertheless, the significant linear trend in the RT data (implying a decrease of RTs from event1 to event3) resembles the typical decrease of RTs with repeated presentation of unexpected (surprising) events.

SAM

It was assumed that pleasant events would elevate UX ratings, whereas unpleasant events would decrease those ratings in comparison to neutral events or no events. Accordingly, a main effect for group was expected concerning SAM ratings (H3). These assumptions could not fully be confirmed. Participants in the positive group did not rate the game significantly higher than participants in the neutral group and the control group (H.3.1). Only participants in the negative group rated the game more negatively than participants in the control group and the positive group (H.3.2). No differences to the neutral group could be confirmed. As discussed in the context of RT results in the preceding section, it

seems as if there has been a difference between the strength of the elicited surprise reactions: pleasant surprise did not have an effect as strong as unpleasant surprise. Accordingly, effects on UX ratings were not as pronounced as in the negative group. Arousal ratings did not differ between the groups at all. Thus H3.3 could not be confirmed. This result seems irritating at first glance, but can be explained in terms of the context of use. All participants had to master the challenge of winning the Tetris game. Although some participants encountered unexpected events in the course of the game, all had to play against the clock. It is possible that this experience on its own led to comparably high arousal levels. An unexpected surprising event did not elevate these arousal levels any further, see Appendix A.6.2.

Product emotions (Mano questionnaire)

A significant difference between groups was expected on the Mano subscale *arousal* (H.4.1). This subscale consists of three items, one of which directly asks participants about their levels if surprise. It was expected that all experimental groups would indicate higher ratings on this subscale than the control group. This could not be confirmed. The same characteristics inherent to the Tetris game as described in the preceding paragraph could have produced this result. The other hypotheses concerning the valence of emotions toward the game could not be confirmed, either (H.4.2 – H.4.5). No differences between the groups concerning the emotional appraisal of the Tetris game could be measured. However, one unexpected result was obtained: the neutral group reported finding the Tetris game significantly less boring than the negative group and the control group. When discussing this finding, it is important to note that all four groups gave considerably low boredom ratings (see Table 4.1). One characteristic that sets the neutral Tetris version apart from the other versions is the implementation of a colour change. Since no other aspects of the game differ between the neutral and the negative group and the control group, the repeated colour change in the neutral group could have had an influence on the subjective level of boredom.

AttrakDiff

A main effect for group was confirmed for the AttrakDiff questionnaire (H.5), resulting from significant differences in the subscales HQI and PQ. It was assumed that participants in the positive group would identify more with the game than participants in the neutral and the control groups (H.5.1). This could not be confirmed. On the other hand, the assumption that participants in the negative group would identify less with the game than participants in the neutral group and the control group was confirmed (H.5.2). No effects

could be found on the overall rating of the product, i.e. the attractiveness subscale (H.5.3 & H.5.4).

One unexpected finding was that participants in the neutral group rated the PQ of the game significantly better than participants in the negative group and the positive group. PQ subsumes factors describing the usability of product. The manipulation of the game score could have been interpreted as a direct conflict with two of the main usability principles, as described in the ISO 9241, section 110: A system should *conform to user expectations*, and *possess controllability*. By giving and taking points without any obvious reason, the game directly contradicted these two principles. As mentioned before, a colour change did not interfere with the task at hand. No immediate effects on PQ are thus observed for the neutral group.

Overall rating

Similar to the findings from the AttrakDiff subscale attractiveness, no differences were found between the positive group and the negative group, or the neutral and control group (H.6.1 & H.6.2). When arguing in line with the CUE model by Mahlke and Thüring (2007), a manipulation of the users' emotions should have led to a change in the overall rating of the game, given high levels of arousal (see section 2.2.2). Although it was expected to find such a difference due to the emotional manipulation of the game, the absence of this effect can be explained in terms of the UX model by Hassenzahl (2005; see section 2.2.1). The intended product character did not match the apparent product character. In the context of games, the emotional manipulation of the Tetris game was not strong enough to elicit different overall product ratings among the groups. In the end, all participants played a subjectively identical Tetris game.

4.4.3 SUMMARY AND CONTRIBUTION

Not all of the hypotheses postulated about RTs and UX ratings of the Tetris game could be confirmed. One difficulty in interpreting the data lay in the partly inconclusive RT results. However, it was revealed that emotional valence, as well as the HQI was affected by the factor group. This effect resulted from the impact of negative surprises in the negative group. Mean ratings differ between this group and the other two groups in the expected direction. But similar to the results of the RTs, no difference could be substantiated between the positive group and the control and neutral groups.

Given the fact that all participants in all groups experienced a positive and financially successful completion of the game, it is even more remarkable that unpleasant surprise

still influenced UX. With respect to UX, the results thus stress two points. First, unexpected events in the course of human computer interaction which entail undesirable consequences should be prevented. They lead to unpleasant surprises which in turn impede users' information processing and have a negative impact on UX. Second, whether unexpected events with desirable consequences lead to positive surprises, mainly depends on the interaction context and on the kind of system under consideration. As the first study shows, an unexpected bonus in a game may not be as surprising as one might suppose. For other systems and in different contexts, such as software in a working environment, an unexpected and beneficial system response may prove as more surprising. Therefore, more research is required to investigate the causes and effects of positive surprise.

From a marketing perspective, the study raises the question whether all positive features of a system should be immediately apparent or whether some of them should be covered. Is it more beneficial to tell customers all positive aspects to prompt them to purchase the system? Or is it better to let them discover some surprising extras later which might pay off in the long run by increasing brand loyalty? Obviously, the presented results are not far-ranging enough to provide a sound answer, but future investigations may shed more light on this issue.

4.4.4 CRITICAL EVALUATION

There are some issues with the implementation of the first study. First, in order to investigate the effects of peasant and unpleasant surprise, it has to be guaranteed that participants indeed feel surprised. Due to faulty EMG data and partly inconclusive RT data, this proved to be difficult. A stronger surprise manipulation should be designed in order to validly elicit surprise.

Second, the design of the unexpected events is problematic in terms of cognitive effort. Whereas the positive and the negative group both received messages with text, the neutral group did not have to process written information. The initial difference in RTs for event1 could be attributed to this characteristic of the event design. Unexpected events should be designed similarly in order to avoid such differences in cognitive effort and processing.

Third, one of the critical issues of study I was the selection of questionnaires. Several overlaps in the measured dimensions of UX were the result. A more conclusive selection of measurement tools should yield more precise hypotheses and differentiated findings.

5 STUDY II - THE INFLUENCE OF UNEXPECTED EVENTS ON UX IN A GAME II

5.1 PROBLEM AND RESEARCH GOALS

The second study was carried out to further investigate the experimental findings and methodical failures from study I. Although descriptive data supported the experimental hypotheses in study I, the surprise manipulations were too weak. A stronger experimental manipulation was considered to be more effective in terms of evoking a surprise reaction in participants, especially for pleasantly surprising events. Furthermore, EMG data needed to be generated differently, in order to get analysable data sets. The general assumptions were still that pleasant and unpleasant surprises have differing effects on the UX ratings of a Tetris game. Three manipulations were chosen to create pleasant, unpleasant, and neutral surprise. To verify a successful surprise manipulation, RTs and facial EMG data were analysed.

The research goals for study II were identical to those for study I (see section 4.1).

The aim in Study II is to elicit surprise by the introduction of unexpected events in the form of motivating and discouraging messages. Such a surprise manipulation is assumed to be more effective and more comparable than a manipulation of the game score.

5.2 METHOD

5.2.1 PARTICIPANTS

All participants were recruited either via bulletins, which were distributed throughout several buildings of the university, or via the *prometei* online platform for study volunteers. As a prerequisite for partaking in the study, participants needed to have prior knowledge of Tetris, as well as a good level of German. Those from Study I were not allowed to participate in Study II. A total of 63 participants contributed to the study. Ten had to be excluded from further analysis due to technical problems or insufficient Tetris skills

resulting in game over. The remaining 53 players (25 male, 28 female) were equally distributed over the three experimental groups, sex being a balanced factor. The average age was 27.3 years (*SD*= 6.7). The sample consisted of well-educated persons. Roughly one third (32.01%) had a university degree and about two thirds (62.26%) had a final secondary school qualification ('Abitur') as their highest educational qualification at the time of the study. Prior knowledge of Tetris was assessed through a 6-point scale in the demographic questionnaire with the anchors no experience at all (1) and a lot of experience (6). Participants were explicitly advised to take earlier experience from their childhood and youth into account when assessing their experience with Tetris. All of them estimated their mastery of Tetris to be at least fair (3 or higher on the experience scale). Average experience was 4.28 (*SD*=0.989). There were no self-reported differences for the factor *experience* with Tetris between the groups. Participants were reimbursed with ten euros for participation (see section independent variables for a more detailed description of the monetary reward structure).

5.2.2 MATERIALS

As an advancement to study I, a more modern EMG recording instrument was used that enabled automatic setting of markers into the EMG data via a stable USB connection. EMG data was recorded using the Healy Health Lab System by Space Bit, from Kowalski Industry Electronic. Data was recorded over *the lateral frontalis* on the right side of the face with Kendall 24mm disposable electrodes (see 4.5). For everything else, the same apparatus and computers were used as in Study I (see section 4.2.).

5.2.3 PROCEDURE

Upon arrival, participants were seated in front of the experimental computer where they filled in the demographic questionnaire and read the instructions for the study (see appendices B. and B.2). Subsequently, electrodes were placed on the face of participants. Strength of the EMG signal was verified on the experimenter PC. After confirmation of EMG signal strength and adjustment of electrode placement if necessary, a training trial was completed. The training version stopped after 300 blocks and participants were not told to reach a specific number of points. If participants did not have further questions after the training trial and proved to have sufficient mastery of the game, the experimental trial was conducted. In total, all participants had to handle 55 Tetris blocks during the experimental session. In contrast to a conventional Tetris games, the score was calculated by adding 10 points for each block which was displayed on the Tetris field. This made it possible for the game score to serve as a discrete indicator for an unexpected event. At the

same time, it was possible to control how many blocks each participant had handled before an event. Unexpected events were presented three consecutive times during one game. RTs were recorded similarly to study I. After completion of the experimental trial, questionnaires were immediately filled in. Electrodes were removed and after a short debriefing, participants received their monetary compensation and left.

5.2.4 Design

Independent variables

Similar to study I, study II comprised a mixed design with the two variables *group* and *event*. *Group* had three levels: a *positive group*, a *negative group*, and a *neutral group.* Contrary to study I, no control group was implemented. It did not yield additional scientific insights and was thus left out. Participants had to reach 550 points in all groups and were confronted with three unexpected events during their interaction with the game. Events consisted of pop-up messages and auditory signals. Each message had the same number of words and visual information to keep the amount of information processing equal for all three groups. The pop-up windows appeared at the outer top right corner of the Tetris game, not covering any part of the square board. In addition to the visual feedback, every group received an auditory feedback consisting of a neutral white noise (control), applause (positive), and booing (negative). Auditory signals were all of the same length (2 seconds) and approximately 60dB loud. Unexpected events are shown in Figure 5.1. Participants were told that they would be reimbursed with seven euros for participation, and would get an additional amount of three euros if they reached their points on time, without producing a game over. Actually, all participants were reimbursed with ten euros. Only participants who reached the goal were included in data analysis.

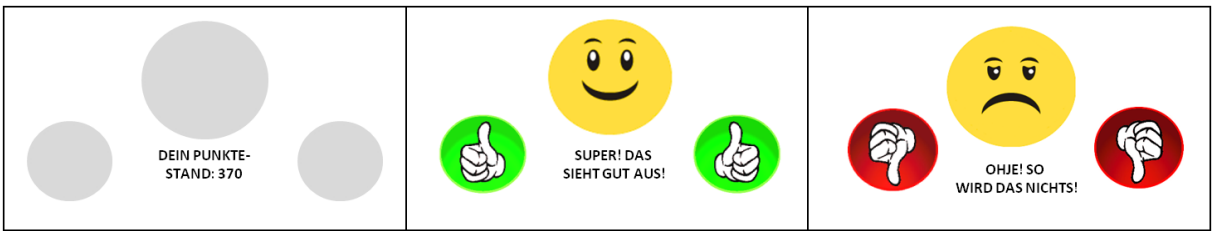


Figure 5.1: Unexpected events for the neutral group (left), the positive group (middle), and the negative group (right).

Negative group (17 participants)

The negative group was surprised by a pop-up window with a negative smiley face and the message 'Ohje! So wird das nichts!' ('Oh no! That's not gonna work!'). Simultaneously, a sound record was played.

Positive group (18 participants)

The positive group was surprised by a pop-up window with a positive smiley face and the message 'Super! Das sieht gut aus!' ('Great. This looks good!'. Simultaneously, a sound record was played.

Neutral group (18 participants)

The neutral group was surprised by a pop-up window with three grey circles and an update about the current score with the message e.g. 'Dein Punkte-Stand: 370' ('your score: 370'). Simultaneously, a white noise sound was played.

The second independent variable was *event*, the time at which an unexpected event occurred. Event had three levels. Participants were confronted with a message on the monitor three times, simultaneously with block 38, 47, and 51. This variable was operationalised as a within-subjects variable.

Dependent variables

There were six dependent variables: two variables to objectively verify the successful evocation of a surprise experience in the participants, and four variables to cover UX ratings. As in study I, RTs were used as an objective measure of surprise. RTs were defined as the time between the appearance of a block with the accompanying event, and the first manipulation of the block via keyboard input. To gain further insight into physical reactions to surprise, facial EMG was conducted, measuring activity of the lateral frontalis (LF) muscle (see Figure 4.5) which is associated with the facial manifestation of surprise reactions (Schwartz et al., 1976; van Boxtel, 2010). To measure UX ratings, three questionnaires were used: the SAM questionnaire for measuring the emotional components of UX, the AttrakDiff mini, and the meCUE questionnaire. For the current study, the modules *User Emotions* and C*onsequences of Use* of the meCUE were employed with a total of 19 items. The meCUE was selected as a replacement of the product emotions questionnaire by Mano (1996). By leaving out the product perception module, possible overlap with the AttrakDiff subscale PQ was avoided. To explicitly measure subjective surprise, a single scale ranging from 1=not surprised at all, to 7=very surprised, was given to participants after completion of all other questionnaires.

5.2.5 HYPOTHESES

All hypotheses concerning user experience are identical to those in study I, except all parts concerning the control group, which has not been implemented in study II. Accordingly, the motivations for identical hypotheses have been left out and can be found in chapter 4.2.5.

Surprise

Following Meyer et al., (1991, 1997), and similar to study I, RT data was employed as an indicator for a successful surprise manipulation with a subsequent surprise reaction in users:

H1: A main effect on RTs is predicted for the factor *event*.

H.1.1 For the positive group, the negative group, and the neutral group, mean RTs will decrease from event 1 through event 2 to event 3.

H2: A main on EMG data effect is predicted for *group*.

H.2.1 For the positive group, the negative group, and the neutral group, mean EMG activity will decrease from event 1 through event 2 to event 3.

These hypotheses are based on findings by Ekman and Friesen (2003). Surprise is accompanied by distinct facial expressions (see section 2.5.6), which should be measurable through EMG. Like RTs, this activity should decrease throughout the emergence of several unexpected events, illustrating the schematization of new information.

User Experience

(Hypotheses are formulated for subscales of questionnaires for which an effect is anticipated, only.)

SAM ratings (measuring emotional reactions to a product):

H3: A main effect is predicted for *group*.

H.3.1 SAM valence ratings made by the positive group will be better than those of the neutral group.

H.3.2 SAM valence ratings made by the negative group will be worse than those of the neutral group.

H.3.3 SAM arousal ratings made by the positive and the negative groups will be higher than those of the control group.

AttrakDiff mini ratings:

H4: A main effect is predicted for *group*.

H.4.1 AttrakDiff ratings on the HQI subscale will be higher for the positive group than for the neutral group.

H.4.2 AttrakDiff ratings on the HQI subscale will be lower for the negative group than for the neutral group.

meCUE ratings:

H5: A main effect for group is expected.

H.5.1 meCUE ratings on the subscale positive emotions, product loyalty, and intention of use will be higher for the positive group than for the neutral group.

H.5.2 meCUE ratings on the subscale negative emotions, product loyalty, and intention of use will be lower for the negative group than for the neutral group.

H.5.3 There will be no effect for the overall rating.

Single item surprise rating:

H.6: A main effect is predicted for *group*.

H.6.1 Ratings will be higher for the positive and the negative group, than for the neutral group.

Due to an action-relevance check (see section 2.5.2) which should be more intensive in the two groups, a higher subjective level of surprise is expected.

5.3 RESULTS

Participants were equally experienced with Tetris between groups ($F(2,50)=1.404$, $p=.255$).

5.3.1 SURPRISE

Raw RT data for all participants, sorted by group already indicated elevated RTs for all three unexpected events (see Figure 5.2). Note how distinct the RT elevations for each event are in both the positive and the negative group. Peaks of RTs are a little less pronounced in the neutral group.

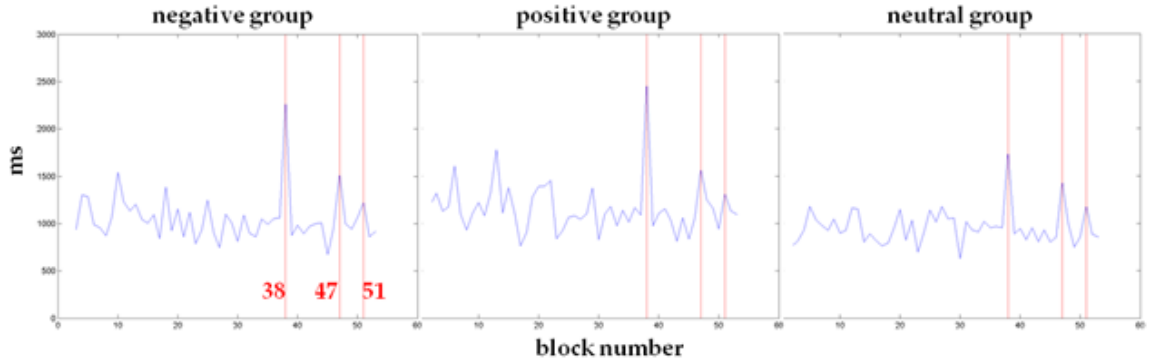


Figure 5.2: Raw RTs per group averaged over all participants in each group.

To properly investigate the effects of the unexpected events on RTs and thus on information processing, a baseline for RTs was calculated as in Study I. Mean RTs from the three blocks prior to the eventful blocks (e.g. block 37, 46, and 51) were averaged. This baseline time was then subtracted from the individual RTs for each block for every participant in each group. Positive values indicate RTs above average RTs for the three blocks prior to the blocks paired with an event. See Figure 5.3 for uncorrected and corrected RTs. Note how the overall pattern of all three groups remains unaffected by the baseline correction. Although all three groups exhibit the usual pattern of RT reduction back to baseline over the three events, a visual inspection indicates a difference between the neutral group, and the negative and the positive group.

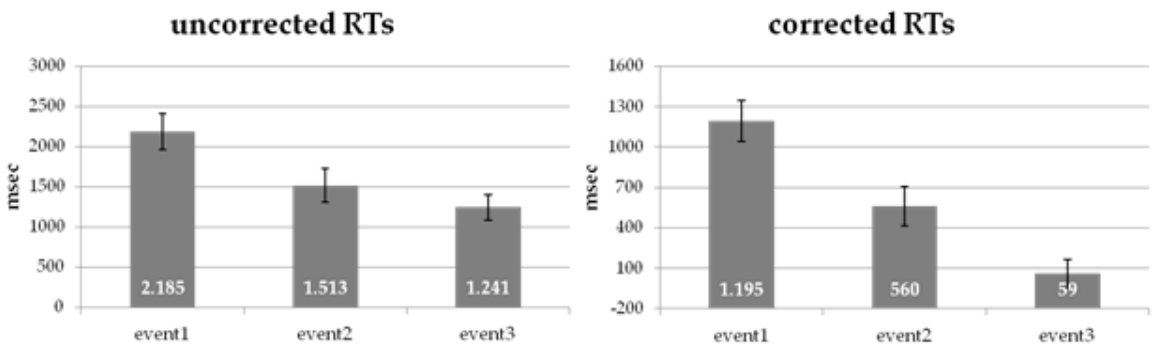


Figure 5.3: Uncorrected RTs (left), corrected RTs (right).

To verify a successful surprise manipulation, RTs were analysed using SPSS. A 3x3 repeated measures analysis of variance (ANOVA) of RTs was carried out with *group* as the between-subjects factor and event as the within-subjects factor. Mauchly's Test of Sphericity indicated that the assumption of sphericity had not been violated ($\chi^2(2) = 1.597$, $p = .450$), hence the degrees of freedom were not adjusted. There was no significant

interaction between group and event (F(4,100)=1.407 p=.237, η_{PART}^2=0.053). To further investigate the apparent RT differences between the groups for event1, a post-hoc one-factorial ANOVA was carried out with *group* as the between-subjects factor. There was a marginally significant effect of group on RTs, F(2,50)=3.04, *p*=.054. There was a significant linear trend F(1,50)=5.446, *p*=.024, indicating that RTs are significantly longer for participants in the positive and the negative group than for participants in the neutral group.

A significant main effect was found for event (F(2,100)=31.25666, p=.000, η_{PART}^2=.385). Contrast analysis attributed this effect to a significant difference between event1 and event2 (p=.000), event1 and event3 (p=.000), and event2 and event3 (p=.002). There was no significant main effect for *group* (F(2,50)=1.040, p=.629, η_{PART}^2=.040), indicating a similar surprise reaction for all participants in all three groups.

5.3.2 EMG DATA

Several participants had to be excluded from EMG data analysis due to technical problems, leaving a group size of 14 (positive group), 14 (negative group), and 13 (neutral group) persons. To reduce movement and blink-related artefacts, as well as the positive skew, EMG data was submitted to a 15-Hz high-pass filter, fully-rectified, and was then subjected to a square-root transformation (Larsen, Norris, & Cacioppo, 2003). EMG reactivity was measured as the difference between the mean activity during 1000ms before the stimuli (e.g. surprise events) and the mean activity during 1500ms after stimuli onset.

The ANOVA on the activity over lateral frontalis with *group* as the between subjects factor

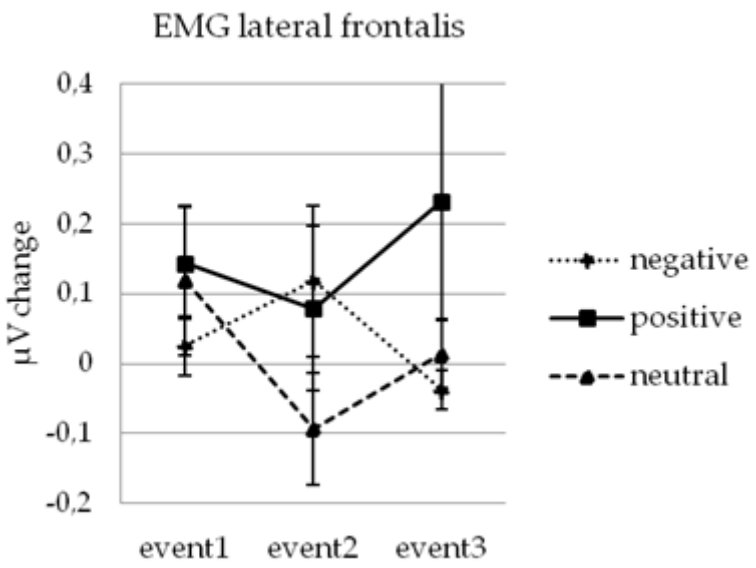


Figure 5.4: µV change over the lateral frontalis muscle for each event with SE.

and *event* as the within-subjects factor revealed no significant main effect for *group* (F(2,38)=3.00, p=.062, η_{PART}^2=.136). Furthermore, there were no significant effects (Greenhouse-Geisser corrected) for *event* (F(3.342, 63.501)=.244, p=.744, η_{PART}^2=.006) and *event***group* (F(3.342, 63.501)=.886, p=.462, η_{PART}^2=.046). Figure 5.4 shows lateral frontalis activities after each unexpected event.

5.3.3 USER EXPERIENCE QUESTIONNAIRES

To investigate surprise effects on UX ratings, a one-factorial MANOVA was carried out with *group* as between-subjects factor (all values z-transformed). The MANOVA revealed a significant multivariate main effect for *group* (Wilk's Λ = 0.330, F(20,82) =3.036, p=.000, η_{PART}^2=.425). Power to detect the effect was .999. Given the significance of the overall test, the univariate main effects were examined. Significant univariate effects for *group* were obtained for SAM Valence (F(2,50)=10.757, p=.000, η_{PART}^2=.301), as well as for SAM Arousal (F(2,50)=6.989, p=.002, η_{PART}^2=.218.), the meCUE product loyalty scale (F(2,50)=3.768, p=.030, η_{PART}^2=.131), and the meCUE negative emotions scale (F(2,50)=3.244, p=.047, η_{PART}^2=.115). No effects were found on any of the AttrakDiff mini scales. Simple contrasts revealed that on the SAM Valence scale, participants in the positive group rated the game significantly better than participants in the negative group (p=.002) and in the neutral group (p=.020). Furthermore, participants in the negative group gave worse ratings than participants in the neutral group (p=0.28). For SAM Arousal, the neutral group indicated lower overall arousal levels than the negative group (p=.003) and the positive group (p=.002). On the meCUE subscale *product loyalty*, participants in the negative group indicated a higher subjective product loyalty than participants in the neutral group (p=.010). On the meCUE subscale *negative emotions*, participants in the negative group indicated a higher level of negative emotions than participants in the neutral group (p=.040) and in the positive group (p=.025). See Table 5.1 for questionnaire ratings for subscales with significant differences between groups. Counterintuitively, a low rating on the SAM valence subscale indicates a happier rating than a high one. Similarly, a low rating on the SAM arousal subscale indicates a higher subjective level of arousal than a high level.

The subjective surprise ratings showed no significant main effect for *group* (F(2,50=1.945, p=.154), although a slight tendency for the positive group and the negative group being subjectively more surprised can be observed (see Figure 5.5).

Table 5.1: Mean UX ratings from questionnaires with SDs.

Questionnaire subscales	*negative (SD)*	*positive (SD)*	*neutral (SD)*
SAM Valence	4.24 (1.64)	2.06 (0.94)	3.17 (1.25)
SAM Arousal	4.29 (0.77)	4.22 (1.66)	5.67 (1.28)
meCUE Product Loyalty	3.82 (0.76)	3.13 (1.34)	2.81 (1.13)
meCUE negative Emotions	2.58 (0.86)	1.92 (0.90)	1.78 (0.77)

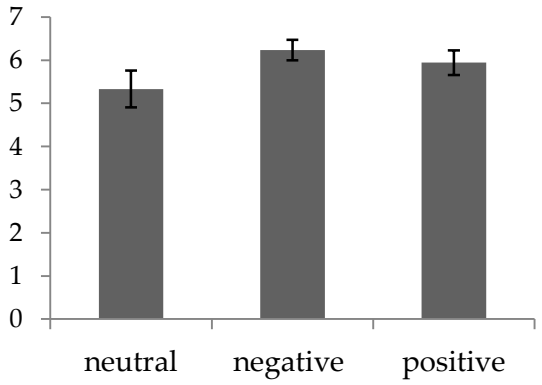


Figure 5.5: Subjective surprise ratings with SE.

5.4 DISCUSSION

The goal of the second study was to test the influence of pleasant and unpleasant surprise on the UX ratings of a digital, interactive game, and to eliminate inconclusive findings from study I.

5.4.1 REACTION TIMES

As in study I, it was assumed that RTs for the reaction to unexpected events would be longer than for expected events and would regress back to a baseline level after repeated presentation of unexpected events. This pattern is typical for RT prolongation associated with unexpected events (see section 2.5.2). To account for inter-individual difference in RTs, a baseline was subtracted from the raw RTs, consisting of the mean three RTs prior to each event. This approach did not have any influence on the overall RT pattern (see Figure 5.3). A distorting effect of this approach can thus be ruled out.

Support for the hypothesis concerning a prolongation of RTs after a surprise was found conducting an ANOVA. The significant prolongation of RTs for surprising events in all

three groups shows the success of the surprise manipulation. A significant decrease of this effect from event1 to event2 to event3 is a sign of the typical disappearance of the surprising characteristic of an unexpected event after several occurrences, confirming H1 and H.1.1. This indicates the typical disappearance of the surprising characteristic of an unexpected event after several occurrences (Meyer et al., 1991, 1997). Based on these RT analyses, it can be assumed that the surprise manipulation was successful for both pleasant and unpleasant surprise. Because visual inspection of the RT data for event1 shows a pronounced difference in RTs between positive and negative, and neutral group, a linear trend analysis was carried out. It indicated that for the first event, participants in the neutral group exhibited significantly lower RTs than participants in the other two groups. In accordance with the cognitive psycho-evolutionary model of Meyer et al. (1997), this can be interpreted as evidence of a less pronounced action-relevance check in the neutral group. Information displayed in the pop-up message for the neutral group was redundant because score count was displayed at all times at the right side of the game. Motivational and discouraging messages on the other hand included information relevant to the participants: the probability of winning the game was conveyed.

5.4.2 EMG

It was assumed that the activity of the lateral frontalis muscle should be elevated for event1, when compared to activation during a baseline phase (H2). A similar pattern of EMG data was expected as for RT data. Due to the schematization of newly acquired information, the activity of the muscle should decrease from event1 over event2 to event3 (H.2.1). Although the activity of the lateral frontalis muscle has been associated with the typical facial expressions of surprise (e.g. opening of the eyes, raising of the eyebrows; see van Boxtel, 2010), no such pattern could be found in the data (see Figure 5.4). Because data was carefully selected and filtered, it is assumed that no erroneous data was included in the EMG analysis. The question thus remains why the measurement of muscle activity did not prove to be associated with surprise reactions. Over the course of eight studies, Reisenzein, Bördgen, Holtbernd and Matz (2006) found a strong dissociation between surprise and facial expression, although participants were indeed experiencing the mental state of surprise. While the experimental manipulations had the predicted effects on other objective behavioural measures (RTs), as well as on subjective measures of surprise, facial expressions were not affected. Additionally, dissociation between facial displays of surprise and beliefs about this display was found: although participants did not exhibit facial components of surprise, they still reported that they had been surprised. One possible explanation for this behaviour is the insufficient sociality hypothesis brought

forward by Reisenzein et al. (2006). Participants might not exhibit facial expressions due to internalized social or personal norms. In addition, facial expressions are used as a means of social communication (Frijda, 1986; Izard, 1977; see section 2.5.6). In an experimental setting, no social communication is involved because the participant is seated alone in front of a computer. Lack of facial expressions could thus be the result of insufficient sociality and lack of social interaction during the experiment.

5.4.3 UX QUESTIONNAIRES

Research on surprise as a design strategy has shown beneficial effects on the appraisal of a variety of non-digital artefacts (see Ludden, 2008, for an overview). The second goal of the study was thus to further investigate findings from study I: concerning the effects of surprise on product ratings and UX. In study II, RT data confirmed a successful surprise evocation in users, making the interpretation of UX questionnaire data more conclusive than in study I

SAM

It was assumed that pleasant events would lead to positive valence ratings, whereas unpleasant events would decrease those ratings in comparison to neutral events. Accordingly, concerning SAM ratings, a main effect was expected for *group* (H3). These assumptions could be confirmed for the SAM valence subscale. Participants in the positive group rated the game as significantly better than participants in the other two groups (H.3.1). Furthermore, participants in the negative group rated the game significantly worse than participants in the neutral and the positive group (H.3.2). Differences on the SAM arousal scale nicely illustrate the importance of valence for surprise. Both surprise manipulations lead to a higher arousal level in the users when compared to the neutral surprise manipulation (H.3.3). All in all, subsequent differences in valence ratings can thus be attributed to pleasant and unpleasant surprise incidents during the interaction with the game. The differences in valence ratings on the SAM subscales support the initial hypotheses of pleasant surprise elevating valence ratings and vice versa. The beneficial effects of surprise on product ratings, as found in the systematic research on non-interactive products (see Ludden, 2008) could thus be transferred to digital interactive products. Additionally, a negative effect of unpleasant surprise on the affective state of users could also be demonstrated.

AttrakDiff mini

No significant difference was found on any of the AttrakDiff subscales.

meCUE

A main effect for group was found for the meCUE questionnaire (H5). The impact of valenced surprises on UX ratings could only be demonstrated for negative surprise. Positive surprise did not lead to more positive emotions concerning the game (H.5.1). The assumption that unpleasant surprise has a negative effect on UX ratings was supported by the differences between groups on the meCUE subscale *negative emotions*. Participants in the negative group indicated higher levels of negative emotions than participants in the other groups (H.5.2). Product loyalty was not elevated by positive surprise (H.5.1), but in the negative group participants indicated a greater inclination to like the product better. Findings from the meCUE questionnaire thus add some interesting information to the insights of emotional product design: product loyalty scores were significantly higher in the negative group than in the neutral group. One possible explanation for this apparent paradox could be the context in which UX was investigated. It is the challenge factor that creates interest for a game (Hsu & Lu, 2004; Sweetser & Wyeth, 2005). Games should always include a challenging factor in order to not become boring (Pagulayan, Keeker, Fuller, Wixon, Romero & Gunn, 2002). In a gaming-context, a challenging or unpleasant surprise could elevate the overall product loyalty of a game, although affect ratings indicate otherwise. Practically speaking, the Tetris game in the negative group actively told participants that they are no good at playing the game. All participants achieved their goal later, thereby proving the games predictions to be wrong. This could have resulted in a higher liking of the game: the game was conquered. This finding also supports a design principle put forward by Fokkinga and Desmet (2013). They propose a rich quality of experience by letting users experience challenging emotions. No differences were found on the subscale *intention of use* (H.5.1). As predicted, no difference could be established between the three groups concerning the overall rating (H.5.3).

Single item surprise rating

It was expected that due to a more intensive action-relevance check, participants of the positive and the negative group would indicate subjectively higher ratings of surprise than participants in the neutral group (H.6.1). Although ratings from the positive and the negative group are slightly elevated, this difference was not significant. This finding can be attributed to the fact that feeling surprised does not always equal awareness of surprise in a person (Chalmers, 1995). Although participants were surprised, they could not report these levels of surprise due to a lack of introspection. The result can be an apparent incongruence between objective measures of surprise (RTs) and subjective measures of surprise (questionnaires).

5.4.4 SUMMARY AND CONTRIBUTION

Eliciting positive emotions through surprise has been shown to work very well in classical product design (Ludden, 2008; Ludden et al., 2008; Ludden et al., 2009). In the UX community, it is largely agreed that a positive UX is a conglomerate of factors such as experience, design, and emotions (Law et al., 2009). One promising way to promote a pleasant UX could be the creation of a positive surprise for users of digital products. In addition to investigating the possibility of this transfer from classical product design studies of pleasant emotions, the severity of possible negative effects following an unpleasant surprise was also investigated. Considering the influence of surprise on UX ratings, the results indicate the success of a transfer of the findings from classical surprising product design. Several points should be stressed while discussing the findings. Pleasant surprise has a positive effect on UX ratings by elevating affect ratings as well as hedonic ratings of a product. Unpleasant surprise has the opposite effect on UX ratings, leaving users in a worse mood after interaction than they were before as has been shown in earlier studies (Gross & Thüring, 2013). Interestingly, although users indicate worse UX ratings after they have been unpleasantly surprised, they show higher product loyalty concerning the Tetris game than users who experienced a pleasant or neutral surprise. According to findings from Fokkinga and Desmet (2013) this finding can be explained in terms of frustration and satisfaction, two feelings that are part of a challenging experience. By telling participants that they will not succeed in the game, a certain frustration is built up which is later resolved into a feeling of achievement when the game has been conquered. The excitement of a computer game results from the paradoxical relation between frustration and achievement. Furthermore, it gives the participant a pleasant feeling in the end, resulting in a higher product loyalty compared to the other games.

From a marketing perspective, Study II thus raises several questions. Since pleasant surprise has a positive impact on UX ratings, possibly increasing the market position of a product, it could be beneficial to create surprise. This could be implemented from the beginning of the use of a product by concealing some positive features of a system so that the user can discover them unexpectedly. But is it more beneficial to tell customers all positive aspects to prompt them to purchase the system, or is it better to let them discover some surprising extras later, which might pay off in the long run by increasing brand loyalty? Furthermore, the challenge factor created by unpleasant surprise during the game seems to have elevated the product loyalty. Outcomes of the study thus did not deliver clear cut implications for product designers in terms of using surprise deliberately as a design feature. Both manipulations had a positive effect on UX, the question remains which manipulation is advisable in which context of use. Obviously, the results are not

far-ranging enough to provide a sound answer, but future investigations may shed more light on this issue.

5.4.5 CRITICAL EVALUATION

The study has some limitations: One major limitation of the study is the choice of application. By choosing Tetris, the research was limited to the domain of digital games. Games are designed to enhance players' emotional states, so experienced gamers could have been conditioned to experiencing surprise and vice versa. Therefore, a generalization of the findings to other product domains is difficult due to the special features of games. A negative surprise which had some positive effects on product loyalty could be absolutely devastating for UX in a goal-directed context, where reaching a goal quickly and efficiently is the user's first priority. The challenge factor of the Tetris game is what makes it interesting for users; however a goal-directed task should be challenge-free to ensure easy, quick, and successful completion, adhering to usability standards. To make up for issues which could possibly arise from the choice of context of use, a change of settings is appropriate for future studies. By extending this study using different technologies in a different domain which is not necessarily goal-directed and with more common-place technologies like mobile phones or tablets, valuable insights could be gained.

6 STUDY III: THE INFLUENCE OF UNEXPECTED EVENTS IN A GOAL-ORIENTED TASK ENVIRONMENT

6.1 PROBLEM AND RESEARCH GOALS

The two previous studies investigated unexpected events in a gaming environment, partly reproducing earlier findings from product design. In the 3rd study, the concept of unexpected design is transferred to a task-, or goal-oriented environment. The research goals of this study are:

- Investigating the influence of unexpected visual elements on the UX of a goal-oriented mobile application
- Investigating different frequencies of these unexpected events in the corresponding effects on UX of a goal-oriented mobile application.

Knowing the possible impact of unexpected and thus surprising behaviour of goal-directed applications on the UX ratings of these applications is valuable to UX research and practice because of two reasons: First, it could give designers a clear directive whether it is useful to deliberately use surprise as a design feature in goal-directed mobile applications. Second, it could help to elevate the market value of a product by following the recommendations for surprise as a design element dependent on context.

Surprise can have a profound influence on users in classical product design (Ludden et al, 2009). There is little literature in the domain of digital interactive products, except for digital interactive games. In two studies, negative and positive surprise had an effect on the UX ratings of a digital desktop game (Gross & Thüring, 2013). To this date, no studies exist about goal-oriented digital mobile applications, to the knowledge of the author. Taking into account the different contexts between a game and a goal-directed or task-

---

Chapter 6 is partly based on: Gross, A. & Silvennoinen, J. (2014). Surprise as a Design Strategy in Goal-Oriented Mobile Applications. In *F. Rebelo & M. Soares (Eds.), Proceedings of the 5th AHFE Conference 19-23 July 2014. Advances in Ergonomics in Design, Usability & Special Populations. Part II.* (pp. 4716-4726).

oriented application, the general hypothesis is that unexpected events during the interaction with a mobile application have an effect on the UX ratings of this application.

Furthermore, it is assumed that surprise conveyed via manipulation of visual design elements will make the task itself surprising, but may have a negative influence on UX. This hypothesis is motivated by the fact that surprising unexpected behaviour might be beneficial for a game but might interrupt the task-flow when a goal-oriented application is used.

Additionally, it is assumed that when the unexpected event occurs in every task it is not seen as surprising anymore, but rather perceived as being annoying, thus having a strong negative effect on UX.

6.2 METHOD

6.2.1 PARTICIPANTS

Participants were recruited at the TU Berlin via the prometei online base for study volunteers (Probandenserver.) A total of 34 participants took part in the study (16 male, 18 female). They were equally distributed over the two experimental groups. Average age of participants was 25.3 years (SD=4.17). All participants had a secondary school qualification ('Abitur') or higher. 85% possessed a smart phone, and 53% were regular users of applications for public transport.

6.2.2 MATERIALS

A public transport application as described in section 3.2.2 was tested on a Samsung Galaxy Note 10.1 tablet (GT-N8000) with an Android version 4.1.2 operating system. The study was recorded with a GoPro Hero3 White™ camera with a wide angle lens, positioned on the desk in front of the tablet. Instructions for each task were printed out and placed in a folder. Participants had to turn over a page to get the instructions for the next task (see Figure 6.1).

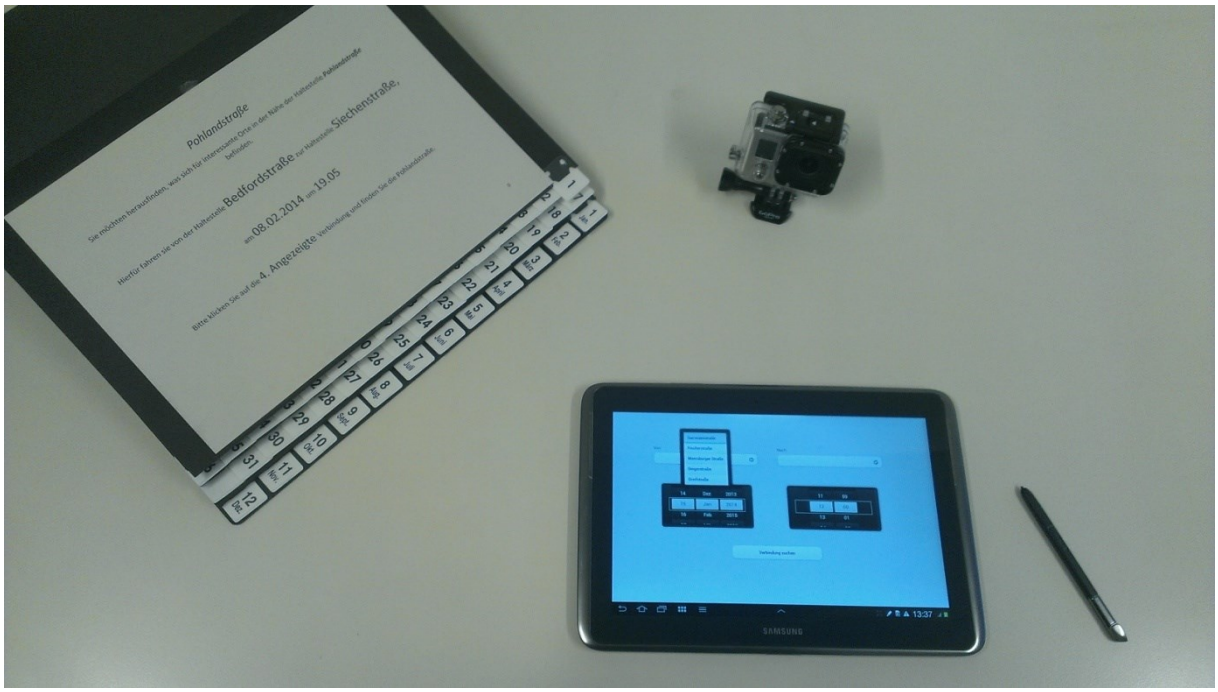


Figure 6.1: Experimental setup with tablet, camera, and folder with experimental tasks.

Task and stimuli

Participants were presented with a sequence of simple data entry and data selection tasks. One task consisted of entering a start time and a date, as well as a start and end station into the first screen of the app. On a next screen, participants were presented with four possible public transport connections of which only one was the correct one. They had to tap on that connection to get to the next screen where they were presented with a detailed list of all stations for the chosen line. One station was marked with an information icon. Participants had to press that icon to get additional information about the surroundings of the station. After pressing the information icon, the new information appeared together with a “next” (German: “weiter”) button. Participants tapped on that button in to reach the start screen of the next task. Figure 6.2 shows all four screens of the application in normal mode, without any surprising events.

The prerequisite for designing unexpected events during the interaction with the application was the unobtrusiveness of the stimuli: the events should not complicate using the app. Furthermore, the functionality and usability of the app should not be altered by including the surprising stimuli. Therefore, unexpected events were created through the manipulation of visual elements in the user interface (UI). In UIs, communication is achieved through the manipulation of visual elements such as shape and colour (Mullet & Sano, 1995). Furthermore, vision is regarded as an essential sensory modality in product experience (e.g. Crilly, Moultrie & Clarkson, 2004).

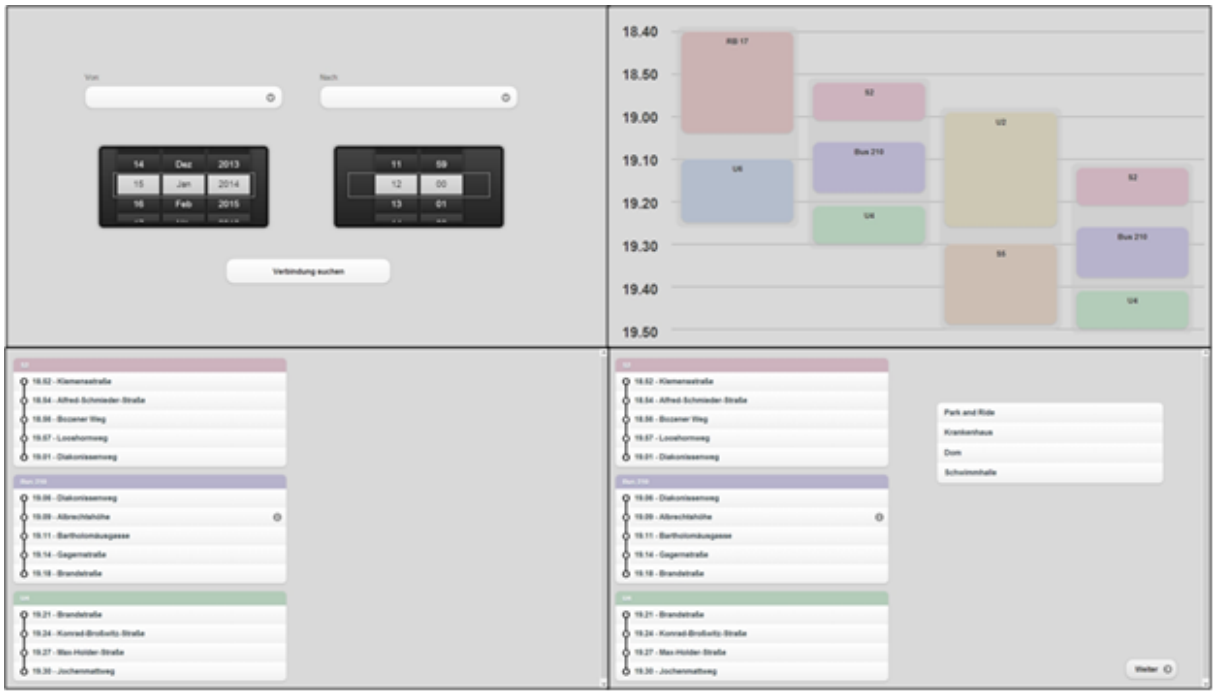


Figure 6.2: UI of the application screens 1 to 4 (screen1: upper left, screen 2: upper right; screen 3: bottom left, screen 4: bottom right)

Mullet and Sano (1995) define the visual characteristics of an object as consisting of shape, colour, texture, size, and orientation. It was chosen to manipulate the visual elements of the representation of one means of transport (e.g. a rectangular form) in one of the displayed connections in each task. Manipulations were developed systematically, always altering two of four visual characteristics for each element (shape, colour, texture, or size). Every visual characteristic was used in the manipulation five times. Additionally, saturation of colour and animated movements were implemented for each visual event. Colour saturation and animated movements have been proven to bind users' attention, making sure participants would not miss the visual events during interaction (Lidwell, Holden & Butler, 2003). In practice, an unexpected event in the application would present itself as follows: during the execution of one task, participants get to a screen with four possible connections displayed (see Figure 6.2, upper left). Every connection consists of several rectangular shapes representing one means of transportation (i.e. bus, metro, tram, etc.). To gather further information about one station, participants have to select one of the rectangles with a tapping gesture. It then either changes shape, texture, or colour saturation. Additionally to two of these changes, it moves across the screen. Figure 6.3 shows screen two of the application with (right) and without (left) an animation, in this case animation number 8 with manipulations of colour and shape.

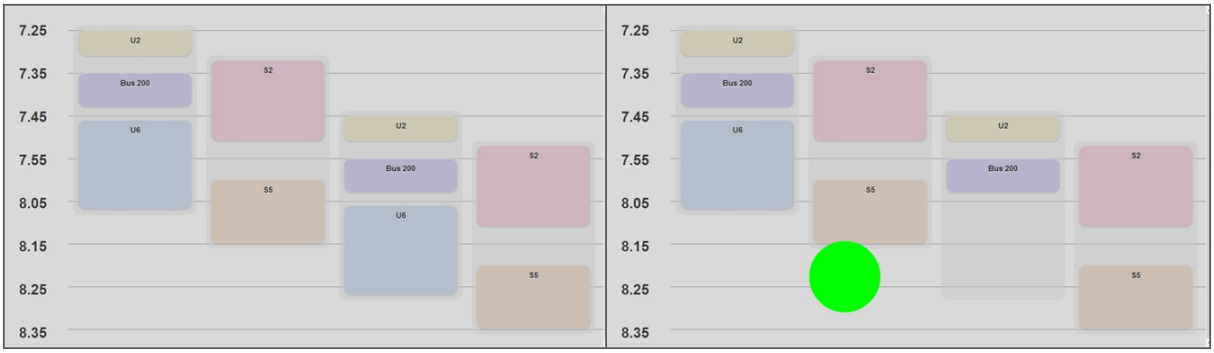
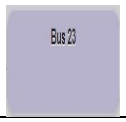
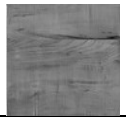
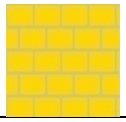
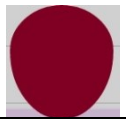
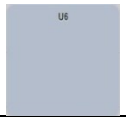
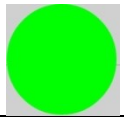
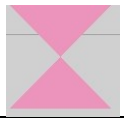
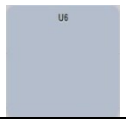
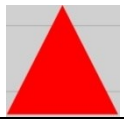


Figure 6.3: Example of a manipulation following a touch on the element "U6".

Table 6.1 shows all ten visual manipulations, including the two characteristics that have been manipulated, the original shape and a description of the animated movement path.

Table 6.1: Original shapes and their corresponding visual manipulations and movement paths.

Original shape	New shape	Visual manipulations	Animated movement path
		texture size	spinning & shrinking
		shape size	spinning & growing
		colour texture	Falling to the bottom
		texture size	Growing over the whole screen, then shrinking
		texture size	Zigzag falling
		texture colour	Flying over whole screen
		colour shape	Flying to the top

		colour shape	Small helix then straight to bottom
		shape size	Top and bottom extend, then fall
		colour shape	Jumping around at the bottom

6.2.3 PROCEDURE

The study was conducted in a laboratory of the Technische Universität Berlin. Completion of the study took about 45 minutes per participant. Users accomplished 25 tasks with a transport application to search and select routes. Upon arrival, participants completed a demographic questionnaire, and were given an introduction to the application. In order to maintain a high level of commitment, participants were asked to imagine having to plan a friend's visit to Berlin. Finding information about the vicinity of several public transportation stations was one part of this planning phase.

In a training phase, participants were given five tasks to solve with the application with the help of the experimenter. During the training, they could familiarize themselves with the data entry format and the design of the application. A block of ten tasks followed, after which the SAM and the meCUE questionnaires were completed (this first block served as baseline). After the baseline, another block with the identical 10 tasks followed, during which unexpected events were encountered, either three or ten times (this second block was called the experimental phase). Task sequence was identical between baseline and experimental phase. After the experimental phase, the SAM and the meCUE questionnaires had to be completed a second time. The sequence of presentation of the tasks was random except for tasks 18, 22, and 25 in group A, and tasks 16, 17, and 18 in group B. All participants were informed about the real purpose of the study after completion.

6.2.4 DESIGN

The study comprised a between-within 2x3 factorial design. Participants were equally distributed over two groups. *Event* was treated as a within–subjects factor with repeated measurements and three levels. *Group* was treated as the between-subjects factor with two levels. See Figure 6.4 for the experimental design.

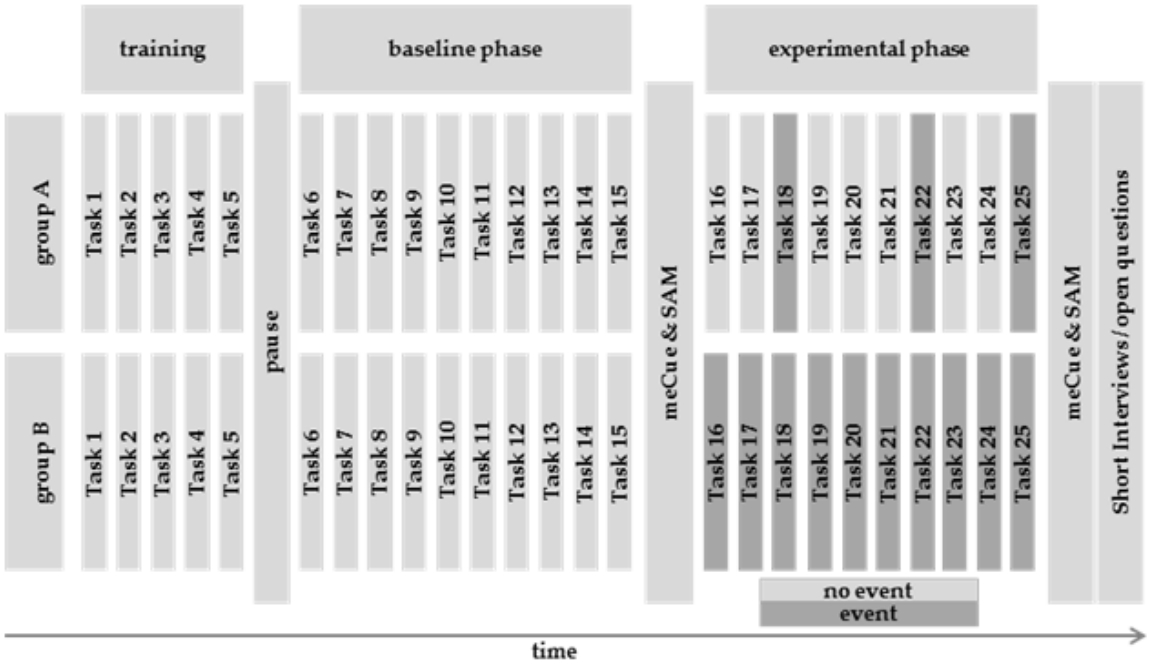


Figure 6.4: Experimental design with moment of appearance of unexpected events.

Independent variables

Group was the first independent variable and had two levels. The difference between the two groups manifested in the experimental phase, only. The first 15 tasks were identical for all participants in both groups, although counterbalanced to avoid sequence effects.

Group A (3 events)

Group A solved a total of 25 tasks. During tasks 18, 22, and 25 of the experimental phase, visual elements of the application were manipulated in an unexpected way.

Group B (10 events)

Group B also solved a total of 25 tasks, thereby encountering 10 unexpected visual events. Events appeared when tapping on one of the four connections, converting the corresponding rectangle shape into a new shape which started moving around on the screen, as described earlier.

Event was the second independent variable and had three levels. Participants encountered several unexpected events during interaction with the application: group A during tasks 18, 22 and 25; group B during all ten tasks in the experimental phase (tasks 16-25).

Dependent variables

There were four dependent variables: to measure valence and UX ratings, the SAM questionnaire and the meCUE questionnaire were used. In order to assess subjective surprise, a single item ranging from 1=not surprised at all, to 7=very surprised, was given to participants after completion of all other questionnaires. RTs were defined as the period between first appearance of the third screen of the application and the subsequent tapping of the information icon displayed for one of the stations (see Figure 6.5).

Figure 6.5: Screen three with information icon displayed for one station.

To properly investigate the effects of the unexpected events on RTs, a baseline for RTs was calculated by averaging individual RTs per participant per group in all ten events from the baseline phase. These times were then subtracted from RTs of the three first unexpected events per group. Because group A experienced only three unexpected events and group B ten unexpected events, only the first three events in group B were compared with the three events in group A. RTs were analysed like this to keep the novelty of the events as comparable as possible. Comparing event 18 from group A with event 18 from group B would not yield the same amount of information since group B would already have seen an unexpected visual event the third time, whereas group A would not have had this experience.

After completion of the questionnaires, participants were given two open questions to gather further insights into their feelings towards the application[13]:

[13] 1. Hat Ihnen die App gefallen? Ja/Nein/Neutral
- Können Sie 3 Dinge nennen warum/warum nicht?

2. Würden Sie diese App weiterhin benutzen? Ja/Nein/Neutral
- Können sie 3 Gründe nennen warum/warum nicht?

1. Did you like the app? yes/no/neutral
    - Could you name three reasons why/why not
2. Would you continue using this app? yes/no/neutral
    - Could you give three reasons why/why not?

Furthermore, they were asked to comment on the animations (the unexpected events) they encountered:

3. How do you rate the animations?

6.2.5 HYPOTHESES

Following Meyer et al., (1991, 1997), and similar to studies I and II, RT data was employed as an indicator for a successful surprise manipulation. The motivations for hypotheses identical to studies I & II have been left out and can be found in chapter 4.2.5.

Surprise

H1: A main effect on RTs is predicted for the factor *event*.

H.1.1 For group A and B, mean RTs will decrease from event 1 over event 2 to event 3.

User Experience

(Hypotheses are only formulated for subscales of questionnaires for which an effect is anticipated.)

SAM ratings:

H2: A main effect is predicted for *group*.

H.2.1 SAM valence ratings made by group A will be better after the experimental phase than after the baseline phase.

H.2.2 SAM valence ratings made by group B will be worse after the experimental phase than after the baseline phase.

H.2.3 SAM arousal ratings made by group A, will be higher than those of group B after the experimental phase.

meCUE ratings:

H3: A main effect is predicted for *group*.

3. Was halten Sie von den Animationen in der App?

H.3.1 meCUE ratings on the subscales positive emotions, product loyalty, and intention of use will be higher for group A than for group B.

H.3.2 meCUE ratings on the subscale negative emotions will be higher for group B than for group A.

H.3.5 meCUE ratings on the overall rating will be higher for group A than for group B.

Single item surprise rating:

H.4: A main effect for group is expected.

H.4.1 Ratings will be higher for group A than for group B.

A repeated presentation of the stimulus will lead to schematization of the event which will thus no longer be appraised as unexpected. It is assumed that repeated presentation of an irrelevant event will be perceived as annoying. The incorporation of the event into the schema of the application will eventually lead to the appraisal of the event as non-surprising. It is assumed that based on the higher frequency of presentation of the event, participants in group B will not rate the event as being surprising after completion of the study.

6.3 RESULTS

Due to an incomplete data set, one participant from group B was excluded from further analysis.

6.3.1 MANIPULATION CHECK

It was assumed that the animations would elicit neutral surprise. Accordingly, manipulations were designed to be neutral, not eliciting pleasant or unpleasant reactions. Pre-test data was collected to ensure that all 10 animations were evaluated equally. Five participants were asked to sort paper representations of the animations into a grid according to how much they liked the animations/shapes, after having experienced all ten animations on the tablet computer (see Figure 6.6). The rating scale ranged from 1=did not like at all, to 7=liked very much. Mean rating of the animations was 3.92 (SD=1.02). Overall, ratings ranged from 2 to 6, mainly varying around the mean value of the scale. Accordingly, it was assumed that all animations were equally neutral in valence.

Figure 6.6: Example of sorting of paper representations of animations according to liking.

6.3.2 REACTION TIMES

As explained, only the first three tasks associated with a visual event were analysed in each group to maintain comparability of surprise levels. A 3(*event*) x 2(*group*) repeated measures ANOVA revealed a significant main effect for *event* (F(1.617, 50.139)=4.256, p=.027 η_{PART}^2=.121; Greenhouse-Geisser corrected) Simple contrasts revealed the source of this effect to be a significant difference of RTs between event1 and event3 (p=.002), as well as between event2 and event3 (p=.033). ANOVA did not reveal a significant main effect for *group* (F(1,31)=.403,p=.530, η_{PART}^2=.013) or an interaction between *group* and *event* (F(1.617, 50.139)=.355,p=.657, η_{PART}^2=.011). A separate ANOVA of RTs for the first event in both groups showed that the seemingly large difference in RTs (859ms vs. 494ms) was not statistically significant (F(1,31)=1.055, p=.312).

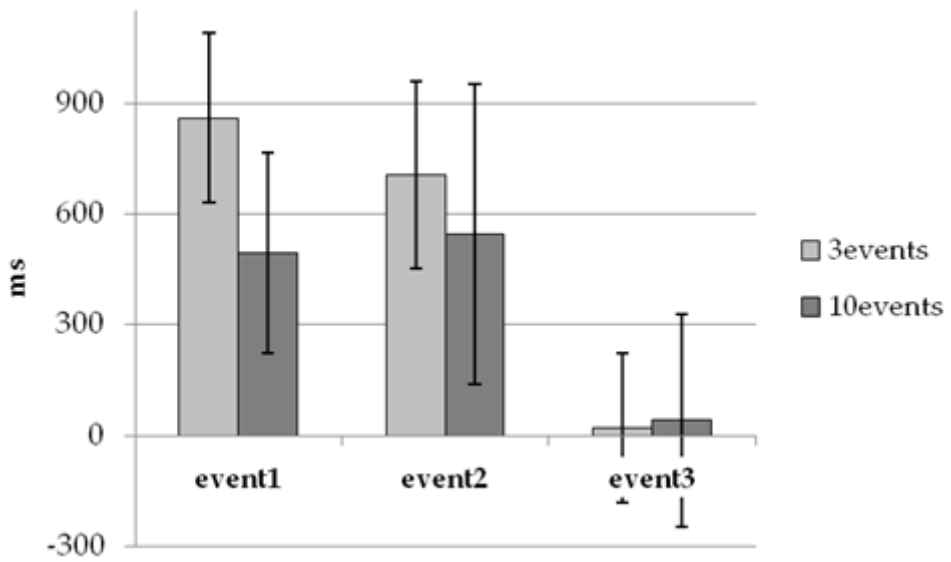


Figure 6.7: Baseline corrected RTs for both experimental groups per event.

6.3.3 USER EXPERIENCE QUESTIONNAIRES

Figure 6.8 shows mean SAM ratings per group with subscales *valence* and *arousal* after the baseline phase and after the experimental phase. It was assumed that both groups would rate the application identically after the baseline phase because no unexpected events had been encountered at that time. A oneway ANOVA revealed no significant differences between groups after the baseline phase for *valence* ($F(1,31)=2.182$, $p=.150$, $\eta_{PART}^2=.066$) and *arousal* ($F(1,31)=1.092$, $p=.591$, $\eta_{PART}^2=.009$), indicating that the two groups did not differ on these scales after having interacted with the application for the first 15 non-eventful trials. A one-way ANOVA with SAM valence ratings of group A as dependent variable and time of measurement as between subjects variable revealed no significant main effect for time of measurement ($F(1,32)=0.631$, $p=.433$, $\eta_{PART}^2=.019$). Again, no effect was found for SAM arousal ratings ($F(1,32)=0.222$, $p=.641$, $\eta_{PART}^2=.007$). A one-way ANOVA with SAM valence ratings of group B as dependent variable and time of measurement as between subjects variable revealed no significant main effect for time of measurement ($F(1,30)=0.942$, $p=.340$, $\eta_{PART}^2=.030$). Similarly, no effect was found for SAM arousal ratings ($F(1,30)=0.053$, $p=.820$, $\eta_{PART}^2=.002$). After the experimental phase, ANOVA revealed a significant main effect between groups on the SAM valence scale ($F(1,31)=9.335$, $p=.005$, $\eta_{PART}^2=.231$), but no significant main effect on the SAM arousal scale ($F(1,31)=.009$, $p=.926$, $\eta_{PART}^2=.000$).

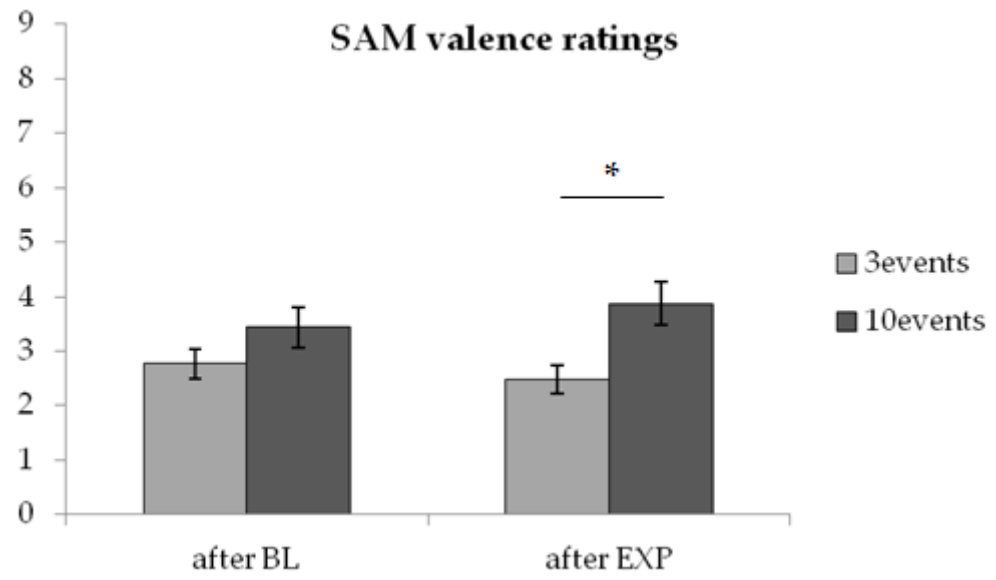


Figure 6.8: Mean SAM valence ratings for group A and B after the baseline phase and the experimental phase with SE.

MeCUE ratings range from 1 (disagreement with a statement) to 7 (agreement with a statement). To compare ratings after baseline phase and after experimental phase, meCUE ratings after the baseline phase were subtracted from ratings after experimental phase. A positive value indicates a higher agreement with the meCUE statements after task completion; a negative value indicates a lower agreement with statements of the meCUE scales after being exposed to the unexpected visual manipulations. Values in

Table 6.2 range from -0.65 to +0.52. Independent T-tests were conducted for every scale separately. A significant effect was found only on subscale *negative emotions*: After completion of all tasks, participants in group A expressed less negative emotions towards the application than participants in group B. Because the Levene's Test for equality of variances was significant, corrected degrees of freedom are reported: (t(20.279)=1.206, p=.032(one-tailed)).

Surprise was reported on a single-item scale ranging from 1=not surprised at all to 7=very surprised. Mean surprise indicated in group A was 5.47 and in group B it was 5.16. An independent samples T-test revealed no significant differences between the two groups, t(33)=.075, p=.512, Cohen's d=0.02611. Formula used to calculate Cohen's d was d=(t*2)/(√ df).

Table 6.2: Absolute differences between BL and EXP for group A and B with SD.

meCUE subscale group	*usability (SD)*	*aesthetics (SD)*	*status (SD)*	*utility (SD)*	*bonding (SD)*
A – 3 events	0.02 (0.52)	0.33 (0.94)	0.12 (0.83)	-0.06 (0.89)	-0.12 (0.33)
B – 10 events	0.02 (0.51)	0.10 (0.65)	-0.28 (0.65)	-0.20 (0.58)	-0.16 (0.36)
	positive emotions (SD)	*negative emotions (SD)*	*intention of use (SD)*	*product loyalty (SD)*	*global rating (SD)*
A – 3 events	-0.07 (0.68)	-0.65 (1.28)	0.08 (0.67)	0.22 (0.47)	0.52 (1.20)
B – 10 events	-0.04 (0.64)	0.00 (0.47)	-0.18 (0.55)	0.08 (0.51)	0.07 (0.67)

6.3.4 OPEN QUESTIONS

Absolute numbers will be reported, rather than percentages given the small sample size (17 in group A, 18 in group B). In group A, 11 participants indicated that they liked the app, compared to 7 participants in group B. Furthermore, 5 participants from group A indicated a neutral opinion, as did 7 participants from group B. One participant from group A did not like the app, compared to 4 participants in group B (see Figure 6.9). Main reasons for a positive evaluation of the application were that it was easy to use, efficient, clearly arranged, and nicely designed. Main reasons given by participants that did not like the app were slowness, imprecise input method and overall operation. Animations were only mentioned by four participants when explaining their like or dislike of the application.

When asking whether participants would continue using the application, a similar pattern was found. In group A, 9 participants indicated that they would continue using the app, compared to 5 participants in group B. 7 participants in group A indicated a neutral opinion, compared to 9 participants in group B. One participant from group A would not continue using the app, compared to 4 participants in group B (see Figure 6.10). Reasons for a prolonged use of the application were similar to reasons for liking the application: that it was easy to use, efficient, clearly arranged, and nicely designed. Furthermore, the additional vicinity information for some stops was appreciated. Reasons given by participants who did not like the app were imprecise input method, the design, the animations, and that there are better applications on the market.

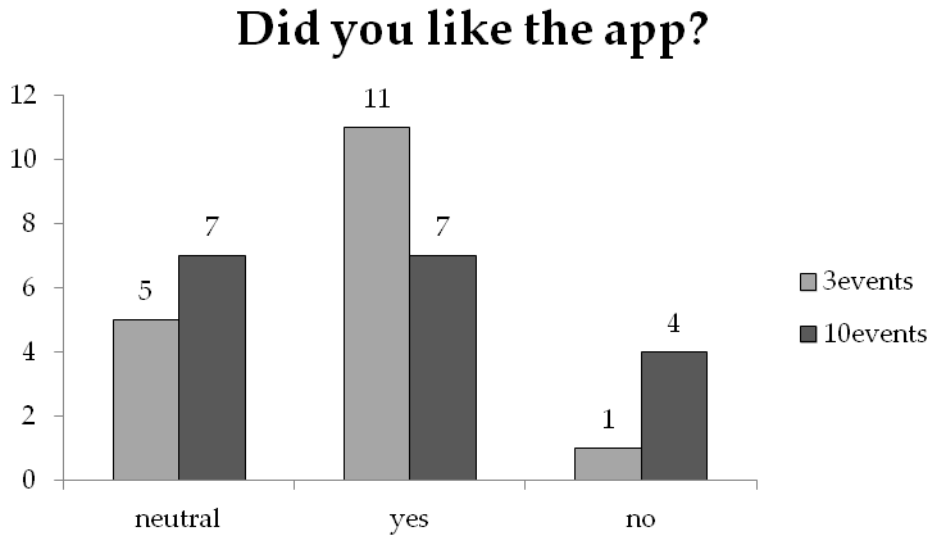


Figure 6.9: Answers to open question 1 (absolute frequencies).

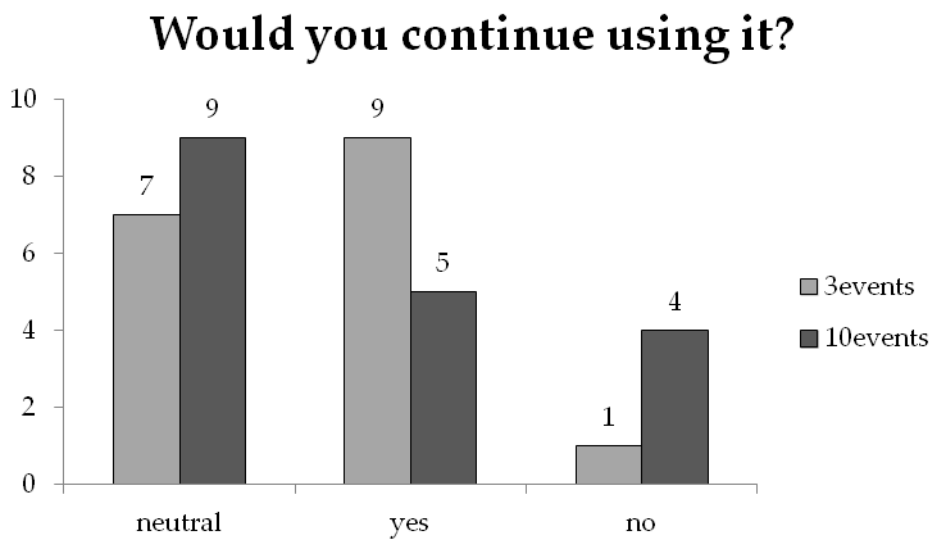


Figure 6.10: Answers to open question 2 (absolute frequencies).

When asked to rate the animations in the application, most participants found the animations as unnecessary, annoying, or irritating (24 participants), whereas only 8 found them as interesting, funny, amusing, or good. 12 participants from group A, and 13 participants from group B did not like the animations. Only 5 participants from group A and 3 from group B liked them. Because expected frequencies for liking of the animations were under 5, a X^2-Test could not be performed (see Table 6.3). Instead, a Fisher's Exact Test was calculated and did not find an effect of group on ratings of the animation ((n=33), p=.381))

Table 6.3: X^2 table of rating of the animation per group.

animation rating group	*negative (expected)*	*positive (expected)*	*total*
A – 3 events	12 (12.9)	5 (4.1)	17
B – 10 events	13 (12.1)	3 (3.9)	16
total	25	8	33

6.4 DISCUSSION

The aim of the third experiment was to investigate the influence of surprise and its frequency on the UX ratings of a digital goal-oriented application. Two different versions of the application were presented to participants, one equipped with three unexpected events, and the other one with ten. It was hypothesised that these events would elicit surprise at first, but would become annoying upon frequent repetition. The resulting effects on RTs and UX ratings will be discussed.

6.4.1 REACTION TIMES

As in studies I and II, it was assumed that RTs for the reaction to unexpected events would be longer than for expected events and would regress back to a baseline level after repeated presentation of unexpected events. This pattern is typical for RT prolongation associated with unexpected events (see section 2.5.2).

Support for the hypothesis concerning a prolongation of RTs after a surprise was provided by an ANOVA. Both groups showed prolongation of RTs for the first unexpected event. A significant decrease of this effect from *event1* to *event2* and then to *event3* indicates the typical disappearance of the surprising characteristic of an unexpected event after several occurrences, confirming H1 and H.1.1. Based on these RTs analyses, it can be assumed that the unexpected event successfully elicited surprise in both groups. As can be seen in Figure 6.7, RTs are back to baseline level at the third event. Accordingly, subsequent presentation of the events would not be surprising for participants in group B. One interesting observation was the apparent difference in RT duration for the first event between the two groups. Although it did not prove to be significant, it could be an indicator for differently perceived subjective surprise. Participants in group A accomplished two more regular tasks after the break between task 15 and 16. Their

cognitive schemas of the application had more time and a bigger pool of information to draw on, in order to stabilize. Participants in group B on the other hand, experienced a new application feature (the animations) directly after the break between tasks 15 and 16. Schema discrepancy could have been smaller because they quickly deduced that they would be confronted with a new version of the application after the break.

6.4.2 User Experience Questionnaires

SAM ratings

If surprise had a beneficial effect on product ratings of digital products, this should have been observable in higher SAM valence ratings of group A after the experimental phase, compared to ratings given after the baseline phase (H.2.1). This assumption could not be confirmed. An unexpected neutral event did not have an effect on the subjective emotional state of the participants. Similarly, it was assumed that frequent unexpected events would first elicit surprise but would later fail to elicit this emotion due to the repeated presentation of the stimuli. Accordingly, it was expected that SAM valence ratings after the experimental phase would be worse than after the baseline phase in group B (H.2.2). Again, this hypothesis could not be confirmed. The high frequency of stimulus presentation to the experimental phase of group B did not significantly affect the subjective emotional state of the participants. In summary, when looking at each group separately, valence ratings were not influenced by the different frequencies of presentation of the event. Nevertheless, a significant difference in SAM valence ratings between the two groups after the experimental phase was found. It is important to keep in mind that the two groups did not give different valence ratings after the baseline phase. The effect of the different experimental manipulations only becomes apparent when looking at the two groups together: group B gives a significantly worse valence rating after having seen an event for ten times, when compared to group A. The slight improvement of valence ratings for group A, and the slight worsening of these ratings for group B manifest an important difference between the groups.

Furthermore, it was assumed that differences in subjectively perceived surprise would be observable in SAM arousal ratings between the two groups after the experimental phase (H.2.3). This hypothesis could not be confirmed, either.

meCUE ratings

It was assumed that the repeated presentation of the events in group B would no longer be perceived as surprising and would rather be annoying. As a result, participants in group A

were expected to rate the application higher on the positive emotions subscale (H.3.1) and lower on the negative emotions subscale (H.3.2), when compared to group B. These assumptions could only partly be confirmed. When comparing the differences between ratings of the application after the baseline phase to the experimental phase, only H.3.2 could be confirmed. Group B expressed a significantly larger negativity of their ratings when compared to group A. In a goal-directed context, the repeated presentation of unexpected stimuli has a negative effect on the subjective emotional status of participants. Furthermore, no other differences were found on the meCUE subscales, partly disconfirming hypotheses H.3.1, and H.3.2. Although unpleasant surprise did elevate product loyalty in a gaming context, this could not be observed in the goal-directed context. A possible explanation is the design of the unexpected event. Whereas valenced surprises were elicited in studies I and II, the design of the event in study III aimed to elicit neutral surprise. Although the investigation of surprise as a design strategy has been shown to have beneficial effects on product ratings (Ludden, 2008; Ludden et al., 2008; Ludden et al., 2009), it seems important to distinguish between its different valences.

Surprise item

It was hypothesised that the two groups would differ in their levels of subjectively experienced surprise (H.4.1). The repeated presentation of the events in group B was expected to lead to a schematization of the events which in turn would lead to a lower rating of subjective surprise levels. This could not be confirmed. Both groups indicated slightly elevated levels of surprise (group A=5.47; group B=5.17) on the seven-point surprise scale. Although slightly higher ratings were given in group A, this difference was not statistically significant.

6.4.3 OPEN QUESTIONS

Although data reported on the open questions consists of frequencies only, it provides a deeper insight into the attitudes participants had towards the application after the encounter of unexpected events. 11 participants in group A liked the application compared to 7 participants in group B. The same pattern could be observed when asking whether participants would continue using the application. However, although pre-testing suggested otherwise, the visual manipulations of the application were mostly rated negatively: 25 out of 33 participants did not like them. This could have had an effect on the overall UX ratings of the application. Although the unexpected events were intended to elicit neutral surprise, participants did not like them. Again, the valence of the unexpected events poses a problem. Although they were designed to be neutral, this effect could not

be guaranteed. Still, participants who did not encounter ten events did give more favourable ratings. It is worth mentioning that among the reasons given for liking or disliking the application, the animations were mentioned by four participants, only. Other aspects of the application like design, efficiency, ease-of-use, overall operation or input method seem to have had a more impact on users' ratings than the animations. The surprise manipulation did not affect users as strongly as hoped, so that resulting effects on the UX of the application were not as strong as anticipated.

6.4.4 SUMMARY AND CONTRIBUTION

The goal of the Study III was to investigate the effect of surprise and repeated surprise on the UX ratings of a digital application in a goal-oriented context. Although RT data supports the success of the intended surprise evocation, effects on UX ratings are small and difficult to interpret.

In summary, differences between groups after the experimental phase could only be observed in terms of SAM valence and meCUE negative emotions ratings. Participants who experienced the visual manipulations ten times reported feeling less happy than participants who encountered three manipulations, only. It is argued that for participants in group A, the visual surprises during interaction with the application had a beneficial effect on subjective emotional ratings. Furthermore, repeated presentation of surprising events had a negative effect on subjective emotional ratings, as has been shown by the meCUE negative emotions scale. This suggests a careful handling of the application of surprise as a design strategy in a goal-oriented context. If it is employed frequently, it is no longer experienced as something unexpected, surprising, and exciting, but rather as an unnecessary and unhelpful feature of the product. Especially in a goal-oriented context, this results in negative product ratings. Although the unexpected event in Study III was designed not to interfere with task completion, participants did not like its emergence.

6.4.5 CRITICAL EVALUATION

The results of this study are partly inconclusive. Designing unexpected events in goal-oriented applications does not seem to have a very distinct effect on users. Although it has been found that surprise has a positive effect on users in classical product design and in the domain of digital games, this effect could only partly be confirmed in a digital goal-oriented application used in this study. Some critical issues emerged during the execution of Study III.

First, it proved to be a lot more difficult to design the unexpected event compared to the first two studies. It was decided to design a neutral surprise instead of a two valenced surprise for various reasons: in a goal-directed application like the one used at hand, it is hard to create surprising events that do not interfere with the task-flow, do not leave room for interpretation of its valence, and are not at the end of the interactions, to allow for RT measurements. The decision to opt for a neutral surprise brought with it new complications. Although a pre-test seemed to confirm the neutral nature of the animations, participants in the study mostly did not like them. Accordingly, the neutrality of the surprise could not be guaranteed.

Second, the emergence of surprise during the execution of the task might have been problematic. A more beneficial result of surprising product design could be possible if the surprising event did not emerge during the execution of a task, but rather after a task is done, e.g. to bridge loading times or to celebrate a successful task completion. Although the unexpected event did not influence performance of the application, it could have elicited the subjective impression of doing so, due to the temporal placement of the event. The timing of the events in study III could not have been different due to the necessity of recording RT data. Nevertheless, a different occurrence, e.g. at the end of a completed task could have produced different results.

7 GENERAL DISCUSSION

Efficient and effective goal accomplishment has long been the single criterion to evaluate the design of an interactive system. As interactive products are used in a growing variety of contexts, the importance of other criteria -beyond usability- has gained momentum. UX is a holistic concept that takes all aspects of interaction into account (Jordan, 1997). An important focus lies on user emotions (Hassenzahl, 2005; Mahlke & Thüring, 2007). Emotional product design stresses the importance of the emotions a user experiences when interacting with a product. Most researchers advocate the elicitation of pleasant emotions in order to elevate the overall UX (Monk et al., 2002; Jordan, 1997; Norman 2002), or even to make up for potential shortcomings of a product (Mun & Hwang, 2003). It seems as if emotions are indeed a rewarding design goal, and their deliberate application as a design feature is advocated. It is common to focus on the design of pleasant or joyful emotional experience. The underlying idea is that the user should experience a positive emotion when interacting with a product, which in turn leads to a positive attitude towards the whole experience and the product at hand. According to Desmet (2012), "...products that evoke positive emotions are bought more often, used more often, and are more pleasurable to use" (p. 1).

New approaches to emotional product design have been proposed over the past three years that promote the inclusion of unpleasant emotions in order to create pleasant experiences. Surprise is especially suitable to be used by designers because of its possible bipolar characteristic. Although it is already used in product design, serves in brand claims to promote favourable impressions, and is implemented in digital products, there are severe shortcomings in the systematic investigation of surprise. The lack of empirical insights into the underlying effects surprise can have on product users is problematic. Although it is mostly seen as a positive emotion, surprise can be experienced as severely negative in situations in which the concerns and expectations of users are not met (Desmet & Hekkert, 2007). Product designers have to be aware of this possibility and the possible interrelations between users' expectations, their concerns and the emotional design of a product.

Three major issues are addressed to close the knowledge gap about surprise as a design strategy in interactive digital products and to give product designers first hints how to use surprise correctly in order to benefit from its potential.

First, the possibility of a deliberate evocation of surprise was examined. Second, different forms of surprise were investigated in order to shed light on the effects of pleasant, unpleasant and neutral surprise on UX. Third, the role of frequency of presentation when working with surprise as a design strategy was investigated.

7.1 DIGITAL PRODUCTS NEED VARIOUS KINDS OF SURPRISES.

The main prerequisite for all three studies was the successful evocation of surprise. For the first two studies, the generation of pleasant, unpleasant, and neutral surprises was accomplished with a computer game through the manipulation of game scores and by pop-up messages, respectively. Based on Reisenzein's BDTE (2009), the valence of surprise was determined according to user's personal expectations and wishes. Success of surprise manipulations was objectively measured through RTs. As described in the cognitive psycho-evolutionary model of surprise, a surprise reaction is characterized by longer RTs due to the interruption of all ongoing cognitive processes. This prolongation quickly vanishes for subsequent similar events due to the schematization of the unexpected information (Meyer, 1991, 1997). Unpleasant surprises were successfully created in Study I by manipulating the game score of a digital game, subtracting a given number of points. This manipulation seemed random to the users because it did not relate to their performance. A corresponding positive surprise manipulation was not successful. Different mental processes when reacting to gains and losses can be the reason for this. As described by Kahneman and Tversky (1979), losses are perceived as being more severe than gains. Accordingly, the negativity of 50 malus points was perceived as much more severe than the benefit of 50 bonus points. Therefore, a different surprise manipulation was chosen for the second study with no numerical gains or losses. Unexpected events that either directly motivated or discouraged the participants both resulted in significantly prolonged RTs, as did neutral unexpected events.

In a third study, similar results were obtained in a more serious, task-directed environment. A neutrally surprising event was designed by manipulating visual characteristics of parts of a mobile application. Prolonged RTs and a subsequent

normalization of the RTs back to a baseline level indicate the successful surprise manipulation in the task-directed environment, as well. Nevertheless, the validation of a successful surprise evocation in Study I and Study II could only be carried out in terms of RT data analysis because other measures of surprise did not show significant differences between groups. EMG data proved unusable in study I and inconclusive in study II. Questionnaire data about surprise did not reveal any differences between groups.

7.2 INVESTIGATING THE EFFECTS OF PLEASANT, UNPLEASANT, AND NEUTRAL SURPRISE ON THE UX OF DIGITAL INTERACTIVE PRODUCTS.

In accordance with reported findings on the effects of surprise on evaluations of decorative, non-interactive products, beneficial effects of pleasant surprise on product evaluations were expected. As predicted, pleasant surprises disguised as unexpected motivating messages had a positive effect on the UX ratings of a game. Likewise, unpleasant surprises had a negative effect on the UX ratings of the game. The interpretation of these results could have been straightforward; if it had not been for the fact that product loyalty ratings were higher for a game with unpleasant surprises when compared to one with pleasant surprises. This finding illustrates the complex nature of emotional design, especially surprising emotional design. It also stresses the possibility for designers to employ negative emotions in order to ultimately create pleasant and rich experiences, as described by Fokkinga and Desmet (2012, 2013), and proposed by Silvia (2009). Of course, games are a special kind of digital interactive product. Negative surprise in the form of a challenge contributes to the overall appreciation of a game, since it constitutes one of its enjoyable characteristics (Hsu & Lu, 2004; Sweetser & Wyeth, 2005). Therefore, the findings from the first two studies need to be expanded and investigated in other contexts of use.

7.3 INVESTIGATE THE INFLUENCE OF FREQUENCY OF PRESENTATION OF A SURPRISING EVENT ON UX OF A DIGITAL INTERACTIVE PRODUCT.

Although some difficulties were encountered during the design of surprise for a digital public transport application, the results suggest a careful application of surprise as a

design strategy in such a context. The problem with surprise is its singularity: as has been shown, users can only be surprised one to two times. Because of the underlying processes (see section 2.5.1), surprise is often described as a once only product emotion (Desmet, 2003). Once a user has become familiar with an unexpected aspect of a product, it will no longer elicit surprise. Findings from the third study suggest that the process of familiarization does take more than one encounter with a surprising aspect of a product. RT data suggest that upon the third encounter, the schematization process is completed and a formerly surprising product feature is no longer appraised as such. Subsequently, the product feature no longer transports possible beneficial effects to the overall UX evaluation of the product. When the feature is rated as being annoying or unnecessary, as in study III, it will lower UX ratings. An appropriate level of novelty is needed in the case of surprise (Hassenzahl et al., 2001). A rough guideline can be deduced from the findings of the last study: in goal-oriented contexts a similar surprise should not be used more than three times. Study III provided first insights into the repeated presentation of unexpected events. However, this repetition of events was within a very short period of time. Investigating these effects of repetitive unexpected events at longer intervals could offer more insights into how surprise influences UX in the long-run: a longitudinal study would be helpful to determine the long-term effects of surprise on UX and user emotions.

7.4 RETURNING TO THE THEORETICAL BASIS

7.4.1 DIFFICULTIES WHEN DESIGNING FOR SURPRISE

It has been shown that surprise is not easy to design for. It is after all a complex emotion. Study I revealed that pleasant surprises in the form of monetary or score manipulations run the risk of not being as strong as their corresponding unpleasant counterparts. In Study III it was demonstrated that the intended valence of a surprise cannot be guaranteed, despite careful design planning. How users appraise an unexpected event is not entirely predictable: given the subjective nature of emotions, all unexpected events that can elicit an emotion are individually interpreted. The result can be fundamentally different emotions. In the case of surprise, this interpretation can go four ways. Either the interpretation of an unexpected event is as intended, meaning that users can enjoy the experience planned by the designers, or it can result in no surprise at all, when unexpectedness is not appraised. Finally, surprise can also take one of the two remaining emotional values that were not intended by the designer: unpleasant surprise instead of a

pleasant one, or vice versa. According to Hassenzahl's model of UX, it is very difficult to design a certain kind of UX, because of inter-individual variances: users react differently to the same product (Hassenzahl, 2005). In order to deliberately use a surprise as a design strategy to generate pleasurable experience, one needs to keep in mind Hassenzahl's distinction between the intended and the apparent product character. The designer's intention is not necessarily congruent with what the user experiences when interacting with a product. In the case of surprise, this is especially important because depending on users' predispositions, personality, or emotional concerns (Desmet et al., 2001), a surprise can lead to emotional reactions that are not desired and possibly even harmful to UX. As has been shown in study III, the neutrally designed event was not perceived as such. Most participants did not like the looks of it and experienced it as being unnecessary. Accordingly, creating a product that is intended to yield pleasant or interesting experience but is instead annoying, frustrating, or unusable to the user has to be avoided. Following this argumentation, a designer can neither design the user nor the use situation, but can rather design *for* UX, with a certain idea in mind. Similarly, Desmet et al. (2001) highlight the diversity of potential users and their different emotional concerns. One aspect that was not considered in the planning of unexpected events and their subsequent consequences were these personal predispositions. Different personality traits can influence the reactions to emotional events in different ways (Desmet & Hekkert, 2007). Hassenzahl and Tractinsky (2006) mention personal predispositions, needs, expectations and motivations as relevant factors for user-product relations. One user is not like another. Accordingly, in the case of surprise, personal predispositions play an important role in whether users appreciate unexpectedness or dislike it.

Ludden stresses the importance of the type of product under evaluation that plays a part in influencing people's appreciation of surprising products (Ludden et al., 2006). For products with a complicated functionality that take up many cognitive resources and thus require full attention from the user, surprise might not be appreciated: it diverts cognitive resources away from the original task. Similarly, Desmet et al. (2001) argue that in order to design for a certain emotion, the context of use needs to be known to the designer. They try to tackle this problem by including the users in the design process.

Furthermore, close attention has to be paid to users, their expectations and concerns, and the context of use. The CUE model by Mahlke and Thüring (2007) illustrates that UX not only depends on the product itself, but also on special characteristics of the user, as well as the situation and thus the context in which the product is used. The way a product is evaluated clearly depends on the personal contexts of use when interacting with certain products (Bargas-Avila & Hornbæk, 2001; Gross & Bongartz, 2013; Hassenzahl et al., 2005;

Law et al., 2009). Especially in goal-oriented contexts, people have precise ideas about what they want to achieve during the interaction with the product. The implementation of surprise in these contexts is especially intricate because it is necessary to avoid interfering with these goals. Practically speaking, an unexpected event that resulted in a pleasant surprise in a gaming context can result in an unpleasant surprise in another context.

Timing is another important factor that needs to be considered when designing surprising products. As shown in the three studies, and in accordance with Meyer at al. (1991, 1997), RTs are longer when something unexpected is encountered. The more relevant the unexpected event is to the actual task at hand, the longer the RT prolongations. Therefore, surprise always equals task interruption. When looking at usability standards, task interruptions should be avoided in order to maintain high levels of usability (DIN EN ISO 9241-210, 2010). In the case of Tetris, unexpected events did not interfere with the task since they did not stop participants from playing the game. Despite the unexpected events, participants were able to go on with the task. During interaction with the mobile application, tasks were subjectively interrupted by the presentation of unexpected events. Although the presence of these events did not slow down the application, they occurred during the execution of a goal-directed task and could have elicited the impression that performance of the application was slowed down. Due to the necessity of recording RT data, the events in Study III could not have been differently timed. Nevertheless, another occurrence, e.g. at the end of a completed task, may have produced other results. Machlovi (2015) describes current successful surprising product design. All of the products mentioned in the blog entry share one characteristic: surprises occurred at a point in time during interaction when users were not actively working on a task, e.g. after completion or during loading times.

7.4.2 EMOTIONAL DESIGN APPROACHES

Several emotional design approaches refer to the importance of well-thought-out emotional design. Norman (2002) for example states that there are three levels of product experience, and that accordingly, there should be three levels of design: the visceral level, which is the most basic one, the behavioural level, which is all about the use of the product, and the reflective level. The reflective level includes self-image, personal satisfaction, and memories. It determines the lasting value of a product to the owner, by influencing memories, self-image, and personal satisfaction. If implemented correctly, surprise can be a design tool that positively reinforces these aspects, making the long-term success of a product more probable. Similarly, surprise can serve as an ideo-pleasure,

according to Jordan's approach (1997). Those pleasures are affiliated with the value a product signifies for a person, which could be positively influenced by surprise. When considering surprise in terms of the model by Desmet and Hekkert (2002), the interplay between users' appraisal and concerns determine the success of surprise as a design strategy for a product. If the surprising features of a product match the concerns of the user, it will be appraised as being beneficial, and vice versa. Users' concerns determine their personal well-being, and thus the success of a product (Desmet & Hekkert, 2007). Accordingly, designers need to carefully consider these concerns and to know their users, as well as the intended context of use. Likewise, if a surprising product feature is too advanced, too novel, or too unexpected it could unsettle users and lower UX ratings. Designers should follow the Most Advanced Yet Acceptable principle (MAYA), finding a balance between novelty and typicality of the unexpected events (Hekkert, Snelders & Wieringen, 2003).

In summary, surprise holds the potential to be a valuable tool for designers to increase the emotional value of a product. Nevertheless, some product features can influence the effectiveness of surprise as a design feature and should be carefully kept in mind.

7.4.3 MOOD AND EMOTION

Strong emphasis must be put on the different characteristics of emotions and moods and their consequential effects on product evaluations when discussing surprise as a design feature. Emotions usually last shorter than moods, have a distinct antecedent event and are accompanied by distinct facial expressions. In contrast, moods are longer lasting affective states, the result of an emotional event of which the user has lost focus. Additionally, they are not necessarily accompanied by distinct facial expressions (Ekman & Davidson, 1994). When investigating surprise as design feature, the present studies aimed at measuring the effect of a successful surprise manipulation by measuring product emotions. As has been discussed above, it was not always possible to establish a relationship between surprise and subjectively relevant emotional states of the users. It might be that instead of influencing product emotions, the implemented surprising events did ultimately influence users' moods. Because users' emotions towards the products were assessed after interaction with the products, the surprising pleasant and unpleasant events could have already turned into a mood, the after-effect of these events (Frijda, 2009). Moods do not necessarily register in terms of facial expressions and were not the focus of the questionnaires presented to participants. They are less easily observable and measurable than emotions. It seems as if the design of emotional experiences always relates emotions with moods and vice versa. Effects of moods and emotions on product

evaluations are similar, whereas the respective assessments differ (Spillers, 2010). Accordingly, the inconclusive results of the studies could be explained in terms of the transition of product emotions to product moods.

Additionally, surprise can be the antecedent for other emotions like joy, fear, anger, amusement, fascination, disappointment, or irritation (Ekman, 2007; Ludden, Hekkert & Schifferstein, 2006; Ludden, Schifferstein, Hekkert, 2009; Scherer, 1987). The three studies assessed the influence of surprise on users' emotions in a different way. Instead of assessing levels of different emotions after the interaction with surprising products, the overall change in emotional valence due to the surprise manipulations was assessed. In order to gain a more detailed view on the after-effects of surprise on product evaluations, an in-depth assessment of possible emotions following surprise can help to clarify the emotional impact surprising products have on users.

7.5 CRITICAL EVALUATION OF THE EMPIRICAL FINDINGS

7.5.1 DESIGN OF SURPRISE

Designing a surprising product feature that has a similar impact on all participants proved to be difficult. Experiencing is something subjective and personal, and it is impossible to anticipate how an experience will feel to another person. Sometimes, theoretical approaches to emotional product design explicitly describe these differences between intended and apparent product characteristics (Hassenzahl, 2005). In addition, factors like users' predispositions, personality, or emotional concerns reinforce these differences.

Studies I, II, and III employed different unexpected events in the hope of creating surprise in users and consequently making a point about the effects of surprising product design. Several difficulties were encountered when designing these unexpected events. Either the manipulations were not strong enough to validly elicit surprise in the users (study I), the context of use had influence on the intended valence of the events (study II), or it was not fully possible to guarantee that participants experienced an unexpected event as neutral (study III). In order to make a valid point about the effects of these unexpected events on UX ratings, surprise reactions to the events should have been congruent with the planned intention. Because of difficulties when designing the events that have been discussed in the preceding sections, it was difficult to interpret the results.

7.5.2 CONTEXT

Different contexts can create different concerns in the users and should therefore be carefully considered by the designers (Desmet, 2003). Although two different contexts of use have been investigated, it is difficult to transfer findings from one context to another. Whereas the valence of surprise was the focus of the studies in the gaming context, the frequency of surprise was the focus of the study in the goal-directed context. When pleasantly surprising events have a beneficial effect on UX ratings in a gaming context, the same events can have a deleterious effect in a goal-oriented context. Similarly, the frequency of neutral events did not prove to have a strong effect on UX ratings of a goal-oriented application. In a different context of use, this could present a different picture. Additionally, the personal performance was the focus of the first two studies and had a direct influence on the outcome of the task. This involvement of performance and the lack of such in the goal-directed context (no monetary reward for good performance) could have established one important difference between the two contexts. Because the experimental manipulations were not similar in all studies, questions like this remain unanswered. In conclusion, it can be stated that the context of use always has an impact on users and therefore needs to be included in the systematic investigation of surprise as a design feature.

7.5.3 INTERPERSONAL DIFFERENCES

Different personality traits, emotional concerns, memories, or experiences result in differently experienced interactions with products (Desmet, Overbeeke & Tax, 2001; Korhonen, Montola & Arrasvuori, 2009). While all participants in all three studies received identical instructions about their interaction with the products, these factors were not treated as independent variables. When designing experiences, there is a subjective dimension rooted in the user that complicates predictions of UX and product evaluations. Including these factors in the description of the participants and including them in the later data analysis could have been helpful in interpreting the results.

7.5.4 METHOD OF MEASUREMENT

To access the minds of participants and their emotional experiences, subjective as well as objective measures were employed, such as questionnaires, RTs and EMG. EMG data proved to yield no valuable insights into the emotional states of the participants. This could have been due to behaviour explained in terms of the insufficient sociality hypothesis (Reisenzein et al., 2006; see section 5.4.2). In Study III, questionnaire data

yielded minimal results, due to lack of diverse response patterns. Lorini and Castelfranchi (2007) describe similar observations when investigating subjective levels of surprise in participants. They conclude that being surprised is not the same as being aware of this surprise. Questionnaires and EMG data have been proven to be useful methods of measurement of surprise in earlier works and were therefore used in the studies, as well. Nevertheless, the employment of additional methods of measurements like electroencephalography (EEG), or eye movement could help to complete the picture when working with surprise.

8 OUTLOOK AND CONCLUSION

The work has been motivated by the notion that surprise as a product feature has a beneficial effect on users: they talk more about the product, they find it more interesting, they have better remembrance of it, and generally like it more. Over the course of three studies, the attempt was made to transfer these findings from the domain of classical product design of decorative, non-interactive objects to the domain of digital, interactive products. Many questions still remain unanswered and much work is still to be done. The influence of factors like context of use or interpersonal differences between users need to be examined in order to extend and refine the findings from the three studies at hand. Further work needs to be devoted to the systematic assessment of these factors and their influence on the effects of surprising product design. This would help to gain a deeper understanding of the dynamics of the effects of surprising products: UX evaluations of interactive products might vary as a function of context, personality, concerns, frequency, and other factors. Additionally, long-term effects of surprising product characteristics were not the focus of this work but are nevertheless of paramount importance to designers. How do the effects of surprise as a product feature vary in the course of time? Is it possible to use the same unexpected event several times over a long period and still generate surprises that are pleasant for users? Do users like to be surprised from time to time or do these features lose their appeal overt the long-term? Only longitudinal studies will be able to answer these questions and provide designers with guidelines on whether and how to implement surprise in their digital products to create satisfied users.

During development of the experimental studies, another aspect of surprise has emerged. Surprise could be a perfect means to measure the original expectations users have about a certain product and how these expectations are fulfilled or disappointed during the interaction. This could unveil possible shortcomings in the product design. Behavioural responses to unexpected product features that the designer did not intend to be surprising are a direct indicator of aspects of a product that do not conform to users' expectations. Surprise can become a dependent variable and could be used more often in experiments in order to better understand the user's needs and expectations. Surprise is only possible when something violates beliefs and expectations and can thus be a direct indicator of

product features that are pleasant to users, and those that are perceived as being unpleasant, unnecessary, or plainly wrong. In the future, designers and researchers should look out for surprise reactions in their users to learn about factors influencing the UX.

Finally, too much novelty can be harmful. When designing for surprising UX, designers should not go overboard in their attempts to create pleasant experiences. Following the MAYA principle when creating unexpected events is a good way to stay in a safe area of design within which users will not be shocked by unexpected events. By doing so, surprise can be established as a unique selling point to set apart one product and its brand from the homogenous mass of competing brands.

In the three studies, different aspects of surprise were systematically varied and their effects on UX ratings of digital interactive products were investigated. Valence, as well as frequency of surprise were investigated in two different contexts of use. In conclusion, surprise is a promising design tool in order to enhance product evaluations and product success. It is already widely used and appreciated by designers for its ability to create rich and enjoyable experiences. Nevertheless, many factors complicate the use of surprise as a design feature and must be carefully studied. Designers need to know about the intricate interplay of these factors and should try to gain insights into the minds of their prospective users before and during the design process. Some researchers already propose design approaches that include the prospective users from an earliest design stage on. In the case of surprise this can help to make sure that surprise is experienced as intended; either as pleasant, or if suitable even as unpleasant. A distinction between participants with different personality traits can further distinguish the effects of surprise on product ratings and should be implemented in future studies.

Surprise is a challenging emotion to work with, but if implemented carefully and with consideration of the users' concerns, it can be a great way to enhance interaction and leave users with exciting and rememberable experiences.

9 REFERENCES[14]

Alben, L. (1996). Quality of experience: defining the criteria for effective interaction design. *Interactions*, *3*(3), 11–15.

Arnold, M. (1945). Physiological differentiation of emotional states. Psychological Review, 35-48.

Bargas-Avila, J. A., & Hornbæk, K. (2011). Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2689–2698). New York, NY, USA: ACM.

Beedie, C., Terry, P., & Lane, A. (2005). Distinctions between emotion and mood. *Cognition and Emotion*, *19*(6), 847–878.

Bias, R. G., & Mayhew, D. J. (2005). *Cost-justifying usability: an update for an Internet age*. Morgan Kaufmann.

Boxtel, A. van (2010) Facial EMG as a Tool for Inferring Affective States. In *Proc. Measuring Behaviour*, Eindhoven, Netherlands.

Bradley, M. M., and Lang, P. J. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry 25,* 1 (1994), 49-59.

Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M., & Ito, T. A. (2010). The Psychophysiology of Emotion. In M. Lewis, J. M. Haviland-Jones, & L. Feldman Barrett (Eds.), *Handbook of Emotions* (pp. 180–195). New York, NY, USA.

Carroll, J. M., & Thomas, J. C. (1988). FUN. *SIGCHI Bull.*, *19*(3), 21–24.

Cornelius, R. R. (2000). Theoretical approaches to emotion. In *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*.

Crilly, N., Moultrie, J., & Clarkson, P. J. (2004). Seeing things: consumer response to the visual domain in product design. *Design Studies*, *25*(6), 547–577.

[14] Internet sources were accessed in the period January – March 2015.

Davidson, R. J. (1994). On Emotion, Mood, and Related Affective Constructs In P. Ekman & R. J. Davidson (Eds.) *The nature of emotion: Fundamental questions.* (pp. 51-55). Oxford University Press.

Davis, F. D. (1993). User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *International Journal of Man-Machine Studies*, *38*(3), 475–487.

Demir, E. (2008). The field of design and emotion: Concepts, arguments, tools, and current issues. *Journal of the Faculty of Architecture*, *25*(1), 135–152.

Derbaix, C., & Vanhamme, J. (2003). Inducing word-of-mouth by eliciting surprise – a pilot investigation. *Journal of Economic Psychology*, *24*(1), 99–116.

Desmet, P. (2002). *Designing Emotion*. Technische Unisversiteit Delft.

Desmet, P. (2003). A multilayered model of product emotions. *The Design Journal*, *6*(2), 4–13.

Desmet, P. M. A. (2012). Faces of product pleasure: 25 Positive emotions in human-product interactions. *International Journal of Design*, *6*(2), 1–29.

Desmet, P. M. A., & Hekkert, P. (2002). The basis of product emotions. *Pleasure with Products, beyond Usability*, 61–68.

Desmet, P., Overbeeke, K., & Tax, S. (2001). Designing products with added emotional value: Development and application of an approach for research through design. *The Design Journal*, *4*(1), 32–47.

Desmet, P. M. A., Porcelijn, R., & Van Dijk, M. B. (2007). Emotional design: application of a research-based design approach. *Knowledge, Technology & Policy*, *20*(3), 141–155.

Desmet, P. M., & Schifferstein, H. N. (2008). Sources of positive and negative emotions in food experience. *Appetite*, *50*(2), 290–301.

DIN EN ISO 9241-210. (2010). Ergonomie der Mensch-System-Interaktion - Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme. Berlin: Beuth.

Ekkekakis, P. (2012). Affect, mood, and emotion. *Measurement in Sport and Exercise Psychology*, 321–332.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*(3-4), 169–200.

Ekman, P. (1993). Facial expression and emotion. *American Psychologist*, *48*(4), 376–379.

Ekman, P. (2007). *Emotions Revealed, Second Edition: Recognizing Faces and Feelings to Improve Communication and Emotional Life*. New York: Holt Paperbacks.

Ekman, P. E., & Davidson, R. J. (1994). *The nature of emotion: Fundamental questions.* Oxford University Press.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face of emotion. *Journal of Personality and Social Psychology*, *17*(2), 124–129.

Ekman, P., & Friesen, W. V. (2003). *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues*. ISHK.

Filipowicz, A. (2006). From positive affect to creativity: The surprising role of surprise. *Creativity Research Journal*, *18*(2), 141–152.

Fokkinga, S., & Desmet, P. (2012). Darker shades of joy: The role of negative emotion in rich product experiences. *Design Issues*, *28*(4), 42–56.

Fokkinga, S. F., & Desmet, P. (2013). Ten ways to design for disgust, sadness, and other enjoyments: A design approach to enrich product experiences with negative emotions. *International Journal of Design*, *7*(1), 19–36.

Frijda, N. H. (1986). *The emotions*. New York: Cambridge University Press.

Frijda, N. H. (2009). Mood. In: D. Sander & K. Scherer (2009). *Oxford Companion to Emotion and the Affective Sciences* (pp. 258-259). Oxford ; New York: Oxford University Press.

Frijda, N. H. (2010). The Psychologist's Point of View. In: M. D. Lewis, J. M. Haviland-Jones & L. F. Barrett. *Handbook of emotions* (pp.68-87). Guilford Press.

Gaver, W. W., Bowers, J., Boucher, A., Gellerson, H., Pennington, S., Schmidt, A., Steed, A., Villars, N., & Walker, B. (2004). The Drift table: Designing for ludic engagement. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems* (pp. 885–900). New York, NY, USA: ACM

Geers, A. L. & Wellman, J. A. (2009). Expectation. In: D. Sander & K. Scherer (2009). *Oxford Companion to Emotion and the Affective Sciences* (pp. 164-165). Oxford ; New York: Oxford University Press

Gross, A., & Bongartz, S. (2012). Why do I like it?: investigating the product-specificity of user experience. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design* (pp. 322–330). ACM.

Gross, A, & Bürglen, J. (in press). The Impact of Expectations on User Experience: Surprising the User. In *Proceedings of HCI 2014: Sand, Sea & Sky: The 28th BCS Conference on Human-Computer Interaction.* Southport, UK, September 9-12, 2014. BCS.

Gross, A. & Silvennoinen, J. (2014). Surprise as a Design Strategy in Goal-Oriented Mobile Applications. In *F. Rebelo & M. Soares (Eds.), Proceedings of the 5th AHFE Conference 19-23 July 2014. Advances in Ergonomics in Design, Usability & Special Populations. Part II.* (pp. 4716-4726).

Gross, A., & Thüring, M. (2013). Encountering the Unexpected: Influencing User Experience through Surprise. In *Proc. of UMAP Annual Conference,* Rome, Italy: Springer, 3-9.

Hancock, P. A., Pepe, A. A., & Murphy, L. L. (2005). Hedonomics: The Power of Positive and Pleasurable Ergonomics. *Ergonomics in Design*, *13*(1), 8–14.

Hassenzahl, M. (2001). The effect of perceived hedonic quality on product appealingness. *International Journal of Human-Computer Interaction*, *13*(4), 481–499.

Hassenzahl, M. (2005). The Thing and I: Understanding the Relationship Between User and Product. In M. A. Blythe, K. Overbeeke, A. F. Monk, & P. C. Wright (Eds.), *Funology* (pp. 31–42). Springer Netherlands.

Hassenzahl, M. (2008). User experience (UX): towards an experiential perspective on product quality. In *Proceedings of the 20th International Conference of the Association Francophone d'Interaction Homme-Machine* (pp. 11–15).

Hassenzahl, M., Beu, A., & Burmester, M. (2001). Engineering joy. *IEEE Software*, *18*(1), 70–76.

Hassenzahl, M., Burmester, D. M., & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In G. Szwillus & J. Ziegler (Eds.), *Mensch & Computer 2003* (pp. 187–196). Vieweg+Teubner Verlag.

Hassenzahl, M. & Monk, A. (2010). The inference of perceived usability from beauty. *Human-Computer Interaction*, 25(3), 235-260.

Hassenzahl, M., & Tractinsky, N. (2006). User experience - a research agenda. *Bahaviour & Information Technology*, *25*(2), 91–97.

Hekkert, P., Snelders, D., & Wieringen, P. C. (2003). "Most advanced, yet acceptable": typicality and novelty as joint predictors of aesthetic preference in industrial design. *British Journal of Psychology*, *94*(1), 111–124.

Helander, M. G., & Po Tham, M. (2003). Hedonomics—affective human factors design. *Ergonomics*, *46*(13-14), 1269–1272.

Hsu, C.-L., & Lu, H.-P. (2004). Why do people play on-line games? An extended TAM with social influences and flow experience. *Information & Management*, *41*(7), 853–868.

International Organization for Standardization (1998). *ISO 9241 - Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability*.

Izard, C. E. (1977). *Human emotions* (Vol. 17). Plenum Press New York.

Izard, C. E. (1992). Basic emotions, relations, among emotions, and emotion-cognition relations. *Psychological Review*, *99*(3), 561–565.

James, W. (1884). What is an emotion? *Mind*, (34), 188–205.

Jordan, P. W. (1997). Putting the pleasure into products. *IEE Review*, *43*(6), 249–252.

Jordan, P. W. (2003). *Designing Pleasurable Products: An Introduction to the New Human Factors*. CRC Press.

Kagan, J. (2009). Temperament. In: D. Sander & K. Scherer (2009). *Oxford Companion to Emotion and the Affective Sciences* (pp. 389-390). Oxford ; New York: Oxford University Press.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, 263–291.

Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. *Cognition*, *11*(2), 143–157.

Kapferer, J.-N. (2008). *The New Strategic Brand Management: Creating and Sustaining Brand Equity Long Term*. Kogan Page Publishers.

Karapanos, E., Hassenzahl, M., & Martens, J.-B. (2008). User experience over time. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems* (pp. 3561–3566). New York, NY, USA: ACM.

Kim, J., & Moon, J. Y. (1998). Designing towards emotional usability in customer interfaces—trustworthiness of cyber-banking system interfaces. *Interacting with Computers*, *10*(1), 1–29.

Klein, J., Moon, Y., & Picard, R. W. (2002). This computer responds to user frustration:: Theory, design, and results. *Interacting with Computers*, *14*(2), 119–140.

Kleinginna, P. R. J., & Kleinginna, A. M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion*, *5*(4), 345–379.

Korhonen, H., Montola, M., & Arrasvuori, J. (2009). Understanding playful user experience through digital games. In *International Conference on Designing Pleasurable Products and Interfaces* (pp. 274–285).

Lallemand, C., Gronier, G., & Koenig, V. (2014). User experience: A concept without consensus? Exploring practitioners' perspectives through an international survey. *Computers in Human Behavior*, *43*, 35–48.

Larsen, J. T., Norris, C. J., & Cacioppo, J. T. (2003). Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology*, *40*(5), 776–785.

Law, E. L.-C., Roto, V., Hassenzahl, M., Vermeeren, A. P. O. S., & Kort, J. (2009). Understanding, scoping and defining user experience: a survey approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 719–728). New York, NY, USA: ACM.

Law, E. L.-C., Vermeeren, A. P., Hassenzahl, M., & Blythe, M. (2007). Towards a UX manifesto. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as we know it. 2,* 205–206.

Lazarus, R. S. (1994). The stable and the unstable in emotion. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion* (pp. 79-85). Oxford, UK: Oxford University Press.

Lazarus, R. S., & Smith, C. A. (1988). Knowledge and appraisal in the cognition-emotion relationship. *Emotion & Cognition*, *2*(4), 281–300.

Leventhal, H., & Tomarken, A. J. (1986). Emotion: Today's Problems. *Annual Review of Psychology*, *37*(1), 565–610.

Lindstedt, J. K., & Gray, W. D. (2013). Extreme Expertise: Exploring Expert Behavior in Tetris. In *Cooperative Minds: Social Interaction and Group Dynamics* 912–917.

Lidwell, W., Holden, K., & Butler, J. (2003). *Universal Principles of Design*. Beverly, MA: Rockport Publishers.

Logan, R. J. (1994). Behavioral and Emotional Usability: Thomson Consumer Electronics. In M.E. Wicklung (Ed), *Usability in Practice: How companies develop user friendly products* (pp. 59-82). Boston: Academic Press.

Logan, R. J., Augaitis, S., & Renk, T. (1994). Design of simplified television remote controls: A case for behavioral and emotional usability. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *38*(5), 365–369.

Lorini, E., & Castelfranchi, C. (2007). The cognitive structure of surprise: looking for basic principles. *Topoi*, *26*, 133–149.

Ludden, G. D. S. (2008). *Sensory incongruity and surprise in product design*. TU Delft, Delft University of Technology.

Ludden, G. D. S., & Schifferstein, H. N. J. (2007). Effects of visual-auditory incongruity on product expression and surprise. *International Journal of Design*, *1*(3), 29–39.

Ludden, G. D. S., & Schifferstein, H. N. J. (2009). Should Mary smell like biscuit? Investigating scents in product design. *International Journal of Design*, *3*(3), 1–12.

Ludden, G. D. S., Hekkert, P., & Schifferstein, H. N. J. (2006). Surprise & emotion. In P. Desmet, M. Karlsson, and J. van Erp (Eds), *Proceedings of the 5th Conference on Design and Emotion Society*. Gothenburg, Sweden: Chalmers University of Technology.

Ludden, G. D. S., Schifferstein, H. N. J., & Hekkert, P. (2008). Surprise as a design strategy. *Design Issues*, *24*(2), 28–38.

Ludden, G. D. S., Schifferstein, H. N. J., & Hekkert, P. (2009). Visual-tactual incongruities in products as sources of surprise. *Empirical Studies of the Arts*, *27*(1), 61–87.

Ludden, G. D. S., Kudrowitz, B. M., Schifferstein, H. N. J., & Hekkert, P. (2012a). Surprise and humor in product design. *Humor*, *25*(3), 285–309.

Ludden, G. D. S., Schifferstein, H. N. J., & Hekkert, P. (2012b). Beyond surprise: A longitudinal study on the experience of visual-tactual incongruities in products. *International Journal of Design*, *6*(1), 1–10.

Macedo, L., Cardoso, A., Reisenzein, R., Lorini, L., & Castelfranchi, C. (2009). Artificial surprise. In: J. Vallverdú, & D. Casacuberta, (Eds.*) Handbook of research on synthetic emotions and sociable robotics: New applications in affective computing and artificial intelligence* (pp. 267-291). Hershey, PA: IGI Global.

Machlovi, H. (2015, January 4). Increasing User Engagement with Rewarding Interactions. [Blog Post]. Retrieved from http://uxmag.com/articles/increasing-user-engagement-with-rewarding-interactions

Mahlke, S., & Thüring, M. (2007). Studying antecedents of emotional experiences in interactive contexts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 915–918). New York, NY, USA: ACM.

Malone, T. W. (1981). Heuristics for designing enjoyable user interfaces: Lessons from computer games. In *Proceedings of the 1982 Conference on Human Factors in Computing Systems* (pp. 63–68). New York: ACM.

Mano, H. (1996). Assessing emotional reactions to TV aids: a replication and extension with a brief adjective checklist. *Advances in consumer research, 23*, 63-69.

Maslow, A. H., Frager, R., Fadiman, J., McReynolds, C., & Cox, R. (1970). *Motivation and personality* (Vol. 2). Harper & Row New York.

Matsumoto, D., & Ekman, P. (2009). Basic Emotions. In D. Sander & K.R. Scherer (Eds.) *Oxford Companion to Emotion and the Affective Sciences*. (pp.69-72). Oxford ; New York: Oxford University Press.

McCarthy, J., & Wright, P. (2004). Technology as experience. *Interactions*, *11*(5), 42–43.

Meyer, W.-U., Niepel, M., Rudolph, U., & Schützwohl, A. (1991). An experimental analysis of surprise. *Cognition & Emotion*, *5*(4), 295–311.

Meyer, W.-U., Reisenzein, R., & Schützwohl, A. (1997). Toward a process analysis of emotions: The case of surprise. *Motivation and Emotion*, *21*(3), 251–274.

Minge, M., & Riedel, L. (2013). meCUE-Ein modularer Fragebogen zur Erfassung des Nutzungserlebens. In Boll, S., Maaß, S., & Malaka, R. (2013). *Mensch & Computer 2013 – Tagungsband: 13. fachübergreifende Konferenz für interaktive und kooperative Medien*. (pp. 89-98). Walter de Gruyter.

Minge, M., Riedel, L., & Thüring, M. (2013). Und ob du wirklich richtig stehst … Zur diskriminativen Validität des User Experience Fragebogens "meCUE". In *Workshopband Mensch & Computer 2013*. Munich: Oldenbourg.

Monk, A., & Frohlich, D. (1999). Guest editorial: Computers and fun. *Personal Technologies*, *3*(3), 91–91.

Monk, A., Hassenzahl, M., Blythe, M., & Reed, D. (2002). Funology: designing enjoyment. In *CHI'02 Extended Abstracts on Human Factors in Computing Systems* (pp. 924–925). ACM.

Mullet, K., & Sano, D. (1995). *Designing visual interfaces: Communication oriented techniques* (Vol. 2550). Englewood Cliffs, NJ: SunSoft Press.

Mugge, R., Schoormans, J. P., & Schifferstein, H. N. (2008). Product attachment: Design strategies to stimulate the emotional bonding to products. In H.N.J. Schifferstein, & P. Hekkert (Eds.) *Product Experience*. (pp. 425–440). Amsterdam, Elsevier.

Mun, Y. Y., & Hwang, Y. (2003). Predicting the use of web-based information systems: self-efficacy, enjoyment, learning goal orientation, and the technology acceptance model. *International Journal of Human-Computer Studies*, *59*(4), 431–449.

Norman, D. A. (1988). *The Design of Everyday Things.* New York, NY: Basic Books.

Norman, D. A. (2002). Emotion & design: Attractive things work better. *Interactions*, *9*(4), 36–42.

Norman, D. A. (2004). *Emotional design: why we love (or hate) everyday things*. New York, NY: Basic Books.

Norman, D. A., & Draper, S. W. (1986). *User Centered System Design; New Perspectives on Human-Computer Interaction*. Hillsdale, NJ, USA: L. Erlbaum Associates Inc.

Norman, D., Miller, J., & Henderson, A. (1995). What You See, Some of What's in the Future, and How We Go About Doing It: HI at Apple Computer. In *Conference Companion on Human Factors in Computing Systems* (p. 155). New York, NY, USA: ACM.

Oliver, R. L., & Winer, R. S. (1987). A framework for the formation and structure of consumer expectations: review and propositions. *Journal of Economic Psychology*, *8*(4), 469–499.

Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions.* New York: Cambridge University Press.

Pagulayan, R. J., Keeker, K., Fuller, T., Wixon, D., Romero, R. L., & Gunn, D. V. (2002). *User-centered design in games*. Hillsdale, NJ, USA: L. Erlbaum Associates Inc.

Pohlmeyer, A. E. (2011). *Identifying Attribute Importance in Early Product Development* (PhD Thesis). Technische Universität Berlin, Berlin.

Reisenzein, R. (1999). Einschätzungstheoretische Ansätze. In *Handbuch der Emotionspsychologie* (pp. 117–138). Weinheim: Psychologie Verlags Union.

Reisenzein, R. (2000). Exploring the strength of association between the components of emotion syndromes: The case of surprise. *Cognition & Emotion*, *14*(1), 1–38.

Reisenzein, R. (2009). Emotions as metarepresentational states of mind: Naturalizing the belief–desire theory of emotion. *Cognitive Systems Research*, *10*(1), 6–20.

Reisenzein, R., Bördgen, S., Holtbernd, T., & Matz, D. (2006). Evidence for strong dissociation between emotion and facial displays: the case of surprise. *Journal of Personality and Social Psychology*, *91*(2), 295.

Robbins, S., Judge, T. A., Millett, B., & Boyle, M. (2013). *Organisational behaviour*. Pearson Higher Education AU.

Roseman, I. J., & Evdokas, A. (2004). Appraisals cause experienced emotions: Experimental evidence. *Cognition & Emotion*, *18*(1), 1–28.

Roto, V. (2007). User experience from product creation perspective. *Towards a UX Manifesto*, 31.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145.

Russell, J. A, & Feldman Barrett, L. (2009). Core affect. In D. Sander & K.R. Scherer (Eds.) *Oxford Companion to Emotion and the Affective Sciences*. (p. 104). Oxford ; New York: Oxford University Press.

Sander & K. Scherer (2009). *Oxford Companion to Emotion and the Affective Sciences.* Oxford ; New York: Oxford University Press.

Schachter, S. (1959). *The psychology of affiliation*. Stanford, CA: Stanford University Press.

Scherer, K. R. (1987). Toward a dynamic theory of emotion. *Geneva Studies in Emotion*, *1*, 1–96.

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. In K. R. Scherer, A. Schorr & T. Johnstone (Eds.). *Appraisal processes in emotion: Theory, Methods, Research* (pp. 92-120). New York and Oxford: Oxford University Press.

Scherer, K. R.(2005). What are emotions? And how can they be measured? *Social Science Information*, *44*(4), 695–729.

Schrammel, J., Geven, A., & Tscheligi, M. (2006). Using Narration to Recall and Analyse User Experiences and Emotions Evoked by Today's Technology. In *International Conference on Design and Emotion*. Citeseer.

Schultz, S. E., Kleine, R. E., & Kernan, J. B. (1989). "These are a few of my favorite things": Toward an explication of attachment as a consumer behavior construct. *Advances in Consumer Research*, *16*(1), 359–366.

Schwartz, G. E., Fair, P. L., Salt, P., Mandel, M. R., & Klerman, G. L. (1976). Facial muscle patterning to affective imagery in depressed and nondepressed subjects. *Science (New York, N.Y.)*, *192*(4238), 489–91.

Silvia, P. J. (2005). What is interesting? Exploring the appraisal structure of interest. *Emotion*, *5*(1), 89.

Silvia, P. J. (2009). Looking past pleasure: Anger, confusion, disgust, pride, surprise, and other unusual aesthetic emotions. *Psychology of Aesthetics, Creativity, and the Arts*, *3*(1), 48–51.

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, *48*(4), 813.

Stein, N. L., Hernandez, M.W. & Trabasso, T. (2010). Advances in Modeling Emotion and Thought. In: M. D. Lewis, J. M. Haviland-Jones & L. F. Barrett (Eds.). *Handbook of emotions* (pp.574-586). Guilford Press.

Sweetser, P., & Wyeth, P. (2005). GameFlow: A Model for Evaluating Player Enjoyment in Games. *Comput. Entertain.*, *3*(3), 3–3.

Tiger, L. (1992). *The pursuit of pleasure*. Transaction Publishers.

Tractinsky, N., Katz, A. S., & Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers*, *13*(2), 127–145.

Tractinsky, N., & Zmiri, D. (2006). Exploring attributes of skins as potential antecedents of emotion in HCI. In P. Fishwick (Ed). *Aesthetic Computing* (pp. 405-521). Cambridge, MA: MIT Press.

Turner, J. H. (2000). On the origins of human emotions: *A sociological inquiry into the evolution of human affect.* Stanford, CA: Stanford University Press.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, 1039–1061.

Vanhamme, J. (2000). The link between surprise and satisfaction: An exploratory research on how best to measure surprise. *Journal of Marketing Management*, *16*(6), 565–582.

Vanhamme, J., & Snelders, D. (2003). What if you surprise your customers...Will they be more satisfied? Findings from a pilot experiment. *Advances in Consumer Research*, *30*, 48–55.

Veryzer, R. W. (1995). The place of product design and aesthetics in consumer research. *Advances in Consumer Research*, *22*, 641–641.

APPENDIX A: STUDY I

A.1 INSTRUCTIONS POSITIVE GROUP

Ablauf:

Im folgenden Experiment werden Sie ein übliches Tetris-Spiel am Computer spielen. Wichtig dabei ist, dass Sie versuchen, **so viele Punkte wie möglich** zu sammeln. Für jeden neuen Stein den Sie ablegen können erhalten Sie **10 Punkte**. Jeder Proband erhält für die Teilnahme einen festen Betrag von **7 Euro**, den er durch besonders erfolgreiches Spielen aufstocken kann. Das heißt, erzielen Sie einen Score von 800 Punkten innerhalb von 5 Minuten, erhalten Sie zusätzlich 3€. Sie spielen also 5 Minuten und müssen versuchen, 800 Punkte zu erreichen. Hierfür ist eine zügige Spielweise erforderlich.

Insgesamt können Sie 10 Euro erspielen.

Bedienung:

Das Spiel wird ausschließlich über die **Pfeil-Tasten der Tastatur gesteuert**. Der Pfeil nach oben dreht den Stein im Uhrzeigersinn, die Pfeile links und rechts lassen den Stein entsprechend auf dem Spielfeld wandern, der Pfeil nach unten lässt den Stein komplett nach unten „fallen".

ACHTUNG: Lässt man den Stein „fallen", kann er unten nicht mehr bewegt werden!

Zusammenfassung:

↑= Drehen des Steines im Uhrzeigersinn

→=Stein nach rechts

←=Stein nach links

↓=Stein nach unten **Achtung**: Der Stein kann unten nicht mehr bewegt werden!

Um sich mit dem Spiel vertraut zu machen werden Sie **zu Beginn ein Probespiel von etwa fünf Minuten absolvieren**. Im Anschluss daran haben Sie auch noch einmal die Möglichkeit, Fragen zu stellen, falls Ihnen etwas unklar sein sollte.

Während des Spiels werden wir Sie zu wissenschaftlichen Zwecken per Video aufzeichnen. Zudem erheben wir die Muskelaktivität über Ihrem rechten und linken Auge. Dafür werden Ihnen fünf Elektroden mittels Kleberingen auf der Stirn befestigt. Die Erfassung der Aktivität erfolgt völlig schmerzfrei.

Bei Fragen wenden Sie sich bitte an die Versuchsleiterin!

Viel Spaß!

A.2 INSTRUCTIONS NEGATIVE GROUP

Ablauf:

Im folgenden Experiment werden Sie ein übliches Tetris-Spiel am Computer spielen. Wichtig dabei ist, dass Sie versuchen, **so viele Punkte wie möglich** zu sammeln. Für jeden neuen Stein den Sie ablegen können erhalten Sie **10 Punkte**. Jeder Proband erhält für die Teilnahme einen festen Betrag von **7 Euro**, den er durch besonders erfolgreiches Spielen aufstocken kann. Das heißt, erzielen Sie einen Score von 500 Punkten innerhalb von 5 Minuten, erhalten Sie zusätzlich 3€. Sie spielen also 5 Minuten und müssen versuchen, 500 Punkte zu erreichen. Hierfür ist eine zügige Spielweise erforderlich.

Insgesamt können Sie 10 Euro erspielen.

A.3 INSTRUCTIONS CONTROL AND NEUTRAL GROUP

Ablauf:

Im folgenden Experiment werden Sie ein übliches Tetris-Spiel am Computer spielen. Wichtig dabei ist, dass Sie versuchen, **so viele Punkte wie möglich** zu sammeln. Für jeden neuen Stein den Sie ablegen können erhalten Sie **10 Punkte**. Jeder Proband erhält für die Teilnahme einen festen Betrag von **7 Euro**, den er durch besonders erfolgreiches Spielen aufstocken kann. Das heißt, erzielen Sie einen Score von 650 Punkten innerhalb von 5 Minuten, erhalten Sie zusätzlich 3€. Sie spielen also 5 Minuten und müssen versuchen, 650 Punkte zu erreichen. Hierfür ist eine zügige Spielweise erforderlich.

Insgesamt können Sie 10 Euro erspielen.

A.4 Consent Form and demographic questionnaire

Herzlich willkommen! **VP:____**

Vielen Dank, dass Sie sich bereit erklärt haben, an der heutigen Studie *„Tetris und räumliche Vorstellung"* teilzunehmen.

Die Dauer des Experiments beträgt insgesamt ca. 30 Minuten, von denen Sie ca. 10 Minuten spielen werden. Für Ihre Teilnahme benötigen wir zuerst Ihre Einverständniserklärung.

Einverständniserklärung

Hiermit erkläre ich mich einverstanden, an der Studie *„Tetris und räumliche Vorstellung"* teilzunehmen. Meine Teilnahme erfolgt freiwillig und mir ist bewusst, dass ich die Studie zu jedem Zeitpunkt abbrechen kann. Die im Zusammenhang mit der Studie erhobenen Untersuchungsdaten werden in anonymisierter Form gespeichert und ausschließlich zu wissenschaftlichen Zwecken ausgewertet. Die Untersuchungsdaten werden streng vertraulich behandelt und können nur von entsprechend autorisierten Personen eingesehen werden.

_____ _____

Ort, Datum Unterschrift

Ein paar Fragen im Voraus:

Alter:_____Jahre

Geschlecht: O männlich O weiblich

Höchster Bildungsabschluss:_____

Kennen Sie Tetris? O Ja O Nein (Dann Frage 2 und 3 überspringen)

1)Wie oft spielen Sie Computerspiele?

nie	O	O	O	O	O	O	sehr oft

2)Wie viel Erfahrung haben Sie mit Tetris (auch frühere Erfahrung)?

gar keine	O	O	O	O	O	O	sehr viel

3)Wie gut würden Sie sich selbst einschätzen?

Gar nicht gut	O	O	O	O	O	O	sehr gut

A.5 UX QUESTIONNAIRES

A.5.1 SAM

VP:_____

Bitte füllen Sie nun den folgenden Fragebogen „SAM“ aus.

SAM bedeutet *Self Assessment Manikin*. SAM stellt Sie und Ihre Gefühle dar. Sie werden SAM heute benutzen, um Ihre emotionale Reaktion auf die Erledigung bestimmter Aufgaben zu ermitteln. Sie können über SAM mitteilen, welches Gefühl eine Interaktion in Ihnen auslöst. Geben Sie dabei bitte nicht eine Bewertung für das Produkt ab, sondern **teilen Sie IHR Gefühl mit, welches die Interaktion mit dem Produkt in Ihnen hervorruft.**

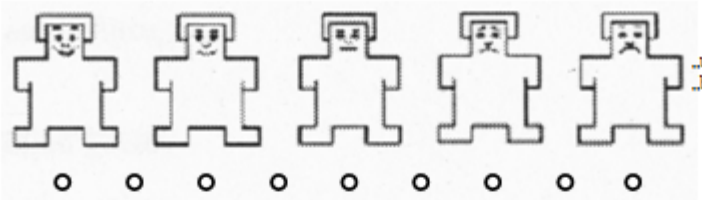
Ein Gefühl kann man aus zwei Blickwinkeln betrachten, nämlich:

→ die Stimmung eines Gefühls (positiv oder negativ)
→ die innere Erregung, die damit einhergeht

Stimmung

Die erste Zeile geht von einem breiten Lachen bis zu einem großen Stirnrunzeln. Diese Zeile versinnbildlicht Stimmungen wie „völlig zufrieden oder stolz“ bis „völlig unzufrieden oder betrübt“.

„zufrieden“ „stolz“

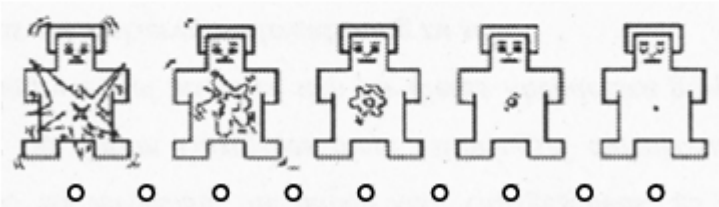


„unzufrieden“ „betrübt“

Erregung

Die zweite Zeile stellt Ihre Erregung dar, die von sehr „erregt“ oder „gespannt“ bis sehr „ruhig“ oder „schläfrig“ geht.

„erregt“ „gespannt“



ruhig“ schläfrig“

Sie geben an wie Sie sich fühlen, indem Sie in **jeder der beiden Zeilen eine Markierung** setzen. Versuchen Sie bitte, immer Ihr erstes spontanes Gefühl darzustellen. Bewerten Sie dabei nicht das Produkt, sondern geben Sie Ihr Gefühl wieder.

A.5.2 Product Emotions MANO

Wie fühlten Sie sich, als Sie das Produkt benutzt haben? **VP:_____**

	überhaupt nicht						voll und ganz
gelassen	○	○	○	○	○	○	○
angeregt	○	○	○	○	○	○	○
unglücklich	○	○	○	○	○	○	○
zufrieden	○	○	○	○	○	○	○
beschwingt	○	○	○	○	○	○	○
träge	○	○	○	○	○	○	○
begeistert	○	○	○	○	○	○	○
lebhaft	○	○	○	○	○	○	○
schläfrig	○	○	○	○	○	○	○
entspannt	○	○	○	○	○	○	○
ängstlich	○	○	○	○	○	○	○
beruhigt	○	○	○	○	○	○	○
ruhig	○	○	○	○	○	○	○
traurig	○	○	○	○	○	○	○
erstaunt	○	○	○	○	○	○	○
überrascht	○	○	○	○	○	○	○
still	○	○	○	○	○	○	○
fröhlich	○	○	○	○	○	○	○
aufgeregt	○	○	○	○	○	○	○
besorgt	○	○	○	○	○	○	○
inaktiv	○	○	○	○	○	○	○
ermüdet	○	○	○	○	○	○	○
betrübt	○	○	○	○	○	○	○
erfreut	○	○	○	○	○	○	○

A.5.3 ATTRAKDIFF & OVERALL RATING

VP:_____

Nachfolgend finden Sie einige Wortpaare, mit deren Hilfe Sie das Spiel bewerten können. Die Wortpaare stellen jeweils extreme Gegensätze dar, zwischen denen eine Abstufung möglich ist.

zum Beispiel:

unsympathisch	○	○	○	○	○	⊗	○	sympathisch

Diese Bewertung bedeutet, dass das Produkt eher sympathisch, aber noch verbesserungsbedürftig ist.

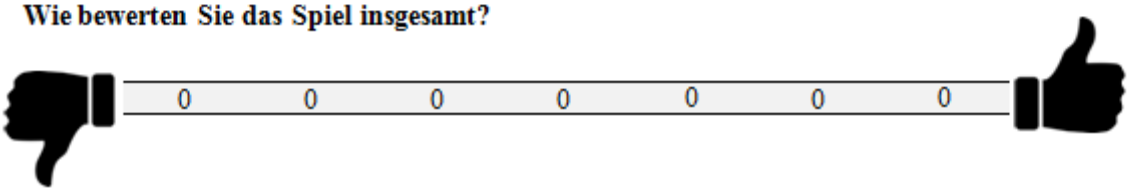
Denken Sie nicht lange über die Wortpaare nach, sondern geben Sie bitte die Einschätzung ab, die Ihnen spontan in den Sinn kommt. Vielleicht passen einige Wortpaare nicht so gut auf das Produkt, kreuzen Sie aber trotzdem bitte immer eine Antwort an. Denken Sie daran, dass es keine "richtigen" oder "falschen" Antworten gibt - nur Ihre persönliche Meinung zählt!

menschlich	○	○	○	○	○	○	○	technisch
isolierend	○	○	○	○	○	○	○	verbindend
angenehm	○	○	○	○	○	○	○	unangenehm
originell	○	○	○	○	○	○	○	konventionell
einfach	○	○	○	○	○	○	○	kompliziert
fachmännisch	○	○	○	○	○	○	○	laienhaft
hässlich	○	○	○	○	○	○	○	schön
praktisch	○	○	○	○	○	○	○	unpraktisch
sympathisch	○	○	○	○	○	○	○	unsympathisch
umständlich	○	○	○	○	○	○	○	direkt
stilvoll	○	○	○	○	○	○	○	stillos
voraussagbar	○	○	○	○	○	○	○	unberechenbar
minderwertig	○	○	○	○	○	○	○	wertvoll
ausgrenzend	○	○	○	○	○	○	○	einbeziehend

nicht vorzeigbar	○	○	○	○	○	○	○	vorzeigbar
zurückweisend	○	○	○	○	○	○	○	einladend
phantasielos	○	○	○	○	○	○	○	kreativ
gut	○	○	○	○	○	○	○	schlecht
verwirrend	○	○	○	○	○	○	○	übersichtlich
abstoßend	○	○	○	○	○	○	○	anziehend
mutig	○	○	○	○	○	○	○	vorsichtig
innovativ	○	○	○	○	○	○	○	konservativ
lahm	○	○	○	○	○	○	○	fesselnd
harmlos	○	○	○	○	○	○	○	herausfordernd
motivierend	○	○	○	○	○	○	○	entmutigend
neuartig	○	○	○	○	○	○	○	herkömmlich
widerspenstig	○	○	○	○	○	○	○	handhabbar

Vielen Dank! ☺

Wie bewerten Sie das Spiel insgesamt?



A.6 ADDITIONAL STATISTICAL DATA

A.6.1.RT DATA

Table A.6.1.1: Descriptive RTs statistics for all four groups during all three unexpected events.

	group	Mean	SD
event1	control	-85.7	329.8
	negative	684.8	581.0
	positive	314.5	439.5
	neutral	189.0	571.7
event2	control	-114.5	539.8
	negative	352.0	599.1
	positive	185.6	457.6
	neutral	126.6	276.7
event3	control	-119.7	401.9
	negative	86.4	623.3
	positive	.2	584.4
	neutral	-240.2	373.0

Table A.6.1.2: MANOVA Test of within-subjects effects of RT data.

		Type III Sum of Squares	df	Mean Square	F	p	part. η2	Observed Power
event	Sphericity Assumed	3950614.3	2	1975307.1	7.813	.001	.112	.947
event * group	Sphericity Assumed	1607828.9	6	267971.5	1.060	.390	.049	.406
Error(event)	Sphericity Assumed	31350147.2	124	252823.8				

a. Computed using alpha = ,05

A.6.2.QUESTIONNAIRE DATA

Table A.6.2.1: MANOVA Test of between-subjects effects.

Source	Wilks' Lambda	df	Error df	F	p	part. η2	Observed Power[a]
Intercept	1.000	15	48	.001	1.000	.000	.050
group	0.272	45	143.376	1.750	.007	.352	.999

a. Computed using alpha = ,05

Table A.6.2.2: Descriptive Statistics of al questionnaire subscale ratings with SDs (note: none=control).

	group	mean	SD
SAM valence	none	2,3	1,1
	negative	3,5	1,7
	positive	2,1	1,3
	neutral	2,7	1,5
SAM arousal	none	4,8	1,7
	negative	4,6	2,0
	positive	4,8	2,5
	neutral	4,8	1,7
AD pragmatic quality	none	5,2	,4
	negative	5,0	,4
	positive	5,2	,7
	neutral	5,6	,5
AD HQ IDENT	none	4,3	,7
	negative	3,7	,8
	positive	4,5	,8
	neutral	4,1	,7
AD_HQ_STIM	none	4,0	1,0
	negative	3,5	1,1
	positive	4,2	1,0
	neutral	3,7	1,3
AD attraction	none	5,3	,6
	negative	5,0	,8
	positive	5,5	,8
	neutral	5,2	,9
Overall-rating	none	5,9	,8
	negative	5,8	,8
	positive	6,2	,7
	neutral	5,7	1,0
MANO arousal	none	3,6	,9
	negative	3,2	,9
	positive	3,3	,9
	neutral	3,1	1,0

	group	mean	SD
MANO elation	none	4,7	1,2
	negative	4,4	1,0
	positive	4,7	,7
	neutral	5,0	,9
MANO quietness	none	3,7	1,0
	negative	4,0	1,0
	positive	3,2	1,2
	neutral	3,6	1,1
MANO boredom	none	2,2	,9
	negative	2,5	1,3
	positive	1,9	1,0
	neutral	1,4	,5
MANO pleasantness	none	5,4	,7
	negative	4,8	,8
	positive	5,0	,8
	neutral	5,1	,8
MANO calmness	none	4,8	1,5
	negative	5,0	1,2
	positive	4,3	1,3
	neutral	4,5	1,4
MANO unhappiness	none	1,1	,3
	negative	1,3	,5
	positive	1,7	1,2
	neutral	1,3	,4
MANO distress	none	1,6	,7
	negative	1,6	,8
	positive	1,7	,8
	neutral	1,5	,8

Table A.6.2.3: MANOVA of questionnaire subscales (note: significant findings are bold).

group	Typ III Sum of Squares	df1	df1	Mean Square	F	p	part. η2
SAM valence	**8,259**	**3**	**62**	**2,753**	**3,008**	**,037**	**,127**
SAM arousal	,062	3	62	,021	,020	,996	,001
AD PQ	**8,392**	**3**	**62**	**2,797**	**3,064**	**,035**	**,129**
AD HQI	**8,855**	**3**	**62**	**2,952**	**3,260**	**,027**	**,136**
AD HQS	4,242	3	62	1,414	1,443	,239	,065
AD attraction	3,784	3	62	1,261	1,277	,290	,058
Overall-rating	4,336	3	62	1,445	1,477	,229	,067
MANO arousal	2,896	3	62	,965	,964	,416	,045
MANO elation	3,612	3	62	1,204	1,216	,311	,056
MANO quietness	3,913	3	62	1,304	1,324	,275	,060
MANO boredom	**10,324**	**3**	**62**	**3,441**	**3,902**	**,013**	**,159**
MANO pleasantness	5,789	3	62	1,930	2,021	,120	,089
MANO calmnes	3,346	3	62	1,115	1,122	,347	,051
MANO unhappiness	6,325	3	62	2,108	2,228	,094	,097
MANO distress	,558	3	62	,186	,179	,910	,009

APPENDIX B: STUDY II

B.1 CONSENT FORM AND DEMOGRAPHIC QUESTIONNAIRE

Herzlich willkommen! **VP:____**

Vielen Dank, dass Sie sich bereit erklärt haben, an der heutigen Studie *„Tetris und räumliche Vorstellungskraft 2“* teilzunehmen.
Die Dauer des Experiments beträgt insgesamt ca. 30 Minuten, von denen Sie ca. 12 Minuten spielen werden. Für Ihre Teilnahme benötigen wir zuerst Ihre Einverständniserklärung.

Einverständniserklärung

Hiermit erkläre ich mich einverstanden, an der Studie *„Tetris und räumliche Vorstellungskraft 2“* teilzunehmen. Meine Teilnahme erfolgt freiwillig und mir ist bewusst, dass ich die Studie zu jedem Zeitpunkt abbrechen kann. Die im Zusammenhang mit der Studie erhobenen Untersuchungsdaten werden in anonymisierter Form gespeichert und ausschließlich zu wissenschaftlichen Zwecken ausgewertet. Ich bin damit einverstanden, dass die Videoaufzeichnung zu wissenschaftlichen Zwecken ausgewertet wird. Die Untersuchungsdaten werden streng vertraulich behandelt und können nur von entsprechend autorisierten Personen eingesehen werden.

_____ _____
Ort, Datum Unterschrift

Ein paar Fragen im Voraus....

Alter:_____ Jahre

Geschlecht: O männlich O weiblich

Höchster Bildungsabschluss:_____

Bei Studium: Studiengang: _____

Kennen Sie Tetris? O Ja O Nein (Dann Frage 2 und 3 überspringen)

1)Wie oft spielen Sie Computerspiele?

nie	O	O	O	O	O	O	sehr oft

2)Wie viel Erfahrung haben Sie mit Tetris (auch frühere Erfahrung)?

gar keine	O	O	O	O	O	O	sehr viel

3)Wie gut würden Sie sich selbst einschätzen?

Gar nicht gut	O	O	O	O	O	O	sehr gut

B.2 INSTRUCTIONS

Herzlich willkommen! **VP:____**

Vielen Dank, dass Sie sich bereit erklärt haben, an der heutigen Studie *„Tetris und räumliche Vorstellungskraft 2“* teilzunehmen.

Die Dauer des Experiments beträgt insgesamt ca. 30 Minuten, von denen Sie ca. 12 Minuten spielen werden. Für Ihre Teilnahme benötigen wir zuerst Ihre Einverständniserklärung.

Einverständniserklärung

Hiermit erkläre ich mich einverstanden, an der Studie *„Tetris und räumliche Vorstellungskraft 2“* teilzunehmen. Meine Teilnahme erfolgt freiwillig und mir ist bewusst, dass ich die Studie zu jedem Zeitpunkt abbrechen kann. Die im Zusammenhang mit der Studie erhobenen Untersuchungsdaten werden in anonymisierter Form gespeichert und ausschließlich zu wissenschaftlichen Zwecken ausgewertet. Ich bin damit einverstanden, dass die Videoaufzeichnung zu wissenschaftlichen Zwecken ausgewertet wird. Die Untersuchungsdaten werden streng vertraulich behandelt und können nur von entsprechend autorisierten Personen eingesehen werden.

_____ _____

Ort, Datum Unterschrift

Ein paar Fragen im Voraus....

Alter:_____ Jahre

Geschlecht: O männlich O weiblich

Höchster Bildungsabschluss:_____

Bei Studium: Studiengang: _____

Kennen Sie Tetris? O Ja O Nein (Dann Frage 2 und 3 überspringen)

1)Wie oft spielen Sie Computerspiele?

nie	O	O	O	O	O	O	sehr oft

2)Wie viel Erfahrung haben Sie mit Tetris (auch frühere Erfahrung)?

gar keine	O	O	O	O	O	O	sehr viel

3)Wie gut würden Sie sich selbst einschätzen?

Gar nicht gut	O	O	O	O	O	O	sehr gut

B.3 UX QUESTIONNAIRES

The SAM questionnaire was identical to the one used in study I (section A.5.1).

B.3.2 MECUE QUESTIONNAIRE

Nachfolgend finden Sie einige Aussagen, mit deren Hilfe Sie das Spiel bewerten können. Kreuzen Sie bitte für jede Aussage an, wie sehr Sie persönlich finden, dass sie auf das Spiel zutrifft.
Es kann sein, dass einige Aussagen nicht so gut zum Produkt passen, kreuzen Sie bitte trotzdem immer eine Antwort an.
Denken Sie nicht zu lange über einzelne Aussagen nach, sondern geben Sie bitte die Einschätzung ab, die Ihnen spontan in den Sinn kommt.
Es gibt keine "richtigen" oder "falschen" Antworten - nur Ihre persönliche Meinung zählt!

	lehne völlig ab	lehne ab	lehne eher ab	weder/ noch	stimme eher zu	stimme zu	stimme völlig zu
Durch das Spiel fühle ich mich ausgeglichen.	O	O	O	O	O	O	O
Durch das Spiel fühle ich mich erschöpft.	O	O	O	O	O	O	O
Das Spiel entspannt mich.	O	O	O	O	O	O	O
Das Spiel stimmt mich euphorisch.	O	O	O	O	O	O	O
Das Spiel nervt mich.	O	O	O	O	O	O	O
Durch das Spiel fühle ich mich fröhlich.	O	O	O	O	O	O	O
Das Spiel macht mich müde.	O	O	O	O	O	O	O
Das Spiel beruhigt mich.	O	O	O	O	O	O	O
Durch das Spiel fühle ich mich passiv.	O	O	O	O	O	O	O
Das Spiel beschwingt mich.	O	O	O	O	O	O	O
Das Spiel verärgert mich.	O	O	O	O	O	O	O
Das Spiel frustriert mich.	O	O	O	O	O	O	O

	lehne völlig ab	lehne ab	lehne eher ab	weder/ noch	stimme eher zu	stimme zu	stimme völlig zu
Wenn ich könnte, würde ich das Spiel täglich nutzen.	○	○	○	○	○	○	○
Ich würde mir genau dieses Spiel jederzeit zulegen.	○	○	○	○	○	○	○
Wenn ich mit dem Spiel zu tun habe, vergesse ich schon mal die Zeit.	○	○	○	○	○	○	○
Ich kann es kaum erwarten, das Spiel erneut zu verwenden.	○	○	○	○	○	○	○
Ich würde das Spiel gegen kein anderes eintauschen.	○	○	○	○	○	○	○
Im Vergleich zu diesem Tetris Spiel wirken andere Tetris Spiele unvollkommen.	○	○	○	○	○	○	○

Globalurteil

Geben Sie bitte abschließend an, wie Sie das Spiel insgesamt bewerten.

B.3.3 ATTRAKDIFF MINI QUESTIONNAIRE

Nachfolgend finden Sie Wortpaare, mit deren Hilfe Sie das Spiel bewerten können. Sie stellen jeweils extreme Gegensätze dar, zwischen denen eine Abstufung möglich ist.

Ein Beispiel:

einfach	○	○	○	○	○	⊗	○	kompliziert

Diese Bewertung bedeutet, dass das Spiel für Sie eher kompliziert ist.

Denken Sie nicht lange über die Wortpaare nach, sondern geben Sie bitte die Einschätzung ab, die Ihnen spontan in den Sinn kommt. Vielleicht passen einige Wortpaare nicht so gut auf das Produkt, kreuzen Sie aber trotzdem bitte immer eine Antwort an. Denken Sie daran, dass es keine "richtigen" oder "falschen" Antworten gibt - nur Ihre persönliche Meinung zählt!

einfach	○	○	○	○	○	○	○	kompliziert
hässlich	○	○	○	○	○	○	○	schön
praktisch	○	○	○	○	○	○	○	unpraktisch
stilvoll	○	○	○	○	○	○	○	stillos
voraussagbar	○	○	○	○	○	○	○	unberechenbar
minderwertig	○	○	○	○	○	○	○	wertvoll
phantasielos	○	○	○	○	○	○	○	kreativ
gut	○	○	○	○	○	○	○	schlecht
verwirrend	○	○	○	○	○	○	○	übersichtlich
lahm	○	○	○	○	○	○	○	fesselnd

B.4 ADDITIONAL STATISTICAL DATA

B.4.1 REACTION TIME DATA

Table B.4.1.1: Descriptive RT data per group and event.

	group	Mean	SD
event1	neutral	794.0	863.4
	negative	1337.8	965.0
	positive	1452.6	697.1
event2	neutral	544.3	602.7
	negative	533.3	765.1
	positive	601.7	1048.6
event3	neutral	83.4	419.5
	negative	69.4	496.3
	positive	21.4	848.6

Table B.4.1.2: ANOVA test of within subjects effects of RT data.

		Type III Sum of Squares	df1	df2	Mean Square	F	p	part. η2	Observed Power
event	Sphericity Assumed	34374068,5	2	100	17187034,2	31.256	.000	.385	1.000
event * group	Sphericity Assumed	3095274,0	4	100	773818,5	1.407	.237	.053	.424

Table B.4.1.3: Within-subjects contrast analysis of RTs for the factor event. Note: significant results are bold.

		Type III Sum of Squares	df1	df2	Mean Square	F	p	part. η2	Observed Power[a]
event	**event1 vs. event3**	**68434328.4**	**1**	**50**	**68434328.4**	**65.813**	**.000**	**.568**	**1.000**
	event2 vs. event3	**13330653.3**	**1**	**50**	**13330653.3**	**10.316**	**.002**	**.171**	**.883**
	event1 vs. event2	**21357223.7**	**1**	**50**	**21357223.7**	**22.081**	**.000**	**.306**	**.996**
event * group	event1 vs. event3	5122792.3	2	50	2561396.2	2.463	.095	.090	.473
	event2 vs. event3	165643.0	2	50	82821.5	.064	.938	.003	.059
	event1 vs. event2	3997386.5	2	50	1998693.3	2.066	.137	.076	.406

Table B.4.1.4: Simple contrasts of RT differences between the three experimental groups.

negative vs. neutral	Contrast Estimate		-217.975
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		-217.975
	Std. Error		159.282
	Sig.		.177
	95% Confidence Interval for Difference	Lower Bound	-537.903
		Upper Bound	101.952
positive vs. neutral	Contrast Estimate		-45.034
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		-45.034
	Std. Error		161.608
	Sig.		.782
	95% Confidence Interval for Difference	Lower Bound	-369.632
		Upper Bound	279.564
negative vs. positive	Contrast Estimate		172.942
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		172.942
	Std. Error		161.608
	Sig.		.290
	95% Confidence Interval for Difference	Lower Bound	-151.657
		Upper Bound	497.540

B.4.2 EMG DATA

Table B.4.2.1: Descriptive statistics of EMG data.

	group	mean	SD
event1	neutral	.1191	.40090
	negative	.0253	.15987
	positive	.1437	.30121
event2	neutral	-.0930	.30033
	negative	.1182	.40615
	positive	.0790	.44145
event3	neutral	.0145	.18091
	negative	-.0375	.10543
	positive	.2321	.63114

Table B.4.2.2: Test of within-subjects effects of event and group on EMG data.

	Type III Sum of Squares	df1	df2	Mean Square	F	p	part. η2	Observed Power
event	.077	1.671	63.501	.046	.244	.744	.006	.084
event * group	.561	3.342	63.501	.168	.886	.462	.045	.245

Computed with alpha = ,05

Table B.4.2.3: test of between-subjects effects of event and group on EMG data.

	Type III Sum of Squares	df1	df2	Mean Square	F	p	part. η2	Observed Power
Intercept	.183	1	38	.183	7.242	.011	.160	.746
group	.151	2	38	.076	3.000	.062	.136	.548

B.4.3 QUESTIONNAIRE DATA

Table B.4.2.4: Descriptive questionnaire data for all questionnaire subscales with SD.

	Group	Mean	SD
AD PQ	neutral	4.92	.77174
	negative	4.97	.68398
	positive	5.39	.76323
AD HQI	neutral	3.83	.74755

	negative	4.24	.97014
	positive	4.31	1.09999
AD HQS	neutral	4.92	.77174
	negative	4.97	.68398
	positive	5.39	.76323
AD attraction	neutral	4.78	.86130
	negative	4.94	1.08804
	positive	5.25	1.27475
SAM valence	neutral	3.17	1.24853
	negative	4.24	1.64048
	positive	2.06	.93760
SAM arousal	neutral	5.67	1.28338
	negative	4.29	.77174
	positive	4.22	1.66470
Surprise item	neutral	5.33	1.81497
	negative	6.24	.97014
	positive	5.94	1.21133
MeCUE positive emotions	neutral	4.64	.91332
	negative	4.25	.64850
	positive	4.49	1.23732
MeCUE negative emotions	neutral	1.97	.77174
	negative	2.58	.86425
	positive	1.92	.90433
MeCUE intention of use	neutral	4.15	1.02404
	negative	4.00	1.04748
	positive	4.50	1.39677
MeCUE product loyalty	neutral	2.82	1.13312
	negative	3.82	.75579
	positive	3.13	1.33917
MeCUE global rating	neutral	2.76	1.04485
	negative	2.56	1.71056
	positive	2.49	2.15310

Table B.4.2.5: MANOVA test of between subjects effects for all questionnaire subscales with event as the within-subjects factor and group as the between-subjects factor. Note: significant results are bold.

Dependent Variable	Type III Sum of Squares	df1	df2	Mean Square	F	p	part. η2	Observed Power
AD PQ	4.156	2	50	2.078	2.172	.125	.080	.424
AD HQI	2.547	2	50	1.274	1.288	.285	.049	.266
AD HQS	4.156	2	50	2.078	2.172	.125	.080	.424
AD attraction	1.756	2	50	.878	.874	.424	.034	.192
SAM valence	**15.644**	**2**	**50**	**7.822**	**10.757**	**.000**	**.301**	**.986**
SAM arousal	**11.361**	**2**	**50**	**5.680**	**6.989**	**.002**	**.218**	**.911**
meCUE positive emotions	1.408	2	50	.704	.696	.503	.027	.161
meCUE negative emotions	**5.973**	**2**	**50**	**2.987**	**3.244**	**.047**	**.115**	**.592**
meCUE intention of use	1.707	2	50	.854	.849	.434	.033	.188
meCUE product loyalty	**6.810**	**2**	**50**	**3.405**	**3.768**	**.030**	**.131**	**.662**
meCUE global rating	.267	2	50	.134	.129	.879	.005	.069

Table B.4.2.5: Simple contrasts between groups to investigate effects of surprise manipulation on all UX questionnaire subscales. Note: significant results are bold.

Group			AD PQ	AD HQI	AD HQS	AD attraction	SAM valence	SAM arousal	meCUE positive emotions	meCUE negative emotions	meCUE intention of use	meCUE global rating	meCUE product loyalty
neutral vs. positive	Contrast Estimate		-.623	-.494	-.623	-.435	.683	1.001	.154	.063	-.301	.166	-.270
	Hypothesized Value		0	0	0	0	0	0	0	0	0	0	0
	Difference (Estimate - Hypothesized)		-.623	-.494	-.623	-.435	.683	1.001	.154	.063	-.301	.166	-.270
	Std. Error		.326	.332	.326	.334	.284	.301	.335	.320	.334	.339	.317
	p		.062	.142	.062	.199	**.020***	**.002****	.648	.845	.372	.626	.399
	95% Confidence Interval for Difference	Lower Bound	-1.278	-1.160	-1.278	-1.106	.112	.397	-.520	-.580	-.973	-.515	-.906
		Upper Bound	.032	.172	.032	.236	1.254	1.604	.827	.705	.370	.847	.367
negative vs. positive	Contrast Estimate		-.552	-.074	-.552	-.285	1.337	.050	-.245	.748	-.428	.044	.595
	Hypothesized Value		0	0	0	0	0	0	0	0	0	0	0
	Difference (Estimate - Hypothesized)		-.552	-.074	-.552	-.285	1.337	.050	-.245	.748	-.428	.044	.595
	Std. Error		.331	.336	.331	.339	.288	.305	.340	.324	.339	.344	.322
	p		.102	.828	.102	.405	**.000*****	.871	.475	**.025***	.213	.900	.070
	95% Confidence Interval for Difference	Lower Bound	-1.216	-.749	-1.216	-.966	.758	-.563	-.928	.097	-1.110	-.647	-.051
		Upper Bound	.113	.602	.113	.396	1.916	.662	.438	1.400	.253	.735	1.241
neutral vs. negative	Contrast Estimate		0.71	.421	0.71	.151	.654	-.951	-.399	.686	-.127	-.123	.865
	Hypothesized Value		0	0	0	0	0	0	0	0	0	0	0
	Difference (Estimate - Hypothesized)		.071	.421	0.71	.151	.654	-.951	-.399	.686	-.127	-.123	.865
	Std. Error		.331	.331	.331	.331	.331	.331	.331	.331	.331	.331	.331
	p		.831	.217	.831	.659	**.028***	**.003****	.247	**.040***	.710	.723	**.010***
	95% Confidence Interval for Difference	Lower Bound	-.593	-.172	-0.032	-236	-1.254	-1.604	-.827	-.705	-.370	-.847	-.367
		Upper Bound	.736	1.160	1.278	1.106	-0.112	-0.397	.520	.580	.973	.515	.906

APPENDIX C: STUDY III

C.1 CONSENT FORM AND DEMOGRAPHIC QUESTIONNAIRE

Herzlich willkommen! **VP:____**

Vielen Dank, dass Sie sich bereit erklärt haben, an der heutigen Studie *„Testung einer mobilen Applikation - ÖPVN"* teilzunehmen. Die Dauer des Experiments beträgt insgesamt ca. 60 Minuten.

Einverständniserklärung

Hiermit erkläre ich mich einverstanden, an der Studie *„Testung einer mobilen Applikation - ÖPVN "* teilzunehmen. Meine Teilnahme erfolgt freiwillig und mir ist bewusst, dass ich die Studie zu jedem Zeitpunkt abbrechen kann. Die im Zusammenhang mit der Studie erhobenen Untersuchungsdaten werden in anonymisierter Form gespeichert und ausschließlich zu wissenschaftlichen Zwecken ausgewertet. Die Untersuchungsdaten werden streng vertraulich behandelt und können nur von entsprechend autorisierten Personen eingesehen werden.

_____ _____

Ort, Datum Unterschrift

Alter:_____ Jahre

Geschlecht: O männlich O weiblich

Besitzen Sie ein Smartphone/Tablet? O Ja O Nein

Benutzen Sie ÖPNV Applikationen? O Ja O Nein
Wenn ja, welche? _____

2)Wie häufig nutzen Sie diese Applikationen?

nie	O	O	O	O	O	O	Mehrmals täglich

Ablauf:

Im folgenden Experiment werden Sie eine ÖPNV Applikation für mobile Endgeräte testen. Die Applikation befindet sich noch in der Entwicklungsphase und ist noch nicht an die real existierenden Datenbanken der Verkehrsbetriebe angeschlossen.

Stellen Sie sich vor, ein guter Freund wird sie Anfang Februar besuchen kommen und sie möchten seinen Besuch so angenehm und interessant gestalten, wie möglich. Hierzu möchten Sie genaue Vorbereitungen treffen, welche Orte Sie zusammen besuchen könnten, wann Sie dies tun können und mit welchen öffentlichen Verkehrsmitteln Sie diese erreichen können.

Hierzu werden Sie einige Aufgaben mit der Applikation ausführen um Informationen über bestimmte Orte an möglichen Reisestrecken in der Stadt zu sammeln.

Zwei Probeaufgabe wird Ihnen gleich vom Versuchsleiter gezeigt werden, dann erhalten Sie die Möglichkeit, anhand von 3 Aufgaben selbst zu üben und sich mit der App vertraut zu machen. Erst nach dieser Eingewöhnungsphase beginnt die eigentliche Evaluation.

Während der Studie werden wir Sie zu wissenschaftlichen Zwecken per Video aufzeichnen. Bei Fragen wenden Sie sich bitte an die Versuchsleiterin!

Viel Spaß!

C.2 UX QUESTIONNAIRES

The SAM questionnaire was identical to the one used in study I (see section A.5.1)

C.2.1 MECUE QUESTIONNAIRE

Fragebogen zur Bewertung interaktiver Produkte

Auf den folgenden Seiten finden Sie verschiedene Aussagen, mit denen Sie die App bewerten können.

Kreuzen Sie bitte für jede Aussage an, wie sehr Sie persönlich finden dass sie auf die App zutrifft. Es kann sein, dass einige Aussagen nicht so gut zur App passen, kreuzen Sie bitte trotzdem immer eine Antwort an.

Denken Sie nicht zu lange über einzelne Aussagen nach, sondern geben Sie bitte die Einschätzung ab, die Ihnen spontan in den Sinn kommt.

Es gibt keine „richtigen" oder „falschen" Antworten – nur Ihre persönliche Meinung zählt!

Beispiel:

	lehne völlig ab	lehne ab	lehne eher ab	weder/ noch	stimme eher zu	stimme zu	stimme völlig zu
Wenn ich könnte, würde ich die App täglich nutzen.	○	○	○	○	○	⊗	○

Diese Bewertung bedeutet, dass Sie die App eher häufiger nutzen würden.

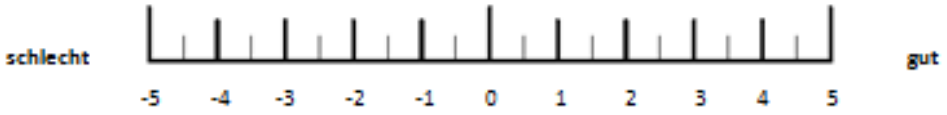
	lehne völlig ab	lehne ab	lehne eher ab	weder/ noch	stimme eher zu	stimme zu	stimme völlig zu
Die App lässt sich einfach benutzen.	O	O	O	O	O	O	O
Die App ist kreativ gestaltet.	O	O	O	O	O	O	O
Die Funktionen der App sind genau richtig für meine Ziele.	O	O	O	O	O	O	O
Die App verleiht mir ein höheres Ansehen.	O	O	O	O	O	O	O
Ohne die App kann ich nicht leben.	O	O	O	O	O	O	O
Das Design wirkt attraktiv.	O	O	O	O	O	O	O
Es wird schnell klar, wie man die App bedienen muss.	O	O	O	O	O	O	O
Durch die App werde ich anders wahrgenommen.	O	O	O	O	O	O	O
Ich halte die App für absolut nützlich	O	O	O	O	O	O	O
Die App ist wie ein Freund für mich.	O	O	O	O	O	O	O
Die App ist stilvoll.	O	O	O	O	O	O	O
Wenn ich die App verlieren würde, würde für mich eine Welt zusammenbrechen.	O	O	O	O	O	O	O
Die Bedienung der Appist verständlich.	O	O	O	O	O	O	O
Meine Freunde dürfen ruhig neidisch auf die App sein.	O	O	O	O	O	O	O
Mithilfe der App kann ich meine Ziele erreichen.	O	O	O	O	O	O	O

	lehne völlig ab	lehne ab	lehne eher ab	weder/ noch	stimme eher zu	stimme zu	stimme völlig zu
Durch die App fühle ich mich ausgeglichen.	O	O	O	O	O	O	O
Durch die App fühle ich mich erschöpft.	O	O	O	O	O	O	O
Die App entspannt mich.	O	O	O	O	O	O	O
Die App stimmt mich euphorisch.	O	O	O	O	O	O	O
Die App nervt mich.	O	O	O	O	O	O	O
Durch die App fühle ich mich fröhlich.	O	O	O	O	O	O	O
Die App macht mich müde.	O	O	O	O	O	O	O
Die App beruhigt mich.	O	O	O	O	O	O	O
Durch die App fühle ich mich passiv.	O	O	O	O	O	O	O
Die App beschwingt mich.	O	O	O	O	O	O	O
Die App verärgert mich.	O	O	O	O	O	O	O
Die App frustriert mich.	O	O	O	O	O	O	O

	lehne völlig ab	lehne ab	lehne eher ab	weder/ noch	stimme eher zu	stimme zu	stimme völlig zu
Wenn ich könnte, würde ich die App täglich nutzen.	O	O	O	O	O	O	O
Ich würde mir genau diese App jederzeit zulegen.	O	O	O	O	O	O	O
Wenn ich mit der App zu tun habe, vergesse ich schon mal die Zeit.	O	O	O	O	O	O	O
Ich kann es kaum erwarten, die App erneut zu verwenden.	O	O	O	O	O	O	O
Ich würde die App gegen keine andere eintauschen.	O	O	O	O	O	O	O
Im Vergleich zu dieser ÖPNV App wirken andere Apps unvollkommen.	O	O	O	O	O	O	O

Globalurteil

Geben Sie bitte abschließend an, wie Sie die App insgesamt bewerten.



C.2.2 Surprise item and open questions

Wir möchten Sie bitten, nachfolgend noch einige Fragen zur Applikation zu beantworten:

Wie überraschend war das Auftreten der Animationen für Sie?

Gar nicht überraschend	○	○	○	○	○	○	○	Sehr überraschend

Hat Ihnen die App gefallen? **Ja** ○ **Nein** ○ **Neutral** ○

Können sie 3 Dinge nennen warum/warum nicht?

1)

2)

3)

Würden Sie diese App weiterhin benutzen? **Ja** ○ **Nein** ○ **Neutral** ○

Können sie 3 Grunde nennen warum/warum nicht?

1)

2)

3)

Was halten Sie von den Animationen in der App?

C.3 INSTRUCTIONS AS SEEN ON THE TABLET

Begrüßung

Herzlich Willkommen zur Studie „Testung einer mobilen Applikation – ÖPNV".

Im Folgenden werden Sie die Funktionen einer neu entwickelten App testen. Dies erfolgt in 3 Phasen:

Erst wird eine „Kennenlernphase" stattfinden, in der Sie in Ruhe und in Ihrem eigenen Tempo einige Beispielaufgaben erledigen können. Auf diese Phase folgen zwei Testphasen, welche beide etwas länger sind als die Kennenlernphase. Zwischen den drei Phasen wird es jeweils eine kurze Pause geben.

Neue Seite:

Um die Funktionsweise der Applikation zu erfahren, werden Sie im Folgenden mehrere Aufgaben erledigen, die mit Reiseplanung zu tun haben.

Bitte stellen Sie sich hierfür Folgendes vor:

Ein guter Freund wird Sie Anfang Februar besuchen kommen und sie möchten seinen Besuch so angenehm und interessant gestalten wie möglich. Hierzu möchten Sie genaue Vorbereitungen treffen, welche Orte Sie zusammen besuchen könnten, wann Sie dies tun können und mit welchen öffentlichen Verkehrsmitteln Sie diese erreichen können.

Zu diesem Zweck werden Ihnen immer wieder kurze Suchaufgaben gestellt werden. Beispiel:

„Sie fahren von der Station **XYZ** nach **XYZ** am 15.02.2014 um 16:15. Klicken Sie auf die 3. Verbindung. Bitte finden Sie heraus, welche Besuchsziele sich in der Nähe der Haltestelle **XY** befinden welche mit der U8 erreicht werden kann."

Neue Seite:

Jede Aufgabe wird Ihnen ausgedruckt vorliegen, so dass Sie jederzeit die Möglichkeit haben, Abfahrts- und Ankunftsort sowie Uhrzeit und Frage erneut anzuschauen.

Bitte behalten Sie im Hinterkopf, dass sich die Applikation noch in der Entwicklung befindet und aus diesem Grund noch nicht an reale Datenbanken angeschlossen ist. Alle Straßennamen und Orte sind deswegen zu diesem Zeitpunkt noch fiktiv.

Neue Seite:

Haben Sie noch Fragen? Dann wenden Sie sich bitte nun an Ihren Versuchsleiter.

Neue Seite:

Es folgen nun einige Probeaufgaben. Bitte versuchen Sie diese fehlerfrei zu bewältigen. Bei Fragen wenden Sie sich bitte jederzeit an Ihren Versuchsleiter.

Trial 1-5

Nach Beendigung der Trainingsaufgaben:

Kurze Pause

Bei Fragen wenden Sie sich bitte nun an den Versuchsleiter.

Neue Seite:

Jetzt folgen 10 weitere Aufgaben. Bitte versuchen Sie diese Aufgaben so schnell und präzise wie möglich zu erledigen. Nach einigen Aufgaben erfolgt eine kurze Abfrage durch den SAM Fragebogen.

Nach Ende dieser Phase werden wir Sie bitten die Zeit zu schätzen, die Sie für die 10 Aufgaben benötigt haben.

Bitte wenden Sie sich während der Bearbeitung nun nicht mehr an den Versuchsleiter.

Nach Beendigung der Baselineaufgaben:

Kurze Pause

Bitte füllen Sie die bereitliegenden Fragebögen aus.

Neue Seite:

Jetzt folgen 10 weitere Aufgaben. Bitte versuchen Sie diese Aufgaben so schnell und präzise wie möglich zu erledigen. Nach einigen Aufgaben erfolgt eine kurze Abfrage durch den SAM Fragebogen. Nach Ende dieser Phase werden wir Sie bitten die Zeit zu schätzen, die Sie für die 10 Aufgaben benötigt haben.

Bitte wenden Sie sich während der Bearbeitung nun nicht mehr an den Versuchsleiter.

Nach Beendigung der Experimentalaufgaben:

Vielen Dank, die Testphase ist nun beendet.

Bitte füllen Sie die bereitliegenden Fragebögen aus.

C.4 TASKS DESCRIPTIONS

Pohlandstraße

Sie möchten herausfinden, was sich für interessante Orte in der Nähe der Haltestelle ***Pohlandstraße*** befinden.

Hierfür fahren sie von der Haltestelle Bedfordstraße zur Haltestelle Siechenstraße,

am 08.02.2014 um 19.05

Bitte klicken Sie auf die 4. Angezeigte Verbindung und finden Sie die Pohlandstraße.

Richardstraße

Sie möchten herausfinden, was sich für interessante Orte in der Nähe der Haltestelle ***Richardstraße*** befinden.

Hierfür fahren sie von der Haltestelle Fischerstraße zur Haltestelle Stoltzestraße,

am 06.02.2014 um 16.55

Bitte klicken Sie auf die 2. Angezeigte Verbindung und finden Sie die Richardstraße.

Kurmarkstraße

Sie möchten herausfinden, was sich für interessante Orte in der Nähe der Haltestelle ***Kurmarkstraße*** befinden.

Hierfür fahren sie von der Haltestelle Gutleutstraße zur Haltestelle Brunnenstraße,

am 02.02.2014 um 06:25

Sturmvogelweg

Sie möchten herausfinden, was sich für interessante Orte in der Nähe der Haltestelle ***Sturmvogelweg*** befinden.

Hierfür fahren sie von der Haltestelle Röhrborngasse zur Haltestelle Lehmansstraße,

am 01.02.2014 um 18.35

Bitte klicken Sie auf die 2. Angezeigte Verbindung und finden Sie die Sturmvogelweg.

Souchaystraße

Sie möchten herausfinden, was sich für interessante Orte in der Nähe der Haltestelle ***Souchaystraße*** befinden.

Hierfür fahren sie von der Haltestelle Friesenstraße zur Haltestelle Zinzendorfweg,

am 18.02.2014 um 12.15

Bitte klicken Sie auf die 4. Angezeigte Verbindung und finden Sie die Souchaystraße.

Maßholderpfad

Sie möchten herausfinden, was sich für interessante Orte in der Nähe der Haltestelle ***Maßholderpfad*** befinden.

Hierfür fahren sie von der Haltestelle Löscherstraße zur Haltestelle Seeadlerstieg,

am 16.02.2014 um 20.45

Tannenkopfweg

Sie möchten herausfinden, was sich für interessante Orte in der Nähe der Haltestelle ***Tannenkopfweg*** befinden.

Hierfür fahren sie von der Haltestelle Schadowstraße zur Haltestelle Simrockstraße,

am 20.02.2014 um 13.10

Bitte klicken Sie auf die 3. Angezeigte Verbindung und finden Sie den Tannenkopfweg.

Pfeiferstraße

Sie möchten herausfinden, was sich für interessante Orte in der Nähe der Haltestelle ***Pfeiferstraße*** befinden.

Hierfür fahren sie von der Haltestelle Gellertstraße zur Haltestelle Steubenstraße,

am 24.02.2014 um 15.50

Bitte klicken Sie auf die 1. Angezeigte Verbindung und finden Sie die Pfeiferstraße.

Albrechtshöhe

Sie möchten herausfinden, was sich für interessante Orte in der Nähe der Haltestelle ***Albrechtshöhe*** befinden.

Hierfür fahren sie von der Haltestelle Klemensstraße zur Haltestelle Jochenmattweg,

am 28.02.2014 um 18.35

Bitte klicken Sie auf die 2. Angezeigte Verbindung und finden Sie die Albrechtshöhe.

Bründlfeldweg

Sie möchten herausfinden, was sich für interessante Orte in der Nähe der Haltestelle ***Bründlfeldweg*** befinden.

Hierfür fahren sie von der Haltestelle Gundhofstraße zur Haltestelle Foockenstraße,

am 12.02.2014 um 10.10

Bitte klicken Sie auf die 2. Angezeigte Verbindung und finden Sie den Bründlfeldweg.

Rathenauplatz

Sie möchten herausfinden, was sich für interessante Orte in der Nähe der Haltestelle ***Rathenauplatz*** befinden.

Hierfür fahren sie von der Haltestelle Lochmühlenweg zur Haltestelle Corinthstraße,

am 03.02.2014 um 09.10

Bitte klicken Sie auf die 4. Angezeigte Verbindung und finden Sie die Rathenauplatz.

Handelsstraße

Sie möchten herausfinden, was sich für interessante Orte in der Nähe der Haltestelle ***Handelsstraße*** befinden.

Hierfür fahren sie von der Haltestelle Neumannstraße zur Haltestelle Ludwigsstraße,

am 22.02.2014 um 13.10

Bitte klicken Sie auf die 3. Angezeigte Verbindung und finden Sie die Handelsstraße.

Dopplerstraße

Sie möchten herausfinden, was sich für interessante Orte in der Nähe der Haltestelle ***Dopplerstraße*** befinden.

Hierfür fahren sie von der Haltestelle Hasenbanckweg zur Haltestelle Regnitzstraße,

am 26.02.2014 um 07.20

Bitte klicken Sie auf die 3. Angezeigte Verbindung und finden Sie die Dopplerstraße.

Gontardstraße

Sie möchten herausfinden, was sich für interessante Orte in der Nähe der Haltestelle ***Gontardstraße*** befinden.

Hierfür fahren sie von der Haltestelle Ferdinandstor zur Haltestelle Reichseltraße

am 14.02.2014 um 10.10

Bitte klicken Sie auf die 2. Angezeigte Verbindung und finden Sie die Gontardstraße.

Gurlittstraße

Sie möchten herausfinden, was sich für interessante Orte in der Nähe der Haltestelle ***Gurlittstraße*** befinden.

Hierfür fahren sie von der Haltestelle Pinkertbrücke zur Haltestelle Quettigstraße,

am 09.02.2014 um 08.30

Bitte klicken Sie auf die 2. Angezeigte Verbindung und finden Sie die Gurlittstraße.

Blauhandgasse

Sie möchten herausfinden, was sich für interessante Orte in der Nähe der Haltestelle ***Blauhandgasse*** befinden.

Hierfür fahren sie von der Haltestelle Lehngutstraße zur Haltestelle Friedensallee,

am 23.02.2014 um 17.50

Bitte klicken Sie auf die 3. Angezeigte Verbindung und finden Sie die Blauhandgasse.

Dunckerstraße

Sie möchten herausfinden, was sich für interessante Orte in der Nähe der Haltestelle ***Dunckerstraße*** befinden.

Hierfür fahren sie von der Haltestelle Eisbachstraße zur Haltestelle Heinrichshöhe,

am 25.02.2014 um 08.30

Bitte klicken Sie auf die 2. Angezeigte Verbindung und finden Sie die Dunckerstraße.

Heisterstraße

Sie möchten herausfinden, was sich für interessante Orte in der Nähe der Haltestelle ***Heisterstraße*** befinden.

Hierfür fahren sie von der Haltestelle Buchgrabenweg zur Haltestelle Schützenplatz,

am 07.02.2014 um 11.20

Bitte klicken Sie auf die 1. Angezeigte Verbindung und finden Sie die Heisterstraße.

Rathenauplatz

Sie möchten herausfinden, was sich für interessante Orte in der Nähe der Haltestelle ***Rathenauplatz*** befinden.

Hierfür fahren sie von der Haltestelle Lochmühlenweg zur Haltestelle Corinthstraße,

am 03.02.2014 um 09.10

Bitte klicken Sie auf die 4. Angezeigte Verbindung und finden Sie den Rathenauplatz.

Altmannbrücke

Sie möchten herausfinden, was sich für interessante Orte in der Nähe der Haltestelle ***Altmannbrücke*** befinden.

Hierfür fahren sie von der Haltestelle Bahnhofstraße zur Haltestelle Bergfelderweg,

am 27.02.2014 um 17.50

Bitte klicken Sie auf die 3. Angezeigte Verbindung und finden Sie die Altmannbrücke.

Oltersteinweg

Sie möchten herausfinden, was sich für interessante Orte in der Nähe der Haltestelle ***Oltersteinweg*** befinden.

Hierfür fahren sie von der Haltestelle Traubelstraße zur Haltestelle Klosterstraße,

am 17.02.2014 um 11.20

Bitte klicken Sie auf die 1. Angezeigte Verbindung und finden Sie den Oltersteinweg.

Mandelbaumweg

Sie möchten herausfinden, was sich für interessante Orte in der Nähe der Haltestelle ***Mandelbaumweg*** befinden.

Hierfür fahren sie von der Haltestelle Feiersteigweg zur Haltestelle Schoberstraße,

am 05.02.2014 um 14.40

Bitte klicken Sie auf die 4. Angezeigte Verbindung und finden Sie den Mandelbaumweg.

Fichtenstraße

Sie möchten herausfinden, was sich für interessante Orte in der Nähe der Haltestelle ***Fichtenstraße*** befinden.

Hierfür fahren sie von der Haltestelle Erlweinstraße zur Haltestelle Dalbergstraße,

am 04.02.2014 um 07.20

Bitte klicken Sie auf die 3. Angezeigte Verbindung und finden Sie die Fichtenstraße.

C.5 ADDITIONAL STATISTICAL DATA

C.5.1 REACTION TIME DATA

Table C.5.1.1: Descriptive statistics of RTs per group and event.

	group	Mean	SD
event1	group A: 3 events	859.12	947.29
	group B: 10 events	494.52	1090.13
event2	group A: 3 events	620.32	1008.47
	group B: 10 events	406.22	1648.41
event3	group A: 3 events	21.65	830.48
	group B: 10 events	41.71	1148.86

Table C.5.1.2: Greenhouse Geisser Test of Sphericity.

	Mauchley's W	Approx. Chi-aquare	df	p	Greenhouse-Geisser
events	.763	8.098	2	.017	.809

Table C.5.1.3: Test of within subjects effects with events as the within-subjetcs factor and group as the between subjects factor.

		Type III Sum of Squares	df1	df2	Mean Square	F	p	part. η2	Observed Power[a]
event	Sphericity Assumed	7416928.61	2	62	3708464.31	4.256	.019	.121	.724
	Greenhouse-Geisser	7416928.61	1.617	50.139	4585785.92	4.256	.027	.121	.656
event * group	Sphericity Assumed	619397.45	2	62	309698.72	.355	.702	.011	.105
	Greenhouse-Geisser	619397.45	1.617	50.139	382965.00	.355	.657	.011	.099

Table C.5.1.4: Within-subjects contrast analysis of RTs for the factor event. Note: significant results are bold.

	events	Type III Sum of Squares	df1	df2	Mean Square	F	p	part. η2	Observed Power
event	**event1 vs. event3**	**13722238,84**	**1**	**31**	**13722238.84**	**11.994**	**.002**	**.279**	**.918**
	event2 vs. event 3	**7646642,44**	**1**	**31**	**7646642.44**	**5.024**	**.032**	**.139**	**.584**
	event1 vs. event2	881904,55	1	31	881904,55	.344	.562	.011	.088
event * group	event1 vs. event3	1219564,30	1	31	1219564.30	1.066	.310	.033	.170
	event2 vs. event3	451940,25	1	31	451940.25	.297	.590	.009	.083
	event1 vs. event2	186687,80	1	31	186687,80	.073	.789	.002	.058

C.5.2 SAM QUESTIONNAIRE DATA

Table C.5.2.1: Descriptive values of the SAM questionnaire for both subscales.

valence	N	mean	SD
3events	17	2.7647	1.09141
10events	16	3.4375	1.50416
arousal			
3events	17	6.1765	2.00734
10events	16	5.8125	1.83371

Table C.5.2.2: Comparisons of SAM subscales rating after the baseline phase (top) and the experimental phase (bottom), comparing ratings between groups. Significant results are bold.

time	source	Type III Sum of squares	df1	df2	Mean square	F	p	part. η2	Observed power
groups after baseline	SAM valence	3.731	1	31	3.731	2.182	.150	.066	.299
	SAM arousal	1.092	1	31	1.092	0.295	.591	.009	.082
groups after experimental	**SAM valence**	**16.257**	**1**	**31**	**16.257**	**9.335**	**.005**	**.231**	**.841**
	SAM arousal	.045	1	31	.045	0.009	.926	.009	.051

Table C.5.2.3: Comparisons of SAM subscale ratings between measurement periods for group A (top), and group B (bottom).

group		source	Type III Sum of squares	df1	df2	Mean square	F	p	part. η2	Observed power
A:3 events	SAM valence	Baseline vs. experimental	.735	1	34	.735	.631	.433	.019	.120
	SAM arousal	Baseline vs. experimental	1.059	1	34	1.059	.222	.641	.007	.074
B: 10 events	SAM valence	Baseline vs. experimental	2.168	1	30	2.168	.942	.340	.030	.156
	SAM arousal	Baseline vs. experimental	.206	1	30	.206	.053	.820	.002	.056

C.5.3: MECUE QUESTIONNAIRE DATA

Table C.5.3: Descriptive questionnaire data for the meCUE questionnaire.

	group	Mean	SD	N
usability	A: 3 events	6.1961	.66728	17
	B: 10 events	6.0392	.69604	17
aesthetics	A: 3 events	3.8235	1.33395	17
	B: 10 events	4.1765	1.35973	17
status	A: 3 events	2.3529	1.23867	17
	B: 10 events	2.3725	1.24656	17
utility	A: 3 events	5.1961	1.24197	17
	B: 10 events	4.8039	1.26962	17
bonding	A: 3 events	1.7059	1.07937	17
	B: 10 events	1.6275	.95657	17
positive emotions	A: 3 events	2.7745	1.19015	17
	B: 10 events	3.0784	1.25155	17
negative emotions	A: 3 events	2.2059	.79828	17
	B: 10 events	2.7843	1.13003	17
intention of use	A: 3 events	3.1961	1.36452	17
	B: 10 events	3.0000	1.12423	17
product loyalty	A: 3 events	3.2549	1.12749	17
	B: 10 events	3.1569	1.39999	17

Table C.5.4: Descriptive questionnaire data of the differences between baseline and experimental phase for the meCUE questionnaire.

	group	N	Mean	Std. Deviation
usability_difference	A: 3 events	17	.0196	.52002
	B: 10 events	17	.0196	.50649
aesthetics_difference	A: 3 events	17	.3333	.92045
	B: 10 events	17	.0980	.65367
status_difference	A: 3 events	17	.1176	.82446
	B: 10 events	17	-.2745	.64802
utility_difference	A: 3 events	17	-.0588	.88377
	B: 10 events	17	-.1961	.57806
bonding_difference	A: 3 events	17	-.1176	.33211

	B: 10 events	17	-.1569	.35586
positive_emo_difference	A: 3 events	17	-.0686	.67971
	B: 10 events	17	-.0392	.63609
negative_emo_difference	A: 3 events	17	-.6471	1.27732
	B: 10 events	17	.0000	.47140
intention_of_use_difference	A: 3 events	17	.0784	.67216
	B: 10 events	17	-.1765	.55425
product_loyalty_difference	A: 3 events	17	.2157	.47054
	B: 10 events	17	.0784	.50730
global_difference	A: 3 events	17	.5147	1.19069
	B: 10 events	17	.0735	.66594

Table C.5.5: Comparison of meCUE subscales after the experimental phase.

		Levene's Test		t-test for Equality of Means						
									95% Confidence Interval	
		F	p	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
usability	Equal variances assumed	.000	1.000	.000	32	1.000	,.00000	.17606	-.35863	.35863
	Equal variances not assumed			.000	31.978	1.000	.00000	.17606	-.35864	.35864
aesthetics	Equal variances assumed	1.419	.242	.859	32	.397	.23529	.27381	-.32244	.79302
	Equal variances not assumed			.859	28.866	.397	.23529	.27381	-.32482	.79541
status	Equal variances assumed	.192	.664	1.542	32	.133	.39216	.25434	-.12591	.91022
	Equal variances not assumed			1.542	30.308	.133	.39216	.25434	-.12704	.91136
utility	Equal variances assumed	.918	.345	.536	32	.596	.13725	.25612	-.38445	.65896
	Equal variances not assumed			.536	27.572	.596	.13725	.25612	-.38776	.66227
bonding	Equal variances assumed	.259	.614	.332	32	.742	.03922	.11805	-.20125	.27969
	Equal variances not assumed			.332	31.849	.742	.03922	.11805	-.20130	.27973
positive emotions	Equal variances assumed	.193	.663	-.130	32	.897	-.02941	.22578	-.48932	.43049
	Equal variances not assumed			-.130	31.860	.897	-.02941	.22578	-.48940	.43057
negative emotions	Equal variances assumed	10.569	.003	-1.959	32	.059	-.64706	.33022	-1.31969	.02558
	Equal variances not assumed			-1.959	20.279	.064	-.64706	.33022	-.33528	.04116

intention of use	Equal variances assumed	.428	.518	1.206	32	.237	.25490	.21130	-.17550	.68530
	Equal variances not assumed			1.206	30.879	.237	.25490	.21130	-.17611	.68591
product loyalty	Equal variances assumed	.015	.903	.818	32	.419	.13725	.16782	-.20458	.47909
	Equal variances not assumed			.818	31.821	.420	.13725	.16782	-.20465	.47916
overall rating	Equal variances assumed	.772	.386	1.333	32	.192	.44118	.33088	-.23281	1.11516
	Equal variances not assumed			1.333	25.118	.194	.44118	.33088	-.24013	1.12248