

Dauerwahrnehmung und Nutzererleben in der Mensch-Computer-Interaktion

vorgelegt von

Dipl.-Psych.

Anna Katharina Trukenbrod (geborene Trapp)

ORCID: 0000-0003-4994-2288

an der Fakultät V – Verkehrs- und Maschinensysteme

der Technischen Universität Berlin

zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften

– Dr. rer. nat. –

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. Markus Feufel

Gutachter: Prof. Dr. Manfred Thüring

Gutachter: Prof. Dr. Florian Klapproth

Tag der wissenschaftlichen Aussprache: 22.11.2019

Berlin 2020

“Allting är mycket osäkert, och det är just det som lugnar mig.”

“Alles ist sehr ungewiss und genau das beruhigt mich.”

Too-ticki in Trollvinter, Tove Jansson, 1957

Danksagung

Eine Dissertation entsteht nie ohne Hilfestellung und für all die geleistete Unterstützung möchte ich mich herzlich bedanken, zunächst bei meinem Betreuer und Erstgutachter Manfred Thüring. Du hast mir Spielraum gegeben, eigene Ideen zu entwickeln und Schwerpunkte in meiner Arbeit zu setzen, während Du gleichzeitig immer als Sparring Partner für einen Gedankenaustausch zur Verfügung standest. Das erfordert Vertrauen und einander Ernst zu nehmen, aber auch zu ertragen, wenn das Gegenüber manchmal Warnungen in den Wind schlägt, scheitert und daran wächst. Für dieses Arbeitsklima und die Unterstützung möchte ich Dir danken. Florian Klapproth hat seine Rolle als Zweitbetreuer und -gutachter bereits früh im Prozess meiner Promotion übernommen. Ich danke Dir, Florian, dafür, dass Du bei Fragen und Diskussionen immer sofort zur Verfügung standest und Dir die Zeit für mich genommen hast.

Ein wichtiger Faktor beim Erstellen einer Dissertation ist das zwischenmenschliche Klima in der Arbeitsgruppe. Meine früheren und derzeitigen Kollegen und Kolleginnen Angelika Engling, Carolin Wienrich, Christin Böschow, Fabienne Roche, Friederice Schröder, Karin Scherrinsky-Pingel, Katharina Jungnickel, Michael Minge, Nikolaus Rötting, Nils Backhaus, Ronny Reckin, Sandra Epple und Stefan Brandenburg haben durch viel Lachen, geteilte Erlebnisse und Verlässlichkeit dafür gesorgt, dass ich mich immer sehr wohl gefühlt habe. Besonders hervorheben muss ich Nikolaus und Nils, die durch ihre wunderbare Art so manches emotionales Auf und Ab aufgefangen haben. Dafür ein großes Dankeschön.

Während meiner Promotionszeit durfte ich mit einer Vielzahl an Studierenden zusammenarbeiten. Ich möchte mich herzlich bei Anna Steinbach, Anne Schmitt, Ektarina Kim, Franziska Ullrich, Florian Grieb, Kevin Nürck, Lena Mrowetz, Lennart Schorling, Maike Rothe und Tatiana Balashova bedanken. Die Betreuung eurer Abschlussarbeiten im Rahmen des Forschungsprojekts, aus dem meine Dissertation gewachsen ist, hat mich viel gelehrt und mir in den meisten Momenten viel Spaß gemacht.

Ein sehr großes Danke geht an meine Familie. Meinem Bruder verdanke ich vermutlich, dass ich die Kognitive Ergonomie als Fachrichtung ausgewählt

habe. Maurits Adam möchte ich danken für das offene Ohr, das aufmerksame Lesen und die vielen geteilten Stunden in Hyrule. Ohne Hans Trukenbrod hätte ich meine Promotion jedoch nicht in dieser Form abgeschlossen. Ich danke Dir für all die Dinge, die Du mir beigebracht hast und für Deine Geduld, wenn ich endlos über Argumentationsketten und Analysen sinniere. Meinen Eltern gehört jedoch der größte Dank. Sie haben immer den perfekten Weg zwischen Unterstützen und Machen lassen gefunden. Auf dem langen Weg schulischer und universitärer Bildung hätte ich mir keine besser Unterstützung wünschen können.

Zusammenfassung

Im Kontext von Mensch-Computer-Interaktionen wird die Interaktionsdauer häufig als objektiver Indikator für die Effizienz oder die Verständlichkeit eines Systems genutzt. Die Dauer wird jedoch auch vom Nutzer subjektiv wahrgenommen. Vor diesem Hintergrund ist es das Ziel der vorliegenden Arbeit zu untersuchen, (1) wie Nutzer Interaktionsdauern wahrnehmen und (2) wie diese Wahrnehmung mit anderen Erlebenskomponenten zusammenhängt. Dazu wurde im Rahmen von vier Studien Dauerwahrnehmung (operationalisiert durch prospektive Dauerschätzung, Dauerbewertung und Zeitflussbeurteilung) zusammen mit anderen Komponenten des Nutzererlebens betrachtet. In den ersten drei Studien absolvierten Nutzer ($N_1 = 61$, $N_2 = 80$, $N_3 = 33$) Aufgaben auf verschiedenen Webseiten. Die Ergebnisse zu Frage (1) zeigten, dass die Dauern meist überschätzt und von der Zweitaufgabe lautes Denken verzerrt wurden. Dies entsprach Vorhersagen kognitionspsychologischer Modelle zu prospektiven Dauerschätzungen. Dauerbewertungen und Zeitflussbeurteilungen wurden hingegen primär durch Aspekte des Systems und nicht der Aufgabe beeinflusst. Die Ergebnisse zu Frage (2) deuteten darauf hin, dass die Dimensionen der Dauerwahrnehmung mit verschiedenen Komponenten des Nutzererlebens kovariierten, vor allem mit der Wahrnehmung instrumenteller Qualitäten und der Valenz. In der vierten Studie ($N_4 = 60$) wurden Dauerbewertung und Zeitflussbeurteilung in einer Simulationsumgebung näher untersucht. Hier zeigte sich, dass die Dauerbewertung primär auf der Dauerschätzung beruht, jedoch auch vom emotionalen Zustand des Nutzers beeinflusst wird. Die Zeitflussbeurteilung wies hier deutliche Unterschiede zur Dauerschätzung und -bewertung auf. Zusammengefasst wird geschlussfolgert, dass (1) Nutzer die Dauer von Interaktionen wahrnehmen und bewerten. Die Bewertung erfolgt sowohl abschließend auf Basis der Gesamtdauer (Dauerbewertung) als auch auf Basis des erlebten Zeitflusses während der Interaktion (Zeitflussbeurteilung). Die korrelativen Zusammenhänge zu Komponenten des Nutzererlebens stützen zudem die Annahme, dass (2) die Wahrnehmung von Dauern, vor allem in Bezug auf die Dauerbewertung, ein integraler Bestandteil des Nutzererlebens ist und als Teil der Wahrnehmung der instrumentellen Qualitäten gesehen werden kann.

Schlagwörter: Nutzererleben, Zeitwahrnehmung, Dauerschätzung, Dauerbewertung, Passage of Time

Abstract

In the field of human-computer-interaction, the duration of the interaction is often used as an objective indicator for the usability of a system. The subjective perception of the duration is, however, mostly ignored. In the presented thesis, I ask (1) how users perceive the duration of an interaction and (2) how this perception correlates with the user experience (UX). In four studies, the UX was measured together with the user's perception of time in terms of an estimated duration, an overall assessment of the duration and a passage of time judgment. In studies 1 to 3 ($N_1 = 61$, $N_2 = 80$, $N_3 = 33$), users solved tasks on different web pages. The results regarding question (1) showed that the duration of the interaction was mostly overestimated. In addition, estimates were distorted when users performed a secondary task (think aloud). These results are in line with the predictions from cognitive models of prospective time perception. In contrast, the overall assessment and the passage of time judgments were primarily affected by the usability of the system and not by the task. The results regarding question (2) indicated that the perception of an interaction's duration covaried with the perception of instrumental qualities as well as the affective state of the user. Study 4 ($N_4 = 60$) focussed on the emergence of overall assessments and passage of time judgments. The results showed that the overall assessment of a duration is strongly based upon the estimated duration but also affected by the user's affective state. Passage of time judgments, however, showed no such effects indicating that they have to be dissociated from the other dimensions of duration perception. In conclusion, (1) when a user performs an interaction with a computer, he/she perceives and evaluates time on different dimensions. (2) These dimensions are an integral part of UX. Especially the overall assessment of the duration is argued to be part of the perception of instrumental qualities.

Keywords: user experience, time perception, estimated duration, overall assessment of a duration, passage of time

Inhaltsverzeichnis

I	Theoretischer Hintergrund	1
1	Einleitung	3
2	Nutzererleben in der Mensch-Computer-Interaktion	7
2.1	Reduktionistische Konzeptionen von UX	8
2.2	Zeitliche Dimensionen des Nutzererlebens	15
2.3	Zwischenfazit zum Nutzererleben	17
3	Erlebte Zeit - prospektive Dauerschätzungen	19
3.1	Modelle und Theorien zur erlebten Zeit	22
3.2	Befunde im prospektiven Paradigma	31
3.3	Zwischenfazit zur erlebten Zeit	42
4	Zeitempfinden	45
4.1	Dauerbewertungen	47
4.2	Zeitflussbeurteilungen	49
4.3	Zwischenfazit zum Zeitempfinden	53
5	Forschungsfragen und Überleitung	55
II	Empirische Arbeiten	59
6	Studie I: Effekte von System und Aufgabe auf Dauerwahrnehmung	61
6.1	Methode	63
6.2	Ergebnisse	75
6.3	Diskussion	82

7	Studie II: Validierung der Effekte von System und Aufgabe auf Dauerwahrnehmung	89
7.1	Methode	90
7.2	Ergebnisse	98
7.3	Diskussion	107
8	Studie III: Korrelationen zwischen Dauerwahrnehmung und UX bei "echten Interaktionen"	117
8.1	Methode	119
8.2	Ergebnisse	126
8.3	Diskussion	130
9	Studie IV: Effekte von System und Dauer auf Dauerbewertung und Zeitflussbeurteilung	137
9.1	Methode	139
9.2	Ergebnisse	148
9.3	Diskussion	156
III	Diskussion	163
10	Allgemeine Diskussion	165
10.1	Dauerwahrnehmung während Mensch-Computer-Interaktionen	167
10.2	Dauerwahrnehmung und Nutzererleben	179
10.3	Anwendungsaspekte der Ergebnisse	183
10.4	Limitationen	187
10.5	Fazit	192
IV	Anhang	195
	Literaturverzeichnis	197
A	Abkürzungen	219
B	Statistische Methoden	221
B.1	Gemischtes lineares Modell - LMM	222
B.2	Korrelationen bei Messwiederholung	233

C Zusammenfassung aller Hypothesen	237
D Anhang Studie I	241
D.1 Versuchsmaterialien	241
D.2 Poweranalysen	243
D.3 Hauptkomponentenanalysen zur Dauerbewertung	244
D.4 Korrelationen der UX-bezogenen Variablen	246
D.5 Analyse der UX-bezogene Variablen	247
E Anhang Studie II	249
E.1 Versuchsmaterialien	249
E.2 Bescheinigung der Ethikkommission	250
E.3 Poweranalysen	251
E.4 Hauptkomponentenanalysen zur Dauerbewertung	252
E.5 Korrelationen der UX-bezogenen Variablen	254
E.6 Analyse der UX-bezogene Variablen	255
F Anhang Studie III	257
F.1 Versuchsmaterialien	257
F.2 Bescheinigung der Ethikkommission	259
F.3 Poweranalysen	260
F.4 Korrelationen der UX-bezogenen Variablen	260
G Anhang Studie IV	261
G.1 Versuchsmaterialien	261
G.2 Bescheinigung der Ethikkommission	263
G.3 Poweranalysen	264
G.4 Korrelationen der UX-bezogenen Variablen	265

Abbildungsverzeichnis

2.1	CUE-Modell nach Minge und Thüring (2018)	14
3.1	Informations-Prozess-Modell der Dauerwahrnehmung nach Treisman (1963)	23
3.2	Scalar Timing Theory nach Gibbon et al. (1984)	26
3.3	Dual-Process Contingency Model nach Zakay (1993a)	29
3.4	Attentional Gate Model nach Zakay und Block (1995)	31
5.1	Operationalisierung Wahrnehmung von Interaktionsdauern	56
5.2	Übersicht über die Studien I bis IV	57
6.1	Studie I: Übersicht über den Versuchsablauf	65
6.2	Studie I: Webseiten	66
6.3	Studie I: Effekte der UVn auf Dauer, Dauerschätzung und PTRs	77
6.4	Studie I: Effekt der UV Usability auf Dauerbewertung	79
6.5	Studie I: Effekte der UVn auf Komponenten des Nutzererlebens	83
7.1	Studie II: Webseiten Aufgabe A	91
7.2	Studie II: Effekte der UVn auf Dauer, Dauerschätzung und PTRs	100
7.3	Studie II: Effekte der UVn auf das Zeitempfinden	102
7.4	Studie II: Effekte der UVn auf Komponenten des Nutzererlebens	106
8.1	Studie III: Übersicht über den Versuchsablauf	121
8.2	Studie III: Effekte der Aufgaben	127
9.1	Studie IV: Annahmen zu Einflüssen auf das Zeitempfinden	138
9.2	Studie IV: Interface des Heizungssystems	141
9.3	Studie IV: Übersicht über den Versuchsablauf	143
9.4	Studie IV: Veranschaulichung der Mediatoranalyse	146

9.5	Studie IV: Effekte der UVn auf Dauerschätzung und Komponenten des Nutzererlebens	150
9.6	Studie IV: Effekte der UVn sowie Dauerschätzung und Valenz auf Dauerbewertung	154
10.1	Studien I bis IV: Streudiagramm der Dauern und Dauerschätzungen	167
B.1	Kodierungsarten für zweistufige, nominalskalierte Prädiktoren	228
B.2	Unterschied zwischen Korrelationen basierend auf gemittelten Werten und Innersubjektkorrelationen	236
D.1	Studie I: Selbstentwickelte Skalen	242
E.1	Studie II: Bescheinigung Ethikkommission	250
F.1	Studie III: Selbstentwickelte zeitbezogene Skalen	258
F.2	Studie III: Bescheinigung Ethikkommission	259
G.1	Studie IV: Selbstentwickelte Skalen	262
G.2	Studie IV: Bescheinigung Ethikkommission	263

Tabellenverzeichnis

3.1	Arten zur Abgabe von Dauerschätzungen.	21
3.2	Studien zum Arousaleffekt im prospektiven Paradigma	34
6.1	Studie I: Korrelationshypothesen	70
6.2	Studie I: Effekte der UVn auf Dauer, Dauerschätzung und PTRs	76
6.3	Studie I: Komponentenladungen erster Messzeitpunkt nach Ex- ploration	78
6.4	Studie I: Effekte der UVn auf Dauerbewertung	80
6.5	Studie I: Innersubjektkorrelationen	81
7.1	Studie II: Aufgaben und Usability-Probleme	93
7.2	Studie II: Übersicht über AVn	94
7.3	Studie II: Korrelationshypothesen	97
7.4	Studie II: Effekte der UVn auf Dauer, Dauerschätzung und PTRs	99
7.5	Studie II: Effekte der UVn auf Dauerbewertung und Zeitfluss- beurteilung	103
7.6	Studie II: Innersubjektkorrelationen	104
7.7	Studien I und II: Zusammenfassung der Ergebnisse	109
8.1	Studie III: Inhalte der Aufgaben	122
8.2	Studie III: Effekt der Aufgaben auf wahrgenommene Usability, Aufgabendauer und Schätzverzerrungen	128
8.3	Studie III: Innersubjektkorrelationen	130
9.1	Studie IV: Effekte der UVn auf Dauerschätzung, wahrgenommene Usability, wahrgenommene Attraktivität und Valenz	149
9.2	Studie IV: Effekte der UVn auf Dauerschätzung und Zeitfluss- beurteilung sowie Mediatoranalysen	153

9.3	Studie IV: Innersubjekt- und reguläre Korrelationen	155
10.1	Vergleich Innersubjektkorrelationen zwischen Studien	169
D.1	Studie I: Ausbalancierung	241
D.2	Studie I: Poweranalysen	243
D.3	Studie I: Komponentenladungen zweiter und dritter Messzeitpunkt nach Exploration	244
D.4	Studie I: Komponentenladungen Aufgaben A, B und C	245
D.5	Studie I: Innersubjektkorrelationen UX-bezogene Variablen	246
D.6	Studie I: Effekte der UVn auf wahrgenommene Usability und wahrgenommene Attraktivität	247
D.7	Studie I: Effekte der UVn auf Valenz und Arousal	248
E.1	Studie II: Ausbalancierung	249
E.2	Studie II: Poweranalysen	251
E.3	Studie II: Komponentenladungen erster bis dritter Messzeitpunkt	252
E.4	Studie II: Komponentenladungen vierter Messzeitpunkt und Aufgaben A und B	253
E.5	Studie II: Komponentenladungen Aufgaben C und D	254
E.6	Studie II: Innersubjektkorrelationen UX-bezogene Variablen	254
E.7	Studie II: Effekte der UVn auf wahrgenommene Usability und wahrgenommene Attraktivität	255
E.8	Studie II: Effekte der UVn auf Valenz und Arousal	256
F.1	Studie III: Häufigkeiten der Aufgabenreihenfolge	257
F.2	Studie III: Poweranalysen	260
F.3	Studie III: Innersubjektkorrelationen UX-bezogene Variablen	260
G.1	Studie IV: Ausbalancierung	261
G.2	Studie IV: Poweranalysen	264
G.3	Studie IV: Innersubjektkorrelationen UX-bezogene Variablen	265

Teil I

Theoretischer Hintergrund: Nutzererleben und Dauerwahrnehmung

Kapitel 1

Einleitung

Zeit spielt in unserem Leben eine wichtige Rolle. Sie bestimmt Teile unserer Handlungen, unserer Handlungsabfolgen und ist eine Voraussetzung für Erinnerung und Planung. Sie dient als Orientierungsmittel (Elias, 1988) oder im Falle von Zeitdauern auch als Signal, dass etwas nicht wie geplant verläuft (Wittmann, 2013). Doch Zeit ist nicht nur eine messbare, physikalische Größe. Sie wird vom Menschen subjektiv wahrgenommen und erlebt. Beispiele hierfür sind Dauerschätzungen, Erwartungen, wann ein Ereignis eintreten sollte, aber auch das Gefühl nicht endender Minuten in einer Wartesituation oder das Erleben von verfliegender Zeit, die mit guten Freunden verbracht wird. Zeit- und Dauerwahrnehmung ist also abzugrenzen von Zeit als messbare Größe. Sie kann nah an der objektiven Zeit sein, ist manchmal aber auch stark verzerrt. Sie ist ein wichtiger Teil unseres Erlebens und kann Einfluss auf Bewertungen, Entscheidungen und das Verhalten haben (Ariely & Zakay, 2001).

Im Kontext der Mensch-Computer-Interaktion werden Dauern meist nur objektiv betrachtet. Die Zeit, die ein Nutzer benötigt, um eine Aufgabe mit einem Computer erfolgreich abzuschließen, wird häufig als Indikator für die Effizienz des Systems interpretiert (Bevan, 2009; Albert, Tullis & Tedesco, 2009; Frøkjær, Hertzum & Hornbæk, 2000; Lesemann, Woletz & Koerber, 2007). Dabei wird Effizienz als eine Kerndimension von Gebrauchstauglichkeit oder Usability verstanden (DIN, EN ISO 9241-11, 2018; Nielsen, 1994, 2012). Aber nicht nur auf die Effizienz wird auf Basis der Aufgabendauer geschlossen. Andere Autoren beschreiben, dass Bearbeitungszeiten ein Indiz für die Verständlichkeit und Intuitivität der Benutzung (Blackler, Popovic & Mahar, 2010) oder

für die Selbstbeschreibungsfähigkeit und Erwartungskonformität des Systems (DIN, EN ISO 9241-11, 2018) sind.

Diese Interpretation von Bearbeitungszeiten erfolgt im Rahmen von Usability- oder UX-Tests auf Basis objektiver Messwerte und wird vom Testleiter durchgeführt. Nutzern stehen die Bearbeitungszeiten jedoch auch zur Verfügung. Auch wenn sie meist keine genaue, objektive Zeitdauer gemessen haben, so haben Nutzer ein Gespür dafür, wie lange eine Interaktion gedauert hat und wie sie diese Dauer bewerten. Eine solche subjektive Dauerwahrnehmung kann in das allgemeine Erleben der Interaktion integriert werden, das sogenannte *Nutzererleben* oder auch *User Experience* (UX) genannt. Sie ist allerdings kein exaktes Abbild der tatsächlichen Dauern, sondern unterliegt dem Einfluss verschiedener Faktoren. Welche Faktoren dies sind und wie sie wirken, kann aus kognitionspsychologischen Forschungsarbeiten und Modellen abgeleitet werden.

Vor dem Hintergrund dieser Grundlagen ist es das Ziel der vorliegenden Arbeit, den Einfluss von Interaktionsdauern auf das Nutzererleben besser zu verstehen, um verlässliche und fundierte Vorhersagen über die Bedeutung von Dauern und Dauerwahrnehmung während einer Mensch-Computer-Interaktion machen zu können.

Hierzu gehört auf der einen Seite die Untersuchung der Dauerwahrnehmung während einer Interaktion. Mögliche Fragestellungen können sich darauf beziehen, ob und wann es zu systematischen Verzerrungen in der Dauerwahrnehmung während einer Mensch-Computer-Interaktion kommt und ob diese mit kognitionspsychologischen Grundlagenbefunden übereinstimmen. Auf der anderen Seite stellt sich die Frage, welche Zusammenhänge es zwischen verschiedenen Dimensionen der Dauerwahrnehmung und des Nutzererlebens gibt. Beide Teilaspekte des Forschungsziels sind an der Schnittstelle zwischen Forschung zum Nutzererleben und kognitionspsychologischer Forschung zur Wahrnehmung von Dauern anzusiedeln. Im theoretischen Hintergrund werden daher beide Forschungsbereiche trotz geringer Überschneidungen behandelt und bilden die Grundlage für die Bearbeitung des Forschungsziels.

Im Fokus der vorliegenden Arbeit stehen einzelne Interaktionen zwischen Nutzern und Computern mit Dauern von wenigen Minuten, in denen Wartezeiten oder Systemverzögerungen (*delays*) keine große Rolle spielen. Forschung

zu den Auswirkungen von Warte- oder Latenzzeiten auf das Nutzererleben gibt es bereits weit mehr (z.B. Hoxmeier & DiCesare, 2000; Sears, Jacko & Borella, 1997; Szameitat, Rummel, Szameitat & Sterr, 2009) als Forschung zum Effekt von Interaktionsdauern. Daher steht Letzteres im Zentrum dieser Arbeit.

Darüber hinaus wird auf Interaktionen fokussiert, in denen die Dauer eine kritische Komponente der Interaktion ist. Beispiele hierfür sind das Suchen von Informationen auf einer Webseite, das Kaufen einer Fahrkarte in einem Online-Shop oder das Bearbeiten eines Dokuments mit einem Textverarbeitungsprogramm. In den beschriebenen Interaktionen hat die Zeitdauer bis zum erfolgreichen Abschluss der Interaktion eine hohe Bedeutung, da die Interaktion primär eine schnelle und einfache Zielerreichung erlauben soll. Zakay (2015) argumentiert, dass eine solche hohe temporale Relevanz zu einer verstärkten Zuwendung von Aufmerksamkeit auf das Verstreichen der Zeit führt, wodurch Dauern *prospektiv* wahrgenommen werden. Dies ist abzugrenzen von einer *retrospektiven* Dauerschätzung, bei der keinerlei Aufmerksamkeit auf das Verstreichen der Zeit während des Erlebens des Intervalls gerichtet wird. Daher steht das Paradigma der prospektiven Dauerwahrnehmung im Fokus der vorliegenden Arbeit.

Im Rahmen der Forschung zur Dauerwahrnehmung hat sich in den letzten Jahren jedoch noch eine weitere Betrachtungsweise in der psychologischen Forschung etabliert: *Zeitempfinden* (*feelings about time*, Wearden, 2015, S. 165). Wearden, O'Donoghue, Ogden und Montgomery (2014) beschreiben dieses Empfinden anhand einer Wartesituation: Auch wenn dem Wartenden¹ bewusst ist, dass er nur fünf Minuten gewartet hat, kann er trotzdem äußern, dass es sich viel länger angefühlt hat. Zeitempfinden kann aber auch während einer Wartedauer entstehen, wenn der Wartende beispielsweise den Minutenzeiger seiner Uhr betrachtet und das Gefühl hat, dass sich jede Minute unendlich lange zieht. Diese Art des Zeitempfindens beschreibt keine Bewertung der Dauer in Bezug auf ihre Länge, sondern ein Gefühl über das Fließen bzw. Vergehen der Zeit (*passage of time*).

In der Mensch-Computer-Interaktion kann das Zeitempfinden als ein Emp-

¹Zur besseren Lesbarkeit wird auf die Verwendung der männlichen und weiblichen Sprachform verzichtet. Sämtliche Personenbezeichnungen gelten gleichwohl für die weibliche und männliche Form (generisches Maskulinum).

finden über die Interaktionsdauer interpretiert werden. Auch wenn es naheliegend scheint, eine zeitliche Erlebenskomponente im Rahmen des Nutzererlebens zu betrachten (Hassenzahl, 2004), gibt es bis jetzt wenig Forschung hierzu (Liikkanen & Gómez Gómez, 2013). Um diese Lücke zu schließen, soll in der vorliegenden Arbeit die Relevanz der Dauerwahrnehmung eines Nutzers für das Entstehen eines spezifischen Nutzererlebens theoretisch erarbeitet und empirisch untersucht werden. Der Forschungsgegenstand Dauerwahrnehmung umfasst dabei die prospektive Dauerwahrnehmung im Sinne von Dauerschätzungen als auch das Zeitempfinden im Sinne einer Dauerbewertung und einer Zeitflussbeurteilung.

Kapitel 2

Nutzererleben in der Mensch-Computer-Interaktion

Nutzererleben oder auch User Experience (UX) genannt, beschreibt das Erleben des Nutzers während einer Interaktion mit einem technischen Gerät oder einer technischen Oberfläche. Diese Umschreibung von UX wirkt zunächst sehr schwammig, da der Begriff Erleben schwer zu definieren ist. Diese Unklarheit spiegelt sich auch in der Vielzahl der Definitionen von UX wider (Lallemant, Gronier & Koenig, 2015; Law, Roto, Hassenzahl, Vermeeren & Kort, 2009). Viele dieser Definitionen haben jedoch gemeinsam, dass sie die Subjektivität und die Ganzheitlichkeit von UX hervorheben (Hassenzahl, Burmester & Koller, 2008). Was bedeutet das?

Die Subjektivität von UX besteht darin, dass das Erleben des Nutzers in den Mittelpunkt gestellt wird (Hassenzahl, 2003, 2008; Hassenzahl, Eckoldt & Thielsch, 2009; Hassenzahl et al., 2008; Law et al., 2009). Dieses ist per se subjektiv, da es nicht objektiv gemessen werden kann, sondern nur durch die Befragung des Nutzers zugänglich wird. Dies findet sich auch in verschiedenen Definitionen zu UX wieder, in denen UX beispielsweise als „Wahrnehmungen und Reaktionen einer Person, die aus der tatsächlichen und/oder der erwarteten Benutzung eines Produkts, eines Systems oder einer Dienstleistung resultieren“ (DIN, EN ISO 9241-11, 2018, S. 7) oder als „momentary, primarily evaluative feeling (good-bad) while interacting with a product and service“ (Hassenzahl, 2008, S. 12) definiert wird. Mit anderen Worten steht bei UX die persönliche Wahrnehmung und das Empfinden im Zentrum des Konstrukts,

wobei Wahrnehmung hier nicht im engen Sinne zu verstehen ist, sondern auch höhere kognitive Prozesse wie Bewerten oder Kategorisieren beinhaltet (Minge & Thüring, 2018).

Ganzheitlich bezieht sich auf die Dimensionen dieser persönlichen Wahrnehmung bzw. des Empfindens. Die DIN, EN ISO 9241-11 (2018) versteht unter diesen Dimensionen „sämtliche Emotionen, Vorstellungen, Vorlieben, Wahrnehmungen, physiologischen und psychologischen Reaktionen, Verhaltensweisen und Leistungen, die sich vor, während und nach der Nutzung ergeben“ (DIN, EN ISO 9241-11, 2018, S. 7). Mit dieser Auflistung wird deutlich, wie umfangreich UX in seinen Teilkomponenten verstanden werden kann. Eine solch holistische Betrachtung von UX ist im Rahmen der Eingrenzung des Konstrukts jedoch nur mäßig zielführend, da die Definition weder spezifiziert, was unter den einzelnen Komponenten zu verstehen ist (Hassenzahl, 2008), noch Schwerpunkte setzt. Daher wird im folgenden auf reduktionistische Konzeptionen von UX eingegangen.

2.1 Reduktionistische Konzeptionen von UX

Im Rahmen einer Strukturierung der Fülle an potentiellen Dimensionen oder Teilaspekten sind verschiedene, reduktionistische Konzeptionen von UX entstanden. Diese benennen wesentliche Komponenten von UX und beschreiben Zusammenhänge zwischen den Komponenten. Ziel dieser Herangehensweise ist es, auf Basis verschiedener Antezedenzen das Erleben während der Interaktion sowie die Konsequenzen der Interaktion erklären und vorhersagen zu können. Zu den Antezedenzen zählen das System, der Nutzer und der Nutzungskontext. Zu den Konsequenzen gehören beispielsweise das zukünftige Nutzungsverhalten oder die Gesamtbewertung des Produkts.

Für das hier vorliegende Forschungsziel bieten reduktionistische Konzeptionen von UX mehrere Vorteile. Zum einen erlauben sie es, das Erleben eines Nutzers während einer Interaktion auf Komponenten herunterzubrechen. Dies erleichtert sowohl die Messung von UX als auch die Untersuchung des Zusammenhangs zwischen der Wahrnehmung von Interaktionsdauern und einzelnen Teilaspekten von UX. Zum anderen erfolgt im Rahmen von reduktionistischen Konzeptionen eine systematische Auseinandersetzung mit Antezedenzen und

Konsequenzen von UX. Diese Systematik kann als Rahmen für die Betrachtung von Interaktionsdauern und somit für die Bearbeitung des vorliegenden Forschungsziels verstanden werden.

2.1.1 Komponenten des Nutzererlebens

Zwei wesentliche Teilaspekte, die von vielen Autoren im Rahmen der reduktionistischen Sichtweise auf UX beschrieben werden, sind die Wahrnehmung von *pragmatischen* bzw. *instrumentellen* Qualitäten der Interaktion und die Wahrnehmung von *hedonischen* bzw. *nicht-instrumentellen* Qualitäten der Interaktion (Bargas-Avila & Hornbæk, 2011; Hassenzahl, 2003, 2008; Mahlke, 2008; Minge, 2008; Minge & Thüring, 2018; van Schaik, Hassenzahl & Ling, 2012; Thüring & Mahlke, 2007). Diese beiden Komponenten werden von manchen Autoren durch eine affektive Komponente ergänzt (z.B. Minge, 2008; Thüring & Mahlke, 2007).

Instrumentelle (pragmatische) Qualitäten der Interaktion beschreiben aufgabenbezogene Qualitäten, die den Nutzer darin unterstützen, bestimmte Aufgaben erfolgreich zu meistern (Aranyi & van Schaik, 2015; Hassenzahl, 2003) und sogenannte „do-goals“ zu erfüllen (Hassenzahl, 2008). Zu diesen Qualitäten gehören die Kontrollierbarkeit (Hassenzahl, 2003; Thüring & Mahlke, 2007), die Nützlichkeit (Hassenzahl, 2003; Minge, Thüring & Wagner, 2016) oder die Usability (Hassenzahl, 2003; Thüring & Mahlke, 2007) des technischen Gerätes, mit dem interagiert wird. Diese Aspekte werden durch den Nutzer wahrgenommen und führen zu einem subjektiven Eindruck der instrumentellen Qualität.

Wie eingangs diskutiert, wird die Dauer einer erfolgreich abgeschlossenen Interaktion häufig objektiv gemessen und als Indikator für die Effizienz des Systems (z.B. Bevan, 2009) oder seine Verständlichkeit (Blackler et al., 2010) genutzt. Hier werden also instrumentelle Qualitäten der Interaktion auf Grund der objektiven Dauer erschlossen. Für das Nutzererlebens sind aber nicht die objektiv beschreibbaren Qualitäten von Bedeutung, sondern die subjektiv wahrgenommenen. Analog dazu steht im Zentrum der vorliegenden Arbeit nicht die objektive Dauer von Interaktionen, sondern ihre Wahrnehmung. Dies zeigt einen ersten Anknüpfungspunkt des vorliegenden Forschungsziels an das Konstrukt Nutzererleben.

Neben der Wahrnehmung instrumenteller Qualitäten nimmt der Nutzer je-

doch auch nicht-instrumentelle (hedonische) Aspekte der Interaktion und des Produkts wahr. Hierzu gehören beispielsweise die Schönheit (Minge, 2008), visuelle Ästhetik (Lavie & Tractinsky, 2004; Minge, 2008; Minge & Thüring, 2018; Thüring & Mahlke, 2007) und haptische Qualität des Produkts (Thüring & Mahlke, 2007) sowie die Identifikation des Nutzers mit dem Produkt (Hassenzahl, 2003; Minge, 2008) oder die Stimulation, die durch das Produkt oder die Interaktion hervorgerufen werden (Hassenzahl, 2003; Minge, 2008). Die Wahrnehmung dieser Produkteigenschaften sind für die auszuführende Aufgabe in der Regel irrelevant, sie dienen vielmehr einer psychologisch-hedonischen Befriedigung des Nutzers (Hassenzahl, 2003) oder der Erfüllung sogenannter „be-goals“ (Hassenzahl, 2008).

Eine weitere Dimension von UX beschreibt die Emotionen und die affektiven Reaktionen des Nutzers während, vor oder nach der Interaktion (Agarwal & Meyer, 2009; Hudlicka, 2003; Mahlke, 2008; Minge, 2008; Thüring & Mahlke, 2007). Emotionen entstehen als Reaktionen auf einen Stimulus oder seine Bewertung und zeigen sich zeitlich begrenzt und synchron auf verschiedenen Dimensionen (Scherer, 2005). Zu diesen Dimensionen gehören körperliche Veränderungen bezogen auf den motorischen Ausdruck und neurophysiologische Komponenten, sowie Handlungstendenzen, kognitive Bewertungen und subjektive Gefühle (Scherer, 2005). Während einer Interaktion mit technischen Produkten können Emotionen von unterschiedlicher *Valenz* sein (z.B. negativ als Frustration oder positiv als Freude) und sich in ihrem Erregungsniveau (*Arousal*) unterscheiden (Vergleich Core Affect, Russell & Barrett, 1999 oder Russell, 2003).

Zusammengefasst werden in reduktionistischen Konzeptionen von UX häufig drei verschiedene Komponenten von UX genannt: die Wahrnehmung instrumenteller Qualitäten, die Wahrnehmung nicht-instrumenteller Qualitäten und die erlebte Emotion. Bei der Interaktion mit einem technischen System entsteht auf allen drei Komponenten eine situations-, interaktions- und personenabhängige Ausprägung, was ein spezifisches Nutzererleben prägt. Dabei stellt sich jedoch die Frage, in wie weit die drei hier vorgestellten Komponenten miteinander interagieren und sich gegenseitig beeinflussen.

2.1.2 Zusammenhänge zwischen den Komponenten des Nutzererlebens

Auch wenn verschiedene Autoren (Hassenzahl, 2003, S. 6; Mahlke, 2008, S. 126; van Schaik et al., 2012, S. 19) teilweise auf Basis empirischer Arbeiten argumentieren, dass instrumentelle und nicht-instrumentelle Qualitäten voneinander unabhängig und unterscheidbar sind, zeigen andere Studien Korrelationen zwischen Maßen dieser beiden Konstrukte. Aranyi und van Schaik (2015) berichteten beispielsweise eine Korrelation von $r = .63$ zwischen instrumentellen und nicht-instrumentellen Qualitäten. Analog hierzu zeigten Lavie und Tractinsky (2004) mittlere bis hohe Korrelationen zwischen der wahrgenommenen Usability und verschiedenen Ästhetik-Dimensionen ($r = .40$ bis $r = .78$). Auch Minge (2008) berichteten mittlere Korrelationen ($r = .37$ bis $r = .53$) zwischen wahrgenommener Usability und wahrgenommener Attraktivität. Diese Zusammenhänge deuten darauf hin, dass instrumentelle und nicht-instrumentelle Qualitäten nicht unabhängig voneinander wahrgenommen werden.

Die Emotion steht ebenfalls in Wechselwirkung mit den andern Komponenten. So zeigten Lavie und Tractinsky (2004) mittlere bis hohe Korrelationen ($r = .44$ bis $r = .68$) zwischen Pleasure und Usability bzw. verschiedenen Ästhetik-Dimensionen. Aranyi und van Schaik (2015) berichteten vergleichbar dazu hohe Korrelationen zwischen positivem Affekt und pragmatischen bzw. hedonischen Qualitäten ($r = .54$ bzw. $r = .67$).

Die beschriebenen Befunde sprechen dafür, dass auf der einen Seite die Unterscheidung zwischen den drei verschiedenen Komponenten des Nutzererlebens sinnvoll ist, da diese Komponenten nicht deckungsgleich sind. Auf der anderen Seite sind die Komponenten jedoch nicht vollkommen unabhängig voneinander, da sie deutlich miteinander kovariieren.

2.1.3 Determinanten und Konsequenzen der Komponenten des Nutzererlebens

Neben den Bestandteilen von UX stellen sich zwei weitere wesentliche Fragen: Erstens, was sind Einflussfaktoren, die auf die Komponenten des Nutzererlebens wirken und so zu einem spezifischen Nutzererlebnis führen? Zweitens, was resultiert aus dem Nutzererleben?

Gemäß der genannten Definitionen (DIN, EN ISO 9241-11, 2018; Hassenzahl, 2008) resultiert Nutzererleben aus der Interaktion oder der erwarteten Interaktion mit einem Produkt. Dabei spielen also die Charakteristiken des Produkts bezüglich des Inhalts, der Gestaltung, der Funktionalität und des Interaktionstyps eine wesentliche Rolle (Hassenzahl, 2003; Mahlke, 2008) und sollten einen Einfluss auf das Entstehen eines spezifischen Nutzererlebnisses haben. Erwartungskonform zeigen viele Studien, dass eine Manipulation der Gebrauchstauglichkeit des Systems zu einer Veränderung der Wahrnehmung der instrumentellen Qualitäten führt (Ben-Bassat, Meyer & Tractinsky, 2006; Hamborg, Hülsmann & Kaspar, 2014; Minge & Thüring, 2018; Thüring & Mahlke, 2007; van der Geest & van Dongelen, 2009). Umgekehrt gibt es auch viel Evidenz dafür, dass eine Manipulation des visuellen Erscheinungsbildes des Systems einen Effekt auf die Wahrnehmung der nicht-instrumentellen Qualitäten hat (Ben-Bassat et al., 2006; Hamborg et al., 2014; Minge & Thüring, 2018; Sonderegger & Sauer, 2010; Thüring & Mahlke, 2007; Tractinsky, Katz & Ikar, 2000).

Es gibt jedoch auch Wechselwirkungen und sogenannte „Halo-Effekte“ zwischen den Komponenten des Nutzererlebens (Minge, 2011; Minge & Thüring, 2018). Minge und Thüring (2018) konnten beispielsweise zeigen, dass sich das visuelle Erscheinungsbild vor der Interaktion auf die erwartete Usability des Systems auswirkt, nicht aber nach einer Interaktion. Zu letzterem Punkt ist die Befundlage jedoch nicht eindeutig. So zeigten andere Autoren, dass sich der Effekt des Erscheinungsbildes auf die wahrgenommene Usability auch noch nach einer Interaktion zeigen kann (Ben-Bassat et al., 2006; Sonderegger & Sauer, 2010; Tractinsky et al., 2000). Frei nach dem Ausdruck „What is beautiful is usable“ (Tractinsky et al., 2000) scheinen gestalterische Aspekte eines Systems also auch die wahrgenommene Usability zu beeinflussen. Gleichzeitig konnten mehrere Autoren zeigen, dass auch die umgekehrte Beeinflussung („What is usable is beautiful“) existiert (Ben-Bassat et al., 2006; Hamborg et al., 2014; Minge & Thüring, 2018).

Neben den System- oder Produkteigenschaften ist für das Nutzererleben jedoch auch von Bedeutung, wer die Interaktion erlebt. Eigenschaften des Nutzers wie Vorerfahrung, Erwartungen, aber auch Bedürfnisse und Attributionsstile können sowohl die Interaktion an sich als auch deren Wahrnehmung

bestimmen. Beispielsweise kann das gleiche System bei zwei Nutzern zu einem sehr unterschiedlichen Erleben führen, wenn sich die Nutzer in Bezug auf ihre do-goals und be-goals unterscheiden (Hassenzahl, 2008). Mahlke (2008) argumentiert darüber hinaus, dass neben den Systemeigenschaften und den Nutzercharakteristika auch Aspekte des Kontexts einen Einfluss auf die Entstehung eines spezifischen Nutzererlebnisses haben. Hierzu gehören die Aufgabe, aber auch die äußeren Bedingungen der Nutzung. Mahlke (2008) grenzt solche Nutzer- und Kontextcharakteristiken von den Systemeigenschaften jedoch ab, da Eigenschaften des Systems direkter auf das Nutzererleben wirken.

Bezogen auf die Frage nach den Konsequenzen, argumentiert Mahlke (2008), dass sich das Nutzererleben auf die allgemeine Bewertung, wie beispielsweise die Akzeptanz des Produkts, die Wahl eines Systems bei verschiedenen Alternativen sowie auf das Nutzungsverhalten und die Nutzungsintentionen auswirkt. Andere Konsequenzen können auch Zufriedenheit oder Vergnügen sein (Hassenzahl, 2003). Insgesamt gibt es also eine Bandbreite an Reaktionen und Konsequenzen, die aus einer spezifischen Ausprägung des Nutzererlebens resultieren können. Sie entstehen jedoch nicht durch eine mathematische Gewichtung der verschiedenen Erlebiskomponenten, sondern auf Basis einfacher Regeln und Heuristiken, die basierend auf der Wahrnehmung der Komponenten angewandt werden (van Schaik et al., 2012).

2.1.4 CUE-Modell: Components of User Experience

Das *Components of User Experience Model* (CUE-Modell) fasst die eben beschriebenen Komponenten des Nutzererlebens und deren Antezedenzen und Konsequenzen in einem Rahmenmodell zusammen (siehe Abb. 2.1, Mahlke, 2008; Minge & Thüring, 2018; Thüring & Mahlke, 2007). Es postuliert, dass auf Basis des Systems, des Nutzers und des Nutzungskontextes Interaktionsmerkmale bzw. -charakteristika entstehen, die der Nutzer in Bezug auf seine instrumentellen und nicht-instrumentellen Qualitäten wahrnimmt. Thüring (2013, S. 115) gibt als Beispiel für ein solches Interaktionscharakteristikum systeminhärente Latenzen. Diese werden wahrgenommen und auf Grund der Vorerfahrungen mit ähnlichen Systemen bewertet (z.B. als lange). Dabei kumulieren sich Bewertungen kleiner Teilschritte in der Interaktion über die gesamte Interaktionslänge auf und resultieren in der Wahrnehmung der instrumentellen

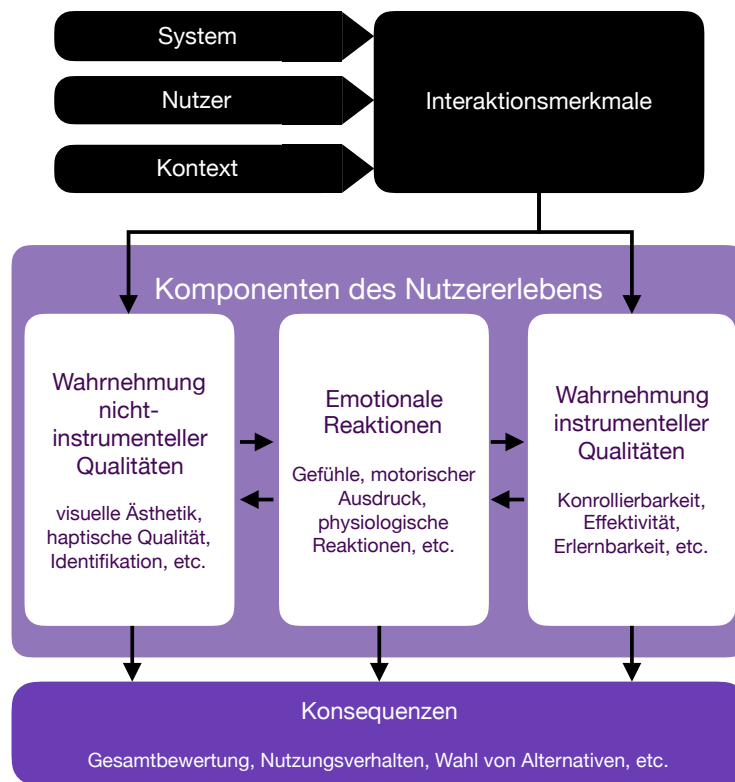


ABBILDUNG 2.1: Components of User Experience Model (CUE-Modell), Eigendarstellung nach Minge und Thüring (2018).

Qualitäten des Systems. Ein ähnlicher Prozess in Bezug auf nicht aufgabenrelevante Aspekte führt gleichzeitig zur Wahrnehmung der nicht-instrumentellen Qualitäten. Ein Beispiel hierfür könnte das Wahrnehmen der Farb- und Liniengestaltung sein, das die Wahrnehmung der visuellen Ästhetik beeinflussen kann.

Laut CUE-Modell beeinflussen sowohl die Wahrnehmung der instrumentellen und der nicht-instrumentellen Qualitäten den emotionalen Zustand des Nutzers (Mahlke, 2008; Thüring & Mahlke, 2007). Gleichzeitig können die erlebten Emotionen während einer Interaktion aber auch auf die Wahrnehmung der instrumentellen und der nicht-instrumentellen Qualitäten einen Einfluss haben (Minge et al., 2016; Minge & Thüring, 2018). Aus dem Nutzererleben folgen die eben diskutierten Konsequenzen, wie das Gesamturteil, das zukünftige Nutzungsverhalten oder die Nutzungsintention.

Zusammengefasst bietet das CUE-Modell ein Rahmenmodell für die Untersuchung des Nutzererlebens, welches im folgenden als Grundlage für die

Beschreibung und Operationalisierung von UX genutzt wird und in welches die Wahrnehmung von Dauern eingepasst werden soll. Dabei wird zunächst an den Interaktionscharakteristika angesetzt, da die Dauer einer Interaktion als ein solches Charakteristikum verstanden wird. Sie sollte durch den Nutzer, das System und den Kontext beeinflusst werden. Evidenz hierfür liefern beispielsweise Studien mit verschiedenen Altersgruppen, die zeigen, dass ältere Nutzer für die gleiche Aufgabe längere Bearbeitungsauern benötigen (z.B. Ziefle & Bay, 2005). Auch Systeme, die sich darin unterscheiden, wie viele Eingaben für die Erreichung eines bestimmten Interaktionsziels notwendig sind, führen zu verschiedenen Interaktionsauern (Ziefle & Bay, 2005).

Die Beschreibung der Interaktionscharakteristika im CUE-Modell ist jedoch relativ unspezifisch. Es wird hierfür weder eine konkrete Definition noch eine Auflistung von verschiedenen Charakteristika gegeben (Mahlke, 2008; Ming & Thüning, 2018; Thüning & Mahlke, 2007), wie es beispielsweise bei der Wahrnehmung der instrumentellen Qualitäten der Fall ist. Daher zeigt die vorliegende Arbeit beispielhaft anhand des ausgewählten Charakteristikums Interaktionsdauer, wie das CUE-Modell durch die Verknüpfung mit Wissen aus dem Grundlagenbereich und empirischen Untersuchungen konkretisiert und erweitert werden kann.

2.2 Zeitliche Dimensionen des Nutzererlebens

Welche Rolle die Wahrnehmung von zeitlichen Aspekten für das Entstehen des Nutzererlebens spielt, ist nicht explizit im CUE-Modell zu finden. Diese Frage kann jedoch bezogen auf verschiedene Dimensionen gestellt werden. Hierzu gehört zum einen die Betrachtung einzelner, zeitbezogener Elemente einer Interaktion, wie zum Beispiel Interaktionsauern, Systemverzögerungen (*delays*) oder Wartesituationen, sowie deren Wahrnehmung und Bedeutung für das Entstehen von UX. Auf der anderen Seite können Veränderungen im Nutzererleben über einen längeren Zeitraum hinweg betrachtet werden.

Letzteres kann mit Hilfe des CUE-Modell am schlechtesten beschrieben werden, da in diesem auf eine einzelne Interaktion fokussiert wird. Zum Nutzererleben über einen langen Zeitraum hinweg gibt es jedoch andere modellbildende Arbeiten, die das CUE-Modell erweitern (z.B. ContinUE-Modell von

Pohlmeyer, Hecht & Blessing, 2009). Darüber hinaus kann hier auf empirische Befunde (z.B. Minge & Thüring, 2018; Walsh et al., 2014; von Wilamowitz-Moellendorff, Hassenzahl & Platz, 2006) und Messmethoden (z.B. Karapanos, Zimmerman, Forlizzi & Martens, 2010; Kujala, Roto, Väänänen-Vainio-Mattila, Karapanos & Sinnelä, 2011; Kujala, Roto, Väänänen-Vainio-Mattila & Sinnelä, 2011) verwiesen werden, die sich mit der dynamischen Veränderung von UX über die Zeit und die wiederholte Nutzung eines Systems beschäftigen.

Die Betrachtung einzelner, zeitbezogener Elemente einer Interaktion kann durch das CUE-Modell im Rahmen der Interaktionscharakteristika besser beschrieben werden. Hierzu gibt es zudem eine Reihe an Untersuchungen zu Systemverzögerungen und Wartedauern bei technischen Interaktionen. Hier standen bis jetzt vor allem negative Konsequenzen von Systemverzögerungen für das Nutzererleben im Fokus (z.B. Hoxmeier & DiCesare, 2000; Sears et al., 1997; Szameitat et al., 2009). Manche dieser Ansätze berücksichtigen dabei nicht nur objektive Systemlatenzen oder Warteintervalle, sondern auch die Dauerwahrnehmung des Nutzers und die Bedeutung dieser Wahrnehmung für die Gestaltung von Wartezeiten (Gorn, Chattopadhyay, Sengupta & Tripathi, 2004; Kurusathianpong & Tangmanee, 2018; Lee, Chen & Ilie, 2012; Seow, 2008). In solchen Arbeiten wird unter Dauerwahrnehmung meist eine Bewertung bezüglich der wahrgenommenen Schnelligkeit des erlebten Zeitflusses (Gorn et al., 2004) oder bezüglich der Attribute langweilig, irritierend oder unakzeptabel (Kurusathianpong & Tangmanee, 2018) verstanden.

In diesem Sinne fordern Liikkanen und Gómez Gómez (2013), dass subjektiv erlebte Zeit (*subjectively experienced time*) bei der Interaktion mit einem technischen System eine weit höhere Bedeutung zugemessen werden sollte als der objektiven Messung von Systemreaktionszeiten (*system response time*). Die Autoren argumentieren, dass das Erleben von Zeit als ein Mediator für das Entstehen von UX dienen könnte. Dabei verstehen sie unter subjektiv erlebter Zeit eine Bewertung der erlebten Dauer, die durch Erfahrungen und den aktuellen Zustand der Person beeinflusst wird (Liikkanen & Gómez Gómez, 2013). So kann beispielsweise das Wissen darüber, warum bei einem Prozess eine Verzögerung auftritt, den emotionalen Zustand und die subjektiv erlebte Dauer beeinflussen. Die Autoren grenzen diese subjektiv erlebte Dauer von der Wahrnehmung von Reaktionszeiten des Systems ab (*subjective response time*),

da letzteres eher einer Dauerschätzung als einer Dauerbewertung entspricht.

In der vorliegenden Arbeiten stehen Interaktionsdauern im Fokus. Diese werden weder von Liikkanen und Gómez Gómez (2013), noch von anderen oben zitierten Autoren betrachtet. Die Wahrnehmung von Interaktionsdauern kann aber genauso zur subjektiv erlebten Zeit beitragen, wie die Wahrnehmung einer Systemverzögerung. So kann ein Nutzer eine Vorstellung davon haben, wie lange es dauern sollte, eine bestimmte Interaktion erfolgreich zu beenden. Das Ändern der Schriftgröße eines Texts in einem Textverarbeitungsprogramm darf beispielsweise nicht länger als zehn Sekunden dauern. Diese Erwartung kann verletzt werden, wenn relevante Interaktionselemente so ungünstig platziert sind, dass die Suche nach ihnen die Interaktionsdauer verlängert. Die Messung des zeitlichen Erlebens der Dauer könnte in diesem Kontext also eine sinnvolle Ergänzung innerhalb der üblichen Komponenten des Nutzererlebens sein.

Zusammengefasst kann UX sowohl als ein Konstrukt gesehen werden, welches über die Zeit hinweg entsteht und sich verändert, als auch als ein Konstrukt, das durch das Erleben von Zeit geformt wird. Die bestehende Forschung zu dieser zweiten Betrachtungsweise konzentriert sich meist auf Warteintervalle. In der vorliegenden Arbeit liegt der Fokus jedoch auf dem Erleben und Wahrnehmen von Zeitdauern einzelner Interaktionen, in denen Wartezeiten keine große Rolle spielen. Hierzu gibt es bis jetzt kaum experimentelle Arbeiten und keine standardisierten Operationalisierungen (Liikkanen & Gómez Gómez, 2013). Das Erleben und Wahrnehmen von Dauern während einzelner Interaktionen kann jedoch mit Konstrukten und Modellen aus der Grundlagenforschung zum Thema Dauerschätzungen und dem Zeitempfinden in Verbindung gebracht werden, wodurch sich ein theoretischer Rahmen für die Untersuchung der subjektiv erlebten Zeit während technischer Interaktionen aufspannt.

2.3 Zwischenfazit zum Nutzererleben

Nutzererleben beschreibt das subjektive Erleben einer Interaktion mit einem System (DIN, EN ISO 9241-11, 2018). Durch seine subjektive Natur ist es personen- und situationsgebunden, wodurch eine positive UX nicht per-se designt werden, sondern lediglich durch die Gestaltung des Systems begünstigt werden kann (Lallemant et al., 2015). Das CUE-Modell (Mahlke, 2008; Minge

& Thüring, 2018; Thüring & Mahlke, 2007) definiert wesentliche Komponenten, Konsequenzen und Antezedenzen von UX und dient der vorliegenden Arbeit als Rahmenmodell. Zu den Komponenten gehören die Wahrnehmung instrumenteller und nicht-instrumenteller Qualitäten sowie die emotionale Reaktion, die auf Basis von Interaktionscharakteristika entstehen. Das Ziel der vorliegenden Arbeit ist es, ein solches Interaktionscharakteristika näher zu untersuchen: die Dauer der Interaktion. Die Interaktionsdauer bezogen auf die Wahrnehmung und Bewertung zeitlicher Aspekte der Interaktion wird im CUE-Modell zwar nicht explizit erwähnt, manche Autoren argumentieren jedoch, dass subjektiv erlebte Zeit eine wesentliche Rolle bei der Entstehung von UX spielt (Liikkanen & Gómez Gómez, 2013). Diese Behauptung soll in der vorliegenden Arbeit unter anderem untersucht werden. Darüber hinaus soll geprüft werden, auf welche Komponenten des Nutzererlebens und in welcher Stärke sich die wahrgenommene Interaktionsdauer auswirkt. Ergebnisse hierzu können dazu beitragen, das CUE-Modell näher zu spezifizieren und gegebenenfalls die Wahrnehmung der Interaktionsdauer einer der Komponenten des Nutzerlebens zuzuordnen. Zunächst muss jedoch geklärt werden, was genau unter Wahrnehmung von Interaktionsdauern zu verstehen ist und welche Erkenntnisse bereits darüber vorliegt, wie Menschen Dauern wahrnehmen.

Kapitel 3

Erlebte Zeit - prospektive Dauerschätzungen

Dauerwahrnehmung ist ein breit untersuchtes Phänomen in der psychologischen Forschungslandschaft (Matthews & Meck, 2016). Forschungsarbeiten hierzu umfassen verschiedene Bereiche, wie zum Beispiel Forschung zur Wahrnehmung von Zeitdauern, Forschung zur Ab- und Zeitfolge von Ereignissen und Forschung zu zeitlichen Perspektiven, mit der Ereignisse und Erlebnisse in Bezug auf Vergangenheit, Gegenwart und Zukunft eingeordnet und bewertet werden (Block, 1990). Im Mittelpunkt dieser Arbeit steht die Forschung zur Wahrnehmung von Zeitdauern, genauer gesagt von Interaktionsdauern bei Mensch-Computer-Interaktionen. Dieser Anwendungskontext ist in Grundlagenarbeiten kaum zu finden. Trotzdem können kognitionspsychologische Modelle und empirische Arbeiten als Rahmen für das vorliegende Forschungsziel verstanden werden.

Grundlegend für das Forschungsfeld Dauerwahrnehmung ist die Frage, wie Menschen dazu in der Lage sind, Dauern wahrzunehmen und anhand dieser Wahrnehmung präzise Erwartungen aufzubauen, Handlungen auszurichten und verschiedene Dauern zu unterscheiden. In einer Vielzahl von Studien wurden diese Fragen experimentell untersucht. Dabei wird typischerweise zwischen zwei Arten von Dauerschätzungen (*Paradigmen*) unterschieden: prospektive und retrospektive Schätzungen. Von prospektiven Schätzungen spricht man, wenn dem Befragten bereits vor dem zu schätzenden Intervall bewusst ist, dass die Zeitdauer von Bedeutung ist (Droit-Volet & Meck, 2007). Man geht

davon aus, dass durch das Wissen über die Relevanz der Dauer, Aufmerksamkeit auf deren Verstreichen gerichtet wird und eine kontinuierliche Überwachung erfolgt. Block (1990) nennt diese Art der Dauerschätzung *experienced time* (*erlebte Zeit*). Bei Experimenten zu retrospektiven Dauerschätzungsaufgaben hingegen wird den Probanden erst nach dem zu schätzenden Zeitintervall die Relevanz der Dauer bewusst. Daher spielen bei diesem Paradigma primär erinnerungsbasierte Mechanismen eine Rolle. Solche Dauerschätzungen werden daher auch als *remembered time* (*erinnerte Zeit*) bezeichnet (Block, 1990).

Die Unterscheidung zwischen den beiden Paradigmen ist notwendig, da Daten darauf hinweisen, dass beide Schätzarten auf unterschiedlichen kognitiven Mechanismen basieren. So demonstrierten Block und Zakay (1997) in einer Meta-Analyse, dass retrospektive Schätzungen im Mittel zu stärkeren Unterschätzungen als prospektive Schätzungen führen. Retrospektive Schätzungen zeigen darüber hinaus im Mittel eine höhere Variabilität zwischen Personen (Block & Zakay, 1997). Auch beeinflussende Faktoren sind bei beiden Paradigmen unterschiedlich. So hat die Verarbeitungsschwierigkeit der zu bearbeitenden Aufgabe (*processing difficulty*) während der Schätzungsaufgabe nur einen systematischen Effekt auf prospektive Schätzungen: Je schwieriger die Verarbeitung, umso kürzer die prospektive Schätzung (Block & Zakay, 1997). Hohe Komplexität der Stimuli hingegen führt nur bei retrospektiven Schätzungen zu einer Verlängerung der Schätzung. Solche Befunde sowie die Meta-Analyse von Block, Hancock und Zakay (2010) stützen die Annahme, dass für beide Paradigmen unterschiedliche Erklärungs- und kognitive Modellansätze benötigt werden.

Neben der Unterscheidung von pro- und retrospektivem Paradigma können verschiedene Arten unterschieden werden, wie Dauerschätzungen abgegeben werden (Allan, 1979; Grondin, 2010; Wearden, 2016). Eine Auswahl davon ist in Tabelle 3.1 zusammengefasst. Alle dort aufgeführten Schätzarten sind im prospektiven Paradigma nutzbar. Im retrospektiven Paradigma hingegen entfällt die Produktion sowie die meisten Sonderformen des Vergleichs, da hier durch die Instruktion Aufmerksamkeit auf das Verstreichen der Zeit gelenkt wird. Im Mittelpunkt der vorliegenden Arbeit stehen verbale Schätzungen. In der hier vorgestellten Literatur werden aber auch anderen Schätzarten genutzt.

Wie in der Einleitung beschrieben, wird der Fokus dieser Arbeit auf erleb-

TABELLE 3.1: Arten zur Abgabe von Dauerschätzungen.

Methode	Beschreibung
Verbale Schätzung	Eine Dauer wird anhand eines Ereignisses (z.B. Abspielen eines Tons) präsentiert und der Proband schätzt die Dauer verbal ein. <i>Abhängige Variable: verbale Dauerangabe [Sek.]</i>
Produktion	Eine Dauer wird verbal vorgegeben und der Proband produziert die Dauer beispielsweise durch das Drücken einer Taste. <i>Abhängige Variable: Dauer [Sek.]</i>
Reproduktion	Eine Dauer wird anhand eines Ereignisses präsentiert und der Proband reproduziert die Dauer beispielsweise durch das Drücken einer Taste. <i>Abhängige Variable: Dauer [Sek.]</i>
Vergleich	Ein konstantes Standardintervall und ein variables Intervall werden pro Trial nacheinander präsentiert. Der Proband gibt in einem binären Antwortformat an, ob das zweite Intervall länger als das Standardintervall war. In einer Variation wird das Standardintervall zunächst erlernt und in der Testphase nur das variable Intervall präsentiert. <i>Abhängige Variable: Wahrscheinlichkeit einer Länger-Antwort ($p(\text{länger})$) oder Parameter der psychometrischen Funktion (z.B. Point of Subjective Equality (PSE) oder Steigung am PSE)</i>
Bisection Task	Sonderform des Vergleichs: Es werden zwei Standardintervalle, ein kurzes und ein langes, mehrfach präsentiert und erlernt. In der Testphase werden neben den beiden Standardintervallen meist fünf dazwischen liegende Intervalle mehrfach und in randomisierter Reihenfolge präsentiert. Für jedes gibt der Proband an, ob das erlebte Intervall eher dem langen oder dem kurzen Standardintervall entspricht. <i>Abhängige Variable: $p(\text{Lang})$, Parameter der psychometrischen Funktion</i>
Generalisation	Sonderform des Vergleichs: Es wird ein Standardintervall erlernt. In der Testphase werden umliegende Intervalle in randomisierter Reihenfolge präsentiert. Für jedes gibt der Proband an, ob er das präsentierte Intervall für das Standardintervall hält. <i>Abhängige Variable: $p(\text{Ja})$</i>

te Zeit, also prospektive Dauerwahrnehmung gelegt, da diese eine hohe Relevanz im Kontext der Mensch-Computer-Interaktion hat. Im Zentrum stehen dabei Interaktionen, bei denen nicht die Interaktion selber das Nutzerbedürfnis befriedigt, wie bei einem Spiel, sondern das Resultat der Interaktion den Grund für die Interaktion darstellt. Beispiele hierfür sind die Informationsbeschaffung auf einer Webseite oder das Ändern der Schriftgröße in einem Textverarbeitungsprogramm. In den hier betrachteten Mensch-Computer-Interaktionen kann davon ausgegangen werden, dass während dem Erleben der Interaktion Aufmerksamkeit auf das Verstreichen der Zeit gelenkt wird (prospektive Dauerwahrnehmung), da eine schnelle Erreichung des Resultats der Interaktion (z.B. finden der Information oder verändern der Schriftgröße) meist ein wesentliches Bedürfnis des Nutzers darstellt. Im Folgenden werden daher Modelle und Befunde zu den kognitiven Prozessen bei prospektiver Dauerwahrnehmung präsentiert.

3.1 Modelle und Theorien zur erlebten Zeit

3.1.1 Informations-Prozess-Modell der Dauerwahrnehmung

Frühe experimentelle Arbeiten im prospektiven Paradigma konzentrierten sich auf die Identifikation der Weber-Konstante bei Intervallen zwischen 400 ms und 15.4 Sek. (für eine Zusammenfassung siehe Treisman, 1963). Im Rahmen dieser Frage schlug Treisman (1963) ein *Informations-Prozess-Modell* (Information-Processing Model of Timing) vor (siehe Abb. 3.1), das auf einer *inneren Uhr* (internal clock) basiert. Aufbauend auf dieser Idee präziserte Treisman (1963) Prozessannahmen zur kognitiven Verarbeitung bei prospektiven Dauerschätzungen.

Die Kernidee des Modells besteht darin, dass Dauerschätzungen auf Basis von intern generierten, regelmäßigen Impulsen entstehen. Die Impulse werden von einem Taktgeber (*pacemaker*) produziert und wandern entlang eines Pfads (*pathway*) (Treisman, 1963, 2013). Ein Zähler (*counter*) greift die Impulse vom Pfad ab und akkumuliert sie, so dass die Menge von Impulsen während eines Intervalls durch eine Anzahl repräsentiert wird. Die Schnelligkeit, mit der der Taktgeber Impulse generiert, ist dabei weitestgehend konstant (Treisman,

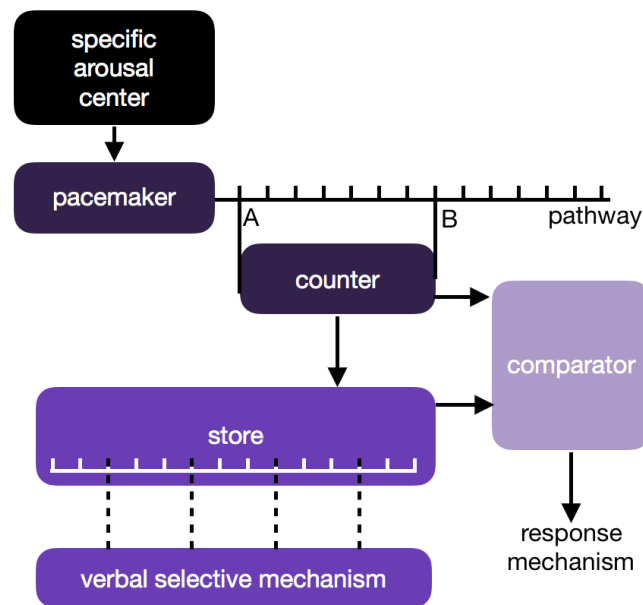


ABBILDUNG 3.1: Informations-Prozess-Modell der Dauerwahrnehmung, Eigendarstellung nach Treisman (1963).

1963), wodurch ein linearer Zusammenhang zwischen der Anzahl der akkumulierten Impulse und der tatsächlich Dauer des Intervalls entsteht (Droit-Volet & Wearden, 2003). Schwankungen in der Frequenz können jedoch durch Änderungen im Arousal-Level der Testperson entstehen und zu Unterschieden zwischen einzelnen Schätzungen und Versuchsdurchgängen führen. Betrachtet man diesen Prozess anhand einer Interaktion zwischen Mensch und Computer, so wird deutlich, dass der kognitive Prozess der Impulsakkumulation parallel zum Erleben der Interaktion abläuft. Kauft ein Nutzer beispielsweise eine Fahrkarte auf einer Webseite, so werden ab dem Zeitpunkt des Interaktionsstarts bis hin zur Beendigung der Interaktion Impulse aufaddiert. Die Anzahl dieser Impulse sollte laut den Modellannahmen des Informations-Prozess-Modells größer sein, wenn die Interaktionsdauer länger ist oder wenn sich das Arousal-Level des Nutzers durch negative Emotionen erhöht hat.

Im Modell wird weiter angenommen, dass der Zähler die akkumulierte Impulsanzahl entweder an einen Arbeitsspeicher (*store*) oder an eine vergleichende Komponente (*comparator*) weitergibt. Der Arbeitsspeicher nutzt eine räumliche Kodierung der Dauern, die vergleichbar mit einem Zahlenstrahl ist. Unterschiedliche Dauern bzw. Impulsanzahlen sind über den logarithmierten Abstand vom Nullpunkt kodiert und mit einer sprachlichen Benennung verbun-

den (*verbal selective mechanisms*). Diese Art der Kodierung führt dazu, dass kurze Dauern in ihrer Repräsentation höher aufgelöst sind als längere Dauern, d.h. sie sind genauer repräsentiert und können leichter voneinander diskriminiert werden. Wird nach dem Kauf der Fahrkarte aus dem obigen Beispiel eine bestimmte Impulsanzahl an den Arbeitsspeicher übergeben, führt dies zu einer gesteigerten Aktivierung an der entsprechenden Stelle des Zahlenstrahls. Diese Stelle kann mit einer verbalen Benennung verbunden sein, wodurch der Nutzer die Dauer der Interaktion benennen kann (z.B. zwei Minuten). Die Aktivierung auf dem Zahlenstrahl unterliegt jedoch einer räumlichen Verteilung. Hierdurch werden auch benachbarte Bereiche des Zahlenstrahls aktiviert, wodurch Fehler in der Schätzung entstehen können. So könnte zum Beispiel die Interaktionsdauer von zwei Minuten fälschlicherweise auf zweieinhalb Minuten geschätzt werden.

Je höher die Aktivierung an einer Stelle des Zahlenstrahls ist, umso wahrscheinlicher ist nicht nur die Aktivierung einer sprachlichen Benennung, sondern auch der Abruf der dort kodierten Dauer durch die vergleichende Komponente. Die vergleichende Komponente empfängt Informationen vom Arbeitsspeicher und vom Zähler und kann diese gegeneinander abwägen. Gibt man einer Versuchsperson beispielsweise die Aufgabe, eine Interaktion nach zwei Minuten zu beenden (Produktion), so wird im Arbeitsspeicher der Bereich um die korrespondierende Impulsanzahl aktiviert. Dadurch kann die entsprechende Impulsanzahl unter der oben beschriebenen Fehleranzahl vom Arbeitsspeicher an die vergleichende Komponente weitergegeben werden und dort mit den akkumulierten Impulsen aus dem Zähler verglichen werden. Dieser Vergleich dient der Entscheidung, wann zwei Minuten vergangen sind und die Interaktion beendet werden soll. Da all diese Prozesse verrauscht sind, entstehen zufällige sowie konstante Fehler in der Dauerschätzung.

Mit diesem Modell lieferte Treisman (1963) ein kognitives Modell einer inneren Uhr, das sehr klare Annahmen aufstellt und starke Ähnlichkeit mit späteren Modellen und Theorien hat, die in der Forschung mit Menschen als auch mit Tieren entstanden sind. Für den Bereich Mensch-Computer-Interaktion zeigt das Modell eine mögliche Erklärung der kognitiven Prozesse auf, die dem Nutzern erlauben das Interaktionsmerkmal Dauer wahrzunehmen und einzuschätzen (vgl. CUE-Modell). Gleichzeitig werden durch die Prozessannahmen

auch Gründe für eine fehlerhafte Schätzung deutlich, wie beispielsweise ein erhöhtes Arousal-Level oder das Rauschen der Aktivierungsprozesse. Über die Interpretation der Dauer durch den Nutzer oder die Einbettung der Dauerwahrnehmung in eine Erlebensbewertung werden jedoch im Informations-Prozess-Modell keine Annahmen getroffen.

3.1.2 Scalar Timing Theory (STT)

Eine wichtige Weiterführung des Modells von Treisman (1963) ist die *Scalar Timing Theory* von Gibbon, Church und Meck (1984). Sie beruht auf der einen Seite auf der Annahme, dass Dauerschätzungen skalare, relativistische Eigenschaften haben und auf der anderen Seite auf ähnlichen Prozessannahmen wie das Informations-Prozess-Modell (Treisman, 1963). “Skalare Eigenschaft” bedeutet, dass die geschätzte Dauer durch eine skalierende Transformation mitwächst, wenn sich die tatsächliche Dauer vergrößert. Dies gilt jedoch nicht nur für die mittlere geschätzte Dauer, sondern auch für deren Standardabweichung. Diese Annahme wurde zuvor im Rahmen der *Scalar Expectancy Theory* (SET) formuliert (Gibbon, 1977). Gibbon (1977) begründete diese Idee auf Basis der Beobachtung, dass Ratten bzw. Tauben bei Zeitschätzaufgaben im Sekunden- bis Minutenbereich zwar relativ präzise Schätzungen abgeben können, die Schätzleistung jedoch abhängig von der zu schätzenden Dauer ist. Mit anderen Worten vergrößert sich mit steigender Dauer nicht nur die mittlere Dauerschätzung, sondern auch die Standardabweichung der Dauerschätzung.

Die Annahme der skalaren Eigenschaften von Dauerschätzungen entspricht in einigen Teilen dem Modell von Treisman (1963, 2013), auch wenn dieser sich mit menschlichen Probanden und im Schnitt kürzeren Zeiten beschäftigte (Gibbon, 1977). Durch die Proportionalität der mittleren Schätzung und der Standardabweichung zum zu schätzenden Intervall entsteht laut der STT ein konstanter Koeffizient der Variation γ mit $\gamma = \frac{SD}{M}$ in der Schätzung (Gibbon, 1977, S. 285). Dieser entspricht einem konstanten Sensitivitätsmaß, das bereits von Treisman (1963) als Weber-Konstante für Dauerschätzungen untersucht wurde (Gibbon, 1977). Die von der STT vorhergesagte Standardabweichung kann darüber hinaus mit der probabilistischen Aktivierung eines Bereichs im Arbeitsspeicher des Informations-Prozess-Modell (Treisman, 1963) verglichen werden. Durch die Abbildung von Zeitdauern auf einem logarithmierten Zah-

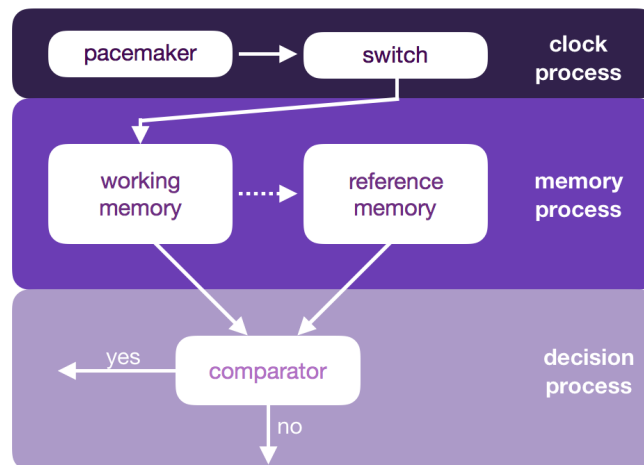


ABBILDUNG 3.2: Scalar Timing Theory, Eigendarstellung nach Gibbon et al. (1984).

lenstrahl im Ansatz von Treisman (1963) sind aktivierte Bereiche von kurzen Dauern höher aufgelöst als längere Dauern. Dadurch kann die vergleichende Komponente des Informations-Prozess-Modells bei kürzeren Dauern aus einer stärker um den Mittelwert konzentrierten Verteilung ziehen, was dazu führt, dass die Varianzen der Schätzungen bei kürzeren Dauern geringer sind. Für die Vorhersage der Dauerwahrnehmung im Rahmen einer Mensch-Computer-Interaktion gilt also für beide Ansätze, dass längere Interaktionen auch länger, aber mit einer größeren Variabilität geschätzt werden. Beide Ansätze, STT und Informations-Prozess-Modell, stimmen daher grob in ihrer Vorhersage überein, auch wenn sie sich in einigen mathematischen Formulierungen der zugrundeliegenden Mechanismen unterscheiden.

Eine weitere Ähnlichkeit zwischen der STT und dem Informations-Prozess-Modell wird deutlich, wenn man die Prozessannahmen (siehe Abb. 3.2) betrachtet. Gibbon et al. (1984) gliederten die Prozesse in drei Schritte, die auf verschiedenen Modulen basieren. Diese Schritte beziehen sich auf die Wahrnehmung (*clock process*), das Erinnern (*memory process*) und das Unterscheiden von Dauern (*decision process*). Sie sind durch Helligkeitsabstufungen in Abbildung 3.2 gekennzeichnet. Die gleiche Kennzeichnung findet sich auch in Abbildung 3.1 und macht deutlich, dass diese Prozessschritte bereits im Modell von Treisman (1963) zu finden sind.

Beim Vergleich der Abbildungen 3.1 (Informations-Prozess-Modell der Dauerwahrnehmung) und 3.2 (STT) fallen große Ähnlichkeiten in den betei-

ligten Modulen auf. Beide Ansätze gehen von einem kontinuierlich arbeitendem Taktgeber (*pacemaker*) aus, dessen Pulse im Zähler (vgl. *counter/accumulator*) akkumuliert werden. Die STT erweitert den clock process um einen Schalter (*switch*). Dieser sorgt dafür, dass nur während des zu schätzenden Intervalls Impulse vom Taktgeber in den Zähler geleitet werden. Der Schalter reagiert dabei mit einer gewissen Latenz auf externe Stimuli (Gibbon et al., 1984), die das zu schätzende Intervall markieren. Der Zähler akkumuliert die Impulse und gibt die Anzahl weiter an das Arbeitsgedächtnis¹.

Die Konzeption des Arbeitsspeichers (*store/working memory*) unterscheidet sich zwischen den beiden Ansätzen. Während Treisman (1963) auf eine räumliche Kodierung zurückgriff, nahmen Gibbon et al. (1984) eine sehr einfache Konzeption des Arbeitsspeichers an, in der lediglich die Anzahl aus dem Zähler übernommen wird (Gibbon et al., 1984, S. 55). Diese Anzahl kann langfristig gespeichert werden, indem sie ins Referenzgedächtnis als Teil des Langzeitgedächtnisses überführt wird. Auf Basis der Referenz und der aktuell akkumulierten Anzahl entscheidet die vergleichende Komponente (*comparator*), ob eine Antwort bzw. Handlung ausgelöst wird oder nicht.

Zusammengefasst stellt die STT ein Prozessmodell dar, welches die kognitiven Prozesse beim Wahrnehmen einer Dauer beschreibt. Sie bietet damit einen Einblick in verschiedene Quellen für Fehler und Varianzen in Dauerschätzungen, wie beispielsweise die skalare Vergrößerung der Schätzvarianz bei längeren Dauern, die Latenz des Schalters, die Speicherung im Referenzgedächtnis oder den vergleichenden Prozess. Bezogen auf den Anwendungskontext Mensch-Computer-Interaktionen zeigt sie damit auf, welche Prozessschritte bei der Wahrnehmung einer Interaktionsdauer durchlaufen werden. Mit ihrer Hilfe kann zum Beispiel vorhergesagt werden, dass sich Dauerschätzungen systematisch verändern, wenn die Latenz des Schalters vergrößert wird. Dies kann beispielsweise durch Ablenkung passieren (Matthews & Meck, 2016; Meck, 1984). Ähnlich wie das Informations-Prozess-Modell ist die STT dazu geeig-

¹Der Zähler wird inzwischen eher dem clock process als dem memory process zugeordnet (Church, Meck & Gibbon, 1994; Droit-Volet & Wearden, 2003; Droit-Volet & Meck, 2007; Gibbon & Church, 1984), was in der Veröffentlichung von Gibbon et al. (1984) noch nicht ganz eindeutig erfolgte. Gibbon et al. (1984) bildeten zwar ebenfalls den Zähler als Teil des clock processes ab (Gibbon et al., 1984, S. 55), schrieben aber wiederholt, dass der Zähler Teil des Arbeitsgedächtnisses sei (Gibbon et al., 1984, S.53, S. 55).

net, Vorhersagen zu machen, wie lange eine Interaktionsdauer wahrgenommen wird und unter welchen Bedingungen eine Verzerrung der Dauerwahrnehmung erfolgt. Sie ist nicht geeignet, Bewertungsprozesse einer erlebten Dauer bezogen auf Akzeptabilität oder Valenz vorherzusagen. Geht man jedoch davon aus, dass solche Bewertungsprozesse auf der wahrgenommenen Dauer beruhen, stellen sowohl Informations-Prozess-Modell als auch STT wichtige Grundlagenmodelle für die Untersuchung der Wahrnehmung von Interaktionsdauern dar.

3.1.3 Dual-Process Contingency Model

Während sich die bisher besprochenen Ansätze ausschließlich auf prospektive Dauerschätzungen unter Berücksichtigung eines internen Taktgebers beziehen, stammt das *Dual-Process Contingency Model* von Zakay (1993a) aus einer Forschungstradition, die sich sowohl mit prospektiven als auch mit retrospektiven Schätzungen beschäftigt. Es entstand basierend auf Studien mit menschlichen Probanden, in denen der Einfluss kognitiver Beanspruchung auf prospektive Dauerschätzungen gezeigt wurde (für eine Meta-Analyse siehe Block et al., 2010). Solche Befunde können durch das Informations-Prozess-Modell der Dauerwahrnehmung oder die STT nicht erklärt werden (Zakay & Block, 1995), da beide Ansätze Aufmerksamkeitsprozesse vernachlässigen. Auch wenn Latenzen bei der Betätigung des Schalters in der STT von Aufmerksamkeitsprozessen beeinflusst werden können (Matthews & Meck, 2016; Meck, 1984), sind diese Modelle und Theorien vielmehr für Situationen konzipiert, in denen die Dauerschätzung die einzige Aufgabe für den Probanden darstellt. Vor allem für den Anwendungsbereich Mensch-Computer-Interaktion stellt dies allerdings eine starke Einschränkung ein. Wie weiter oben bereits diskutiert, muss die Akkumulation der Impulse als Teil der Dauerschätzungsaufgabe parallel zur Interaktion geschehen. Hier ist die Dauerschätzung also nicht die einzige Aufgabe des Probanden.

Um prospektive Schätzungen auch unter Ablenkung prozessorientiert darstellen zu können, wird im Dual-Process Contingency Model nicht nur ein Impulsgeber als Informationsquelle für Dauerschätzungen, sondern auch ein erinnerungsbasierter Prozess angenommen (siehe Abb. 3.3). Der Impulsgeber generiert ähnlich wie im Informations-Prozess-Modell der Dauerwahrnehmung

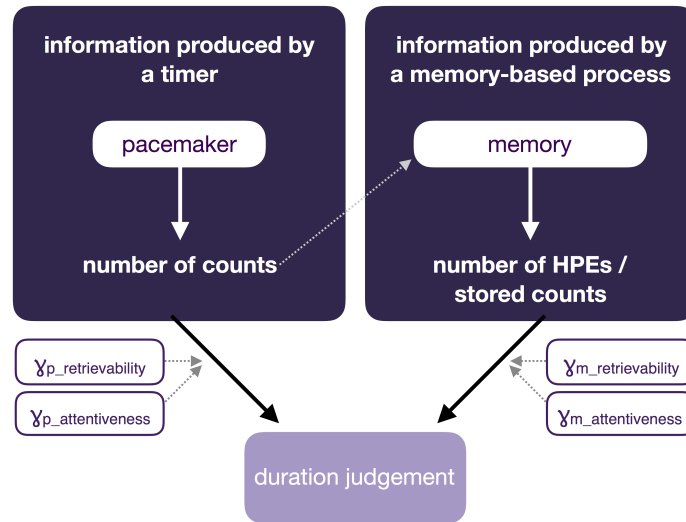


ABBILDUNG 3.3: Dual-Process Contingency Model, Eigendarstellung nach Zakay (1993a). HPE steht für high-priority event und γ ist die Gewichtung der Informationsquelle abhängig von der Zugänglichkeit (retrievability) und der zugewiesenen Aufmerksamkeit (attentiveness).

und in der STT eine Impulsanzahl, die nur kurzzeitig zur Verfügung steht. Weitere Ähnlichkeiten zeigen sich darin, dass der Impulsgeber durch einen dem Schalter ähnlichen Mechanismus auf Null zurückgesetzt werden kann und seine Impulse für eine langfristige Speicherung, wie sie beispielsweise für eine Reproduktion oder einen Vergleich benötigt wird, an das Langzeitgedächtnis weitergeben kann. Der erinnerungsbasierte Prozess hingegen unterscheidet sich von der Konzeption der Speicherungsebene (*memory process*) der STT, da er als eigenständige Informationsquelle neben dem Impulsgeber zur Verfügung steht und nicht, wie in der STT angenommen, als Folgeprozess basierend auf den Informationen des Impulsgebers fungiert.

Im Dual-Process Contingency Model wird angenommen, dass beide Informationsquellen, Impulsgeber und Erinnerung, zu Zeitschätzungen beitragen. Die Stärke dieses Beitrags hängt davon ab, wie einfach die Informationen der jeweiligen Quelle zugänglich sind (*retrievability*) und wie viel Aufmerksamkeit der entsprechenden Quelle zugewiesen wird bzw. wurde (*attentiveness*). In einem rein prospektiven Paradigma wird alle Aufmerksamkeit auf die verstreichende Zeit und damit den Impulsgeber gelegt. Dadurch werden die Informationen, die durch den Taktgeber entstehen, besonders stark gewichtet. Bei ei-

ner retrospektiven Schätzung hingegen wird keinerlei Aufmerksamkeit auf den Impulsgeber gelegt, wodurch der erinnerungsbasierte Prozess zur Schätzung genutzt wird. Hier werden aus dem Gedächtnis (*memory*) hervorstechende Ereignisse, sogenannte *high-priority events* (*HPEs*), dekodiert, die während des zu schätzenden Intervalls enkodiert wurden. Beispiele für HPEs sind kontextuelle Veränderungen (Zakay, 1993a), wie das Erscheinen einer neuen Internetseite, wenn bei einem Kaufvorgang auf einer Webseite der Warenkorb oder die Eingabemaske zum Bezahlen geöffnet wird. Die Menge der Informationen bezüglich der Impulse und der HPEs dient als Grundlage für die Dauerschätzung.

Im Vergleich zu den vorher diskutierten Ansätzen ist das Dual-Process Contingency Model weit weniger spezifisch, da es auf der einen Seite keine mathematische Ausformulierung der einzelnen Prozesse und auf der anderen Seite keine genauen Definitionen von hervorstechenden Ereignissen (HPEs) bietet. Trotzdem leistet es einen entscheidenden Beitrag zum Verständnis der beteiligten kognitiven Prozesse bei der Wahrnehmung von Interaktionsdauern, da es höhere kognitive Funktionen, wie die Aufmerksamkeitsallokation, berücksichtigt. Damit macht es auch Situationen zugänglich, in denen eine Dauerschätzung neben einer Zweitaufgabe abgelegt wird, wie bei einer Mensch-Computer-Interaktion. Hier muss Aufmerksamkeit kontinuierlich zwischen der Schätz- und der Zweit-Aufgabe bzw. Interaktion verteilt werden und es liegt kein rein prospektives Paradigma mehr vor.

3.1.4 Attentional Gate Model (AGM)

Das *Attentional Gate Model* (AGM, Block & Zakay, 1996; Zakay & Block, 1995, 1997) ist eine Weiterentwicklung und Kombination des Dual-Process Contingency Models (Zakay, 1993a) und der STT (Gibbon et al., 1984). Wie in Abbildung 3.4 zu sehen ist, übernimmt es alle Module der STT (Vergleich 3.2) und erweitert den Prozessschritt zur Wahrnehmung der Dauer (*clock process*) um ein Tor (*gate*), das durch Aufmerksamkeitsprozesse beeinflusst wird.

Die Funktionsweise des Tors ist als ein Engpass zu verstehen, durch den die Impulse des Taktgebers (*pacemaker*) auf dem Weg zum Zähler (*cognitive counter*) hindurch müssen und dessen Breite bzw. Durchlässigkeit von der Menge der zugewendeten Aufmerksamkeit bedingt wird. Ist die gesamte Aufmerksamkeit auf das Verstreichen der Zeit gerichtet, öffnet sich das Tor vollständig

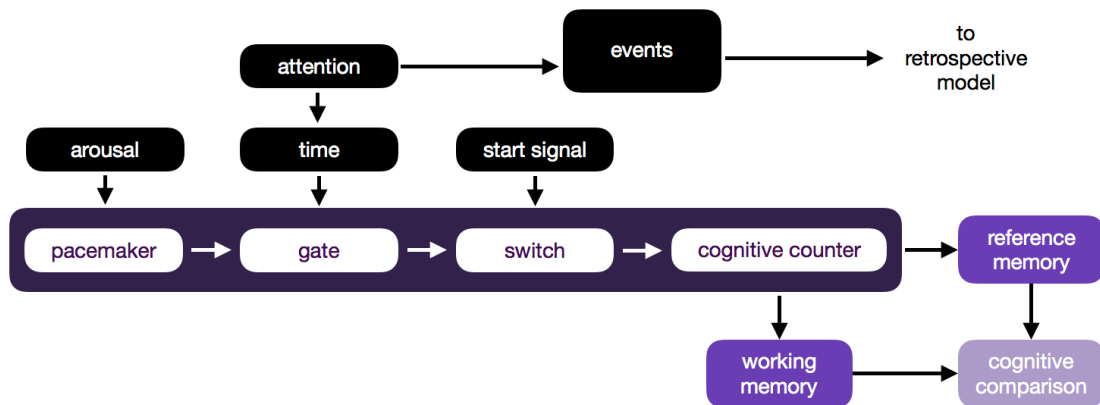


ABBILDUNG 3.4: Attentional Gate Model, Eigendarstellung nach Block und Zakay (1996) und Zakay und Block (1995, 1997).

und alle ankommenden Impulse können passieren. Sobald jedoch eine Aufgabe oder ein Ereignis Teile der Aufmerksamkeit für sich beanspruchen, steht weniger Aufmerksamkeit für die Dauerschätzung zur Verfügung und das Tor schließt sich graduell. Dadurch gelangen manche Pulse nicht mehr zum Zähler, sondern gehen verloren. Bei vollständiger Ablenkung durch ein anderes Ereignis schließt das Tor vollständig: Es gelangen keine Impulse in den Zähler und eine prospektive Dauerschätzung wird unmöglich.

Zusammengefasst wird im AGM die Annahme eines skalaren Zeitgebers (Pacemaker), der durch Arousal beeinflusst wird, aus früheren Modellen übernommen, während zusätzlich der Einfluss der Aufmerksamkeitsallokation berücksichtigt wird. Dadurch ist es für die Vorhersage von Dauerschätzungen im Anwendungskontext Mensch-Computer Interaktion besonders relevant. Solche Interaktionen kennzeichnen sich durch eine Aufteilung der Aufmerksamkeit auf verschiedene Teilaufgaben, die neben der Dauerschätzung absolviert werden. Darüber hinaus können während der Interaktion Emotionen auftreten, die eine Veränderung im Arousal-Level mit sich bringen.

3.2 Befunde im prospektiven Paradigma

Die vorgestellten Ansätze sind kognitive Modelle der Dauerwahrnehmung, die einen modularen Aufbau mit spezialisierten Komponenten zeigen (Ivry & Schlerf, 2008; Droit-Volet & Wearden, 2003), bei denen ein interner Taktgeber als zentrale Komponenten angenommen wird. Daher werden sie auch als

„timing-with-a-timer“-Modelle bezeichnet (Block & Zakay, 1996). Auf Grund der Ähnlichkeiten zwischen den Ansätzen lassen sich (teilweise modellübergreifende) Vorhersagen darüber machen, welche Eigenschaften Dauerschätzungen haben und welche Faktoren einen systematischen Einfluss zeigen sollten. Wichtige Vorhersagen betreffen die skalare Eigenschaft, Effekte des Arousal-Levels sowie der Aufmerksamkeit.

3.2.1 Skalare Eigenschaften

Der interne Taktgeber von „timing-with-a-timer“-Modellen geht mit der Annahme einher, dass eine Dauerschätzung mit der zu schätzenden Dauer skaliert. Daher sollten längere Dauern auch zu einer längeren Dauerschätzung führen. Dies konnte in vielen Experimenten und Studien sowohl im Sekundenbereich (z.B. Droit-Volet, Bigand, Ramos & Bueno, 2010; Droit-Volet, Ramos, Bueno & Bigand, 2013; Gil & Droit-Volet, 2011, 2012) als auch im Minutenbereich (z.B. Droit-Volet, Monceau, Berthon, Trahanias & Maniadakis, 2018) gezeigt werden. Darüber hinaus leitet sich aus dem Informations-Prozess-Modell (Treisman, 1963) und der STT (Gibbon et al., 1984) ab, dass Dauerschätzungen mit zunehmender objektiver Dauer variabler werden sollten, wozu es ebenfalls eine Vielzahl an experimentellen Befunden gibt (z.B. Allan, 1979; Brown, 1997; Droit-Volet et al., 2018; Wearden & Lejeune, 2008).

3.2.2 Effekt von Arousal auf prospektive Dauerschätzungen

Basierend auf dem Informations-Prozess-Modell der Dauerwahrnehmung (Treisman, 1963), der STT (Gibbon et al., 1984) und dem AGM (Block & Zakay, 1996; Zakay & Block, 1995, 1997) ist ein Effekt des Arousal-Levels auf prospektive Dauerschätzungen zu erwarten: Je höher das Arousal-Level ist, desto schneller produziert der Taktgeber Impulse und desto mehr von ihnen werden pro Zeiteinheit im Zähler akkumuliert. Daher sollte die gleiche Zeitdauer bei hohem Arousal als länger eingeschätzt werden als bei geringerem Arousal. Dies gilt für verbale Dauerschätzungen und Vergleiche. Bei Produktionen sollte das zu produzierende Intervall unter erhöhtem Arousal früher beendet werden, da hier die Impulsanzahl, die aus dem Referenzgedächtnis für

den Vergleich abgerufen werden muss, früher erreicht wird.

Diese Vorhersage wurde für verschiedene Schätzarten untersucht. Ein Beispiel für eine solche Studie ist die Arbeit von Gil und Droit-Volet (2012). Die Autorinnen präsentierten ihren Versuchspersonen für eine Dauer zwischen 200 ms und 1600 ms emotionsgeladene Bilder und ließen diese Dauer verbal schätzen. Die Bilder unterschieden sich bezüglich ihrer dargestellten Basisemotion (Ekel, Angst und Trauer) und dem Arousal-Level, welches sie hervorrufen sollten. Die Ergebnisse zeigten, dass Bilder, die a priori als stärker erregend klassifiziert worden waren und somit ein hohes Arousal-Level auslösen sollten, zu einer längeren Schätzung führten als weniger erregende Bilder bzw. als neutrale Bilder.

Neben solchen Studien zu verbalen Dauerschätzungen gibt es jedoch auch eine Vielzahl an Studien mit Produktionen, Reproduktionen und vergleichende Schätzarten, wie zum Beispiel der Bisection Task, Generalisationen sowie andere Vergleichsarten (Beschreibung der Schätzarten siehe Tab. 3.1, S. 21). In Tabelle 3.2 ist eine Auswahl an Studien zusammengetragen, die sich diesen verschiedenen Zeitschätzarten zuordnen lassen. Die Befunde entsprechen für Dauern bis zu zwei Sekunden den schätzart-spezifischen Vorhersagen bezüglich des Effekts des Arousal-Levels auf die Dauerschätzung. Studien, die dies stützen, sind in Tabelle 3.2 durch fetten Schriftsatz gekennzeichnet.

Für längere Dauern ist die Befundlage jedoch nicht so eindeutig. Hier berichten manche Autoren den erwarteten Effekt (Castellà et al., 2017; Droit-Volet et al., 2013; Fayolle et al., 2015), während andere keine bzw. keine eindeutige Evidenz für eine Veränderung der Dauerschätzungen durch Arousal bei Intervallen von zwei bis acht Sekunden finden (Angrilli et al., 1997; Nather et al., 2011; Noulhiane et al., 2007; Trapp & Thüring, 2016). Die Unklarheit in der Befundlage bei längeren Dauern kann auf der einen Seite durch die Induktionsart erklärt werden, mit der das Arousal-Level manipuliert wird (Droit-Volet et al., 2018). Auf der anderen Seite können sich die kognitiven Prozesse bei Dauerschätzungen von Intervallen unter zwei Sekunden von denen bei längeren Intervallen unterscheiden (Ulbrich, Churan, Fink & Wittmann, 2007).

Bezogen auf den ersten Erklärungsansatz stellt sich die Frage, welche Induktionsmethoden für Arousal typischerweise genutzt werden. In diesem Kontext wird das Arousal-Level meist als „globales Erregungsniveau“ (Müsseler, 2008,

S. 158) verstanden, welches Teil eines *Kernaffects* ist (*core affect*, Russell, 2003). Der Kernaffect beschreibt einen bewussten, neurophysiologischen Zustand, der als ein einfaches, unreflektiertes Gefühl wahrgenommen wird und sich aus einer hedonischen Komponente (Valenz) und dem Arousal (Erregungszustand) zusammensetzt (Russell, 2003). Daher werden zur Steigerung des Arousal-Levels häufig affekt- bzw. emotionsinduzierende Stimuli genutzt. Hierzu gehört in Studien zum Arousaleffekt auf Dauerschätzungen die Präsentation von emotionsgeladenen Bildern (Angrilli et al., 1997; Gil & Droit-Volet, 2012; Grommet et al., 2011; Grondin et al., 2014), von emotionalen Gesichts-

TABELLE 3.2: Übersicht über Studien zum Arousaleffekt im prospektiven Paradigma. Durch fetten Schriftsatz sind solche Studien gekennzeichnet, die den Arousaleffekt für Dauern bis zu 2 Sekunden zeigen. Nicht fett sind Studien mit längeren Dauern, für die eine weniger eindeutige Befundlage vorliegt.

verbale Schätzung	Gil & Droit-Volet, 2011; Gil & Droit-Volet, 2012; Noulhiane, Mella, Samson, Ragot & Pouthas, 2007
Produktion	Castellà, Cuello & Sanz, 2017; Gil & Droit-Volet, 2011
Reproduktion	Angrilli, Cherubini, Pavese & Manfredini, 1997; Gil & Droit-Volet, 2011; Noulhiane et al., 2007; Trapp & Thüring, 2016
Bisection Task	Droit-Volet, Brunot & Niedenthal, 2004; Droit-Volet, Mermillod, Cocenas-Silva & Gil, 2010; Droit-Volet, Bigand et al., 2010; Droit-Volet, Fayolle & Gil, 2011; Droit-Volet et al., 2013; Droit-Volet & Gil, 2015; Fayolle, Gil & Droit-Volet, 2015; Gil & Droit-Volet, 2009; Gil & Droit-Volet, 2011; Grommet et al., 2011; Nather, Bueno, Bigand & Droit-Volet, 2011; Tipples, 2008
Generalisation, Vergleich und ähnliche Methoden	Gil & Droit-Volet, 2011; Grondin, Laflamme & Gontier, 2014; Mella, Conty & Pouthas, 2011

ausdrücken (Droit-Volet et al., 2004; Gil & Droit-Volet, 2011; Grondin et al., 2014; Tipples, 2008), von Personen, deren Körperhaltung spezifisch für eine Emotion ist (Droit-Volet & Gil, 2015), sowie die Präsentation von emotionsauslösenden Geräuschen (Mella et al., 2011; Noulhiane et al., 2007; Trapp & Thüning, 2016). Dass solche Induktionsmethoden tatsächlich das physiologische Erregungsniveau verändern, wurde vor allem für emotionsgeladene Bilder und Geräusche gezeigt (Angrilli et al., 1997; Brouwer, Van Wouwe, Muehl, Van Erp & Toet, 2013; Mella et al., 2011).

Solche Stimuli haben jedoch häufig keine hohe, persönliche Relevanz für die Probanden. Dadurch kann zwar eine kurzfristige Änderung des Arousal-Levels hervorgerufen werden, diese bleibt aber nicht lange bestehen (Droit-Volet et al., 2018). Dies wird durch Studien gestützt, in denen mit stärkeren Induktionsmethoden auch für längere Dauern ein Effekt des Arousal-Levels gezeigt werden konnte: Fayolle et al. (2015) nutzten Stromschläge und fanden den Effekt bei Intervallen bis zu acht Sekunden in einer Bisection Aufgabe. Castellà et al. (2017) ließen Probanden kurz nach der Überquerung einer Hängebrücke aus Seilen (Höhe circa 20 m) ein einminütiges Intervall schätzen. Sie berichteten längere verbale Schätzungen unter dieser Bedingung mit hohem Arousal im Vergleich zu einer Schätzung unter weniger erregenden Bedingungen, in der die Probanden einen Waldweg entlang liefen. In einer Studie mit Ratten, denen zur Steigerung des Arousal-Levels intermittierend Kokain verabreicht wurde, konnte der verkürzende Effekt von erhöhtem Arousal auf Dauerproduktionen ebenfalls bei Intervallen bis zu 90 Sekunden gezeigt werden (Matell, King & Meck, 2004). Zusammenfassend stützen diese Studien die Erklärung, dass das häufige Ausbleiben eines signifikanten Effekts von Arousal auf Dauerschätzungen von Intervallen über zwei Sekunden zumindest teilweise auf eine zu geringe Veränderung des Arousal-Levels zurückgeführt werden kann.

Eine Alternativerklärung besteht in der Annahme, dass sich die beteiligten kognitiven Prozesse bei längeren Dauerschätzungen verändern (Noulhiane et al., 2007; Trapp & Thüning, 2016). So könnten beispielsweise Gedächtnis- und Aufmerksamkeitsprozesse bei langen Dauern eine größere Rolle spielen als bei kurzen (Ivry & Schlerf, 2008). Ulbrich et al. (2007) untersuchten diese Annahme in einem quasi-experimentellen Versuchsdesign. Dazu teilten sie ihre Stichprobe anhand der Arbeitsgedächtniskapazität in zwei Gruppen ein (nied-

rige vs. hohe Arbeitsgedächtniskapazität) und ließen die Probanden Dauern von ein bis fünf Sekunden reproduzieren. Die Autoren argumentierten, dass sich die Schätzungen zwischen den beiden Gruppen bei längeren Intervallen (> 3 Sekunden) stärker unterscheiden müssten als bei kürzeren, wenn Prozesse des Arbeitsgedächtnisses bei längeren Dauern eine höhere Relevanz haben. Hier sollten Versuchspersonen mit einer höheren Arbeitsgedächtniskapazität längere Reproduktionen zeigen, da sie mehr Impulse abspeichern können als Probanden mit einer niedrigeren Kapazität. Das Befundmuster entsprach dieser Erwartung und stützt damit die Annahme, dass Dauerschätzungen von Intervallen über zwei Sekunden stärker von Gedächtnisprozessen abhängig sind als solche unter zwei Sekunden.

Bei wesentlich längeren Intervallen scheinen noch weitere Prozesse an Bedeutung zu gewinnen. Droit-Volet et al. (2018) untersuchten Dauern zwischen zwei und 32 Minuten. Sie konnten zeigen, dass die Schätzungen nicht nur durch das selbstberichtete Arousal, sondern auch durch die Zeitflussbeurteilung vorhergesagt wurden. Dieses wurde jedoch nicht experimentell, sondern rein korrelativ untersucht. Das Gefühl von sich ziehender oder schnell vergehender Zeit wird in Kapitel 4 erneut aufgegriffen und näher diskutiert.

Zusammengefasst stützen Befunde zum Effekt von Arousal auf Dauerschätzungen die Annahme, dass prospektive Dauerschätzungen auf einem Taktgeber basieren, dessen Frequenz vom aktuellen Arousal-Level abhängt. Das Arousal-Level kann durch Manipulation des emotionalen oder physiologischen Zustands verändert werden, wobei auf Grund der typischen Induktionsmethoden (Bilder, Töne) meist nur kleine und kurzfristige Änderungen hervorgerufen werden. Der Arousaleffekt erscheint zudem vor allem bei Dauerschätzungen von unter zwei Sekunden eine wesentliche Rolle zu spielen. Je länger die Dauer wird, umso mehr gewinnen andere Prozesse, wie Aufmerksamkeits- und Gedächtnisprozesse sowie die Wahrnehmung des Zeitflusses, an Bedeutung. Für Interaktionsdauern im Kontext Mensch-Computer-Interaktion erscheint der Arousaleffekt daher nur mäßig relevant, da er nur bei sehr starken Veränderungen des Arousal-Levels oder bei sehr kurzen Interaktionsdauern zu erwarten ist.

3.2.3 Effekt von Aufmerksamkeit auf prospektive Dauerschätzungen

Basierend auf dem AGM können spezifische Vorhersagen für den Effekt von Aufmerksamkeitsverschiebungen auf prospektive Dauerschätzungen gemacht werden. So sollte, bei gleichem Arousal-Level, eine Dauer verbal kürzer geschätzt werden, wenn mehr Aufmerksamkeit von der Zeitschätzaufgabe abgezogen wird. Laut dem AGM geschieht dies, da nicht alle Impulse das Tor passieren, die während des Verstreichens der Dauer generiert wurden. Dadurch ist die Anzahl der akkumulierten Impulse geringer und es wird eine kürzere Dauer geschätzt. Bei einer Produktion hingegen entsteht eine längere Schätzung, wenn die Aufmerksamkeit nicht alleine auf die Zeitschätzaufgabe gerichtet wird. Hier muss länger produziert werden bzw. müssen mehr Impulse generiert werden bis die Referenzanzahl erreicht ist, da einige Impulse verloren gehen. Ein ähnlicher Effekt lässt sich auf Basis der STT vorhergesagen, sobald man Latenzen in der Öffnung des Schalters berücksichtigt, die durch Aufmerksamkeitsprozesse beeinflusst werden (Matthews & Meck, 2016; Meck, 1984). Diese Vorhersagen wurden experimentell bereits in vielen Kontexten bestätigt (für Reviews siehe Brown, 2008; Grondin, 2010; Matthews & Meck, 2016; Zakay, 2005). Dabei konzentrieren sich einzelne Arbeiten meist auf eine von zwei verschiedenen Arten von Aufmerksamkeit: *selektive* oder *geteilte* Aufmerksamkeit (Matthews & Meck, 2016).

Selektive Aufmerksamkeit beschreibt die Auswahl eines Stimulus, eines Objekts oder eines Ortes aus einer Menge möglicher Stimuli/Objekte/Orte und das Belegen dieser Auswahl mit Aufmerksamkeit (Matthews & Meck, 2016; Müsseler, 2008). Hierdurch entsteht eine Fokussierung, die auf der einen Seite mit einer höheren Verarbeitungskapazität für die selektierten Informationen und auf der anderen Seite mit einem “Ausblenden” anderer Informationen einhergeht. Die Zuwendung von Aufmerksamkeit auf ausgewählte Stimuli oder Bereiche kann willkürlich gesteuert werden (*endogene* Aufmerksamkeitssteuerung). Sie kann aber auch unwillkürlich durch die Eigenschaften eines Stimulus entstehen (*exogene* Aufmerksamkeitssteuerung; für einen Überblick siehe z.B. Hagendorf, Krummenacher, Müller & Schubert, 2011).

In Bezug auf Dauerwahrnehmung stellt sich die Frage, auf welche Informa-

tion Aufmerksamkeit während einer Schätzaufgabe selektiv gerichtet wird und wie diese Aufmerksamkeitszuwendung die Dauerschätzen beeinflusst. Typische Untersuchungen nutzen hierfür zum Beispiel Ereignisse, die Aufmerksamkeit kurzfristig von der Schätzaufgabe abziehen (exogene Steuerung): Block, George und Reed (1980) instruierten ihre Probanden die Aufmerksamkeit auf die Zeit zu legen, während sie einen Topf gefüllt mit Flüssigkeit beobachteten. Die Flüssigkeit konnte dabei entweder kurz aufkochen oder nicht. Anschließend wurde die wahrgenommene Beobachtungsdauer reproduziert. Die Ergebnisse zeigten kürzere Reproduktionen bei Aufkochen der Flüssigkeit, da es hier zu einer kurzfristigen, exogenen Aufmerksamkeitverschiebung kam. Der Befund passt zu einer geringeren Anzahl von akkumulierten Impulsen bei Ablenkung und einer entsprechend als kürzer erlebten Dauer. Andere Studien ließen ihre Versuchspersonen auf den Ort oder die Modalität des zeitgebenden Stimulus fokussieren (endogene Steuerung, Mattes & Ulrich, 1998). Hier zeigte sich die erwartete Verlängerung der Dauerschätzung. Zusammengefasst führt also eine selektive Zuwendung der Aufmerksamkeit auf zeitlich relevante Aspekte, wie den Zeitfluss und den zeitgebenden Stimulus oder Ort, zu einer Verlängerung der erlebten Zeit (Matthews & Meck, 2016).

Ähnlich zu Befunden der selektiven Aufmerksamkeit unterstützen auch Arbeiten zur geteilten Aufmerksamkeit die Vorhersagen des AGM. Hier wird häufig das *Dual-Task-Paradigma* genutzt, in dem neben der Zeitschätzaufgabe eine Zweitaufgabe durchgeführt wird. Die Spannweite der bereits untersuchten Zweitaufgaben ist dabei enorm. Sie umfasst Aufgaben zur motorischen Steuerung, perzeptuellen Diskrimination, verbalen Verarbeitung, visuellen Suche, mentalen Arithmetik, Entscheidungen sowie erinnerungsbasierte Aufgaben (für Reviews siehe Brown, 1997, 2008). Die Befundlage zeigt fast einheitlich, dass verbale Dauerschätzungen innerhalb eines Dual-Task-Paradigmas kürzer und ungenauer sind sowie eine höhere Variabilität aufweisen als Dauerschätzungen, die ohne Zweitaufgabe abgegeben werden. Für Produktionen zeigen sich entsprechend längere Schätzungen, wenn eine Zweitaufgabe zusätzlich erledigt werden muss. Brown (1997) nennt diesen Effekt den *Interferenzeffekt*. In 91% der von Brown (2008) zusammengefassten Arbeiten wird Evidenz für diesen Effekt berichtet.

Auch im Kontext der Mensch-Computer-Interaktion gibt es Arbeiten, die

den Interferenzeffekt zeigen. Hertzum und Holmegaard (2015) untersuchten verbale Dauerschätzungen bei einem Problemlöse-Spiel an einem Computer, wobei sie die Hälfte der Stichprobe instruierten, während der Bearbeitung laut zu denken. Lautes Denken ist im Kontext der Untersuchung von Mensch-Computer-Interaktionen eine häufig angewandte Methode, bei der Nutzer während einer Interaktion ihre Gedanken verbalisieren (*concurrent think aloud*). Diese Gedanken beziehen sich meist darauf, was der Nutzer tut, wie er versucht die Aufgabe zu lösen, welche Erwartungen er hat und warum er bestimmte Handlungen durchführt (Tullis & Albert, 2010). Auch wenn lautes Denken eine reliable Methode zur Detektion von Usabilityproblemen und Gründen für Fehlnutzungen bietet (McDonald, Zhao & Edwards, 2013), so stellt es eine Zweitaufgabe und eine Erhöhung der mentalen Beanspruchung dar. In ihrem Experiment verglichen Hertzum und Holmegaard (2015) die Schätzungen der Aufgabendauern zwischen den Gruppen mit und ohne lautes Denken. Trotz gleicher Aufgabendauern in den beiden Gruppen berichten die Autoren kürzere Schätzungen unter der Bedingung mit lautem Denken. Dies entspricht dem Interferenzeffekt, wie er von Brown (1997, 2008) beschrieben wurde.

Eine wichtige Frage ist jedoch, wie stark die Interferenz ist und von welchen Aufgabencharakteristika ihre Stärke beeinflusst wird. Im AGM wird die Aufteilung der Aufmerksamkeit zwischen der Zeitschätzung und dem Event bzw. der Zweitaufgabe als abstufbar beschrieben, wobei graduell mehr Aufmerksamkeit auf die eine oder die andere Aufgabe gelegt werden kann. Dies entspricht der Konzeption von Aufmerksamkeit als eine begrenzte und teilbare Ressource (Kahneman, 1973). Laut dem AGM sollte der Interferenzeffekt also umso stärker sein, je weniger Aufmerksamkeit auf die Zeitschätzaufgabe gelegt werden kann. Mit anderen Worten sollten Zweitaufgaben, die mehr Aufmerksamkeit benötigen, zu einem größeren Interferenzeffekt führen.

Brown (1997) untersuchte diese Vorhersage experimentell, indem er zu einer Zeitschätzaufgabe verschiedene Zweitaufgaben gab, die in den Stufen leicht und schwer absolviert werden mussten. Die Dauerschätzaufgabe bestand darin zwei Minuten lang zwei- bzw. fünfsekündige Intervalle zu produzieren. Zu den Zweitaufgaben gehörten eine motorische Trackingaufgabe, eine visuelle Suche sowie eine Kopfrechenaufgabe. Sie wurden gewählt, da sie kontinuierlich absolviert und in ihrer Schwierigkeit manipuliert werden konnten. In einer zusätz-

lichen Baselinemessung wurde die Leistung beim alleinigen Durchführen der Dauerschätzung gemessen. Die Ergebnisse zeigten für die visuelle Suche und die Kopfrechenaufgabe, dass die Produktionen länger wurden, je schwieriger die Aufgabe war (gradueller Anstieg der Schwierigkeit von keiner über leichte zu schwerer Zweitaufgabe). Für alle genutzten Zweitaufgaben zeigte sich darüber hinaus, dass die Dauerschätzungen mit steigender Zweitaufgabenschwierigkeit variabler wurden. Zusammengefasst stützen die Befunde die Vorhersagen des AGMs: Je anspruchsvoller eine Zweitaufgabe ist, umso weniger Aufmerksamkeit steht für die Dauerschätzung zur Verfügung und umso länger sind die Produktionen bzw. umso kürzer wird die Dauer verbal geschätzt.

Im Anwendungskontext von Fahrsimulationsstudien wurde diese Abstufung des Interferenzeffekt bereits genutzt, um mentale Beanspruchung indirekt zu messen. Baldauf, Burgard und Wittmann (2009) baten ihren Probanden beispielsweise ein Intervall von 17 Sek. zu produzieren, während sie eine simulierte Fahrt auf einer geraden Strecke, in einer Stadt und auf einer kurvigen Straße mit Gegenverkehr absolvierten. Die Ergebnisse zeigten, dass das Fahren in der Stadt und auf einer kurvigen Straße im Vergleich zur Fahrt auf gerader Strecke eine höhere mentale Beanspruchung auslöste (gemessen mit physiologischen und subjektiven Maßen). Wie erwartet zeigten sich für diese Fahrsituationen auch längere produzierte Dauern. Lind und Sundvall (2007) konnte den gleichen Effekt auch im Kontext von Systemevaluationen und für verbale Dauerschätzungen zeigen: Je höher die subjektiv berichtete mentale Beanspruchung war, um so stärker wurden die Interaktionsdauern unterschätzt.

Diese und ähnliche Ergebnisse (z.B. Block, 1992; Brown, 1985; Brown & Boltz, 2002; Zakay, 1993b) legen nahe, dass die Schwierigkeit der Zweitaufgabe bzw. die mentale Beanspruchung durch die Zweitaufgabe ein wesentlicher Einflussfaktor für die Stärke des Interferenzeffekts ist. Auf Basis einer Meta-Analyse fassten Block et al. (2010) die Effekte von mentaler Beanspruchung auf prospektive Dauerschätzungen wie folgt zusammen: Prospektive Dauerschätzungen werden stärker beeinträchtigt, wenn a) die Verarbeitung der Zweitaufgabe eine aktive im Vergleich zu einer passiven Verarbeitung voraussetzt, b) die Aufmerksamkeit zwischen Aufgaben geteilt werden muss im Vergleich zu einer einheitlichen Aufmerksamkeitsallokation und c) die Verarbeitungsschwierigkeit der Zweitaufgabe erhöht wird. Brown (2008) ergänzte hierzu,

dass auch Übung sowohl der Zweitaufgabe als auch der Zeitschätzaufgabe zu einer Abschwächung der Interferenz führen kann (*Abschwächungseffekt* oder *Attenuation effect*), da sich durch Übung die mentale Beanspruchung reduzieren lässt.

Eine entgegengesetzte Interferenz, das heißt eine Verschlechterung der Leistung in der Zweitaufgabe durch die Schätzaufgabe, ist jedoch nur bei bestimmten Zweitaufgaben zu finden. Grundsätzlich spricht eine bidirektionale Interferenz dafür, dass ähnliche Ressourcen bei beiden Aufgaben genutzt werden (Wickens, 2002, 2008). Brown (2006) fasste hierzu 33 verschiedene Studien zusammen und kam zu dem Ergebnis, dass vor allem solche Zweitaufgaben von der Zeitschätzung beeinträchtigt werden, die eine höhere kognitive Verarbeitung voraussetzen. Hierzu gehören Kopfrechenaufgaben, Lesen und stark aufmerksamkeitsbindende Aufgaben im Gegensatz zu wahrnehmungsbasierten Aufgaben, Mustererkennung und Tracking. Daher kann angenommen werden, dass Zeitschätzungen, ähnlich wie die eben genannten Aufgaben, primär auf Exekutivfunktionen basieren (Brown, 2008; Zakay & Block, 2004). Unter Exekutivfunktionen werden dabei solche kognitiven Prozesse verstanden, "die für die zielbezogene Kontrolle anderer, aufgabenspezifischer Prozesse verantwortlich sind" (Müsseler, 2008, S. 727). Hierzu gehören Prozesse der Handlungsplanung, der Inhibition inadaquater Reaktionen oder der Aufrechterhaltung von aufgabenrelevanten Informationen (Müsseler, 2008). Zu diesen Exekutivfunktionen kann also auch das prospektive Wahrnehmen von Dauern gezählt werden.

Ein wichtiger Punkt zum Effekt von Aufmerksamkeit auf prospektive Dauerschätzungen, der bis jetzt unerwähnt geblieben ist, sind die Zeitintervalle, die in den diskutierten Arbeiten betrachtet wurden. Diese reichen von Millisekunden (Mattes & Ulrich, 1998) über wenige Sekunden (Brown, 1997, 2006) bis hin zu Minuten (Block et al., 1980; Block, 1992). Die Spanne zeigt auf, dass der Effekt von Aufmerksamkeit auf sehr verschiedenen Zeitskalen gefunden werden kann und anders als der Effekt von Arousal nicht spezifisch für bestimmte Intervalle ist.

Zusammengefasst legen die Befunde nahe, dass prospektive Dauerschätzungen eng mit Prozessen der Aufmerksamkeitssteuerung verbunden sind. Dabei ist es ähnlich wie bei Geschmack oder Schmerzen: Die Zuwendung der Auf-

merksamkeit zu der Wahrnehmung hin verstärkt das Erlebnis (Brown, 2008). Dadurch entsteht eine Verlängerung der erlebten Zeit, je mehr Aufmerksamkeit dem Verstreichen von Zeit geschenkt wird. Dies steht mit den Vorhersagen des AGM im Einklang und muss nicht, anders als der Effekt des Arousal, in seiner Gültigkeit für bestimmte Zeitintervalle eingeschränkt werden. Darüber hinaus zeigen die hier diskutierten Befunde, dass Dauerschätzungen eine höhere kognitive Verarbeitung (vgl. Exekutivfunktionen) beanspruchen.

3.3 Zwischenfazit zur erlebten Zeit

Die Fähigkeit Dauern zu schätzen ist eine kognitive Leistung, die mit vielen anderen kognitiven Prozessen verbunden ist. Hierzu gehören die Wahrnehmung und Verarbeitung der Stimuli, Emotionen, die Aufmerksamkeitsallokation, die mentale Beanspruchung und andere Exekutivfunktionen. Die vorgestellten Modelle zu prospektiver Dauerwahrnehmung beschreiben die kognitiven Prozesse, die für eine Dauerschätzung nötig sind und berücksichtigen dabei Teile der eben genannten Ko-Prozesse bzw. Einflussfaktoren. Zusammengefasst postulieren die Modelle einen internen Taktgeber, der für prospektive Dauerschätzungen genutzt wird und auf dessen Basis sich skalare Eigenschaften von Dauerschätzungen ableiten lassen. Darüber hinaus ist der Taktgeber in seiner Frequenz sensitiv gegenüber dem Arousal-Level. Die Registrierung der Impulse des Taktgebers ist hingegen abhängig vom Ausmaß der Aufmerksamkeit, die auf die Dauerschätzung gerichtet wird. Diese Modellvorhersagen lassen sich weitestgehend empirisch stützen. Der Einfluss des Arousal-Levels scheint aber nur für kurze Dauern relevant zu sein, da er bei Dauern über zwei Sekunden selten nachgewiesen werden konnte.

Teil des Forschungsziels der vorliegenden Arbeit ist es zu untersuchen, in wie weit die Dauer einer Interaktion als ein Interaktionscharakteristikum von Nutzern wahrgenommen werden kann bzw. wann und ob es zu Verzerrungen kommt. Die vorgestellten Modelle und empirischen Arbeiten beziehen sich nicht explizit auf eine solche Anwendungsdomäne. Daher ist es nicht klar, wie gut sich die Annahmen und Befunde auf den Bereich Mensch-Computer-Interaktion übertragen lassen und wie gut sie in diesem Bereich verlässliche Vorhersagen treffen können. Auf Basis der im vorliegenden Kapitel diskutie-

ren „timing-with-a-timer“-Modellen (Block & Zakay, 1996) und Befunden zu prospektiven Dauerschätzungen können jedoch Erwartungen an die Dauerwahrnehmung eines Nutzers während einer technischen Interaktion abgeleitet werden:

- Auf Grund der **skalaren Eigenschaften** (Vergleich Information-Process Model of Timing und STT) kann erwartet werden, dass Zeitdauern von Interaktionen grundsätzlich wahrgenommen und geschätzt werden können. Längere Dauern sollten entsprechend zu längeren Schätzungen führen. Dies gilt für Dauern im Minuten-Bereich, wie er im Kontext Mensch-Computer-Interaktion von Interesse ist, wie auch für Dauern im Sekunden-Bereich (Droit-Volet et al., 2018). Die Schätzungen sollten jedoch mit steigender Interaktionsdauer eine größere **Varianz** aufweisen.
- **Emotionen** werden häufig während einer technischen Interaktion erlebt (Backhaus & Brandenburg, 2014; Brandenburg & Backhaus, 2016; Brave & Nass, 2003; Jokinen, 2015; Lewis & Mayes, 2014). Sie bilden eine wesentliche Komponente des Nutzererlebens, welche im Kapitel 2 bereits diskutiert wurde. Auf Basis der taktgeber-basierten Modelle könnte eine Veränderung der Dauerschätzungen bei einer Veränderung des emotionalen Status des Nutzers erwartet werden. Ob Emotionen, die während einer mehrminütigen Mensch-Computer-Interaktion erlebt werden, tatsächlich die Schätzung der Interaktionsdauer beeinflussen, ist aber fragwürdig. Auf Basis der hier diskutierten Studien kann angenommen werden, dass die Schnelligkeit des Taktgebers bei Intervallen über zwei Sekunden keinen wesentlichen Einfluss auf die Schätzung mehr hat, da Prozesse der Aufmerksamkeit sowie des Arbeitsgedächtnisses eine zunehmend wichtige Rolle übernehmen und die Veränderungen im Affekt meist zu schwach sind.
- **Aufmerksamkeitsprozesse** spielen ebenfalls eine wichtige Rolle während der Mensch-Computer-Interaktion. Um sein Interaktionsziel zu erreichen, muss der Nutzer dieses im Gedächtnis halten, während er es gleichzeitig in Unterschritte gliedert, die er in Zusammenarbeit mit dem System bearbeitet und überwacht. Dabei konfrontiert ihn sowohl die Ausführung und Planung der Teilschritte als auch das Design der Interakti-

on mit variierender **mentaler Beanspruchung**. In Untersuchungen zur Usability oder zum Nutzererleben bei Mensch-Computer-Interaktionen wird die mentale Beanspruchung darüber hinaus manchmal durch Zweitaufgaben, wie zum Beispiel lautes Denken, zusätzlich erhöht. Auf Basis der in diesem Kapitel diskutierten Arbeiten und des AGM sollten Dauerschätzungen durch variierende mentale Beanspruchung und durch Zweitaufgaben beeinflusst werden. Auch wenn es hierzu im Kontext Mensch-Computer-Interaktion bereits erste Arbeiten gibt (Hertzum, Hansen & Andersen, 2009; Hertzum & Holmegaard, 2015; Lind & Sundvall, 2007), ist die Befundlage noch sehr begrenzt.

Um das Forschungsziel zu erreichen konzentriert sich die vorliegende Arbeit daher auf die Prüfung der genannten Punkte (skalare Eigenschaften, Emotion und Aufmerksamkeit) im prospektiven Paradigma. Durch die Untersuchung der Eigenschaften von Dauerschätzungen im Anwendungskontext können Aussagen darüber getroffen werden, wie Nutzer Dauern wahrnehmen und welche Bedeutung die Dauerwahrnehmung für das Nutzererleben hat.

Kapitel 4

Zeitempfinden

Wie in der Einleitung bereits argumentiert, ist die menschliche Dauerwahrnehmung mehr als nur das reine Schätzen von Dauern, da sie häufig zu einer kognitiven Bewertung führt. Wearden (2015) bezeichnet eine solche Bewertung als *Zeitempfinden* (*feeling about time*) und nennt das Gefühl von sich ziehender Zeit, während man in einer Warteschlange steht, als ein Beispiel. Dieses Beispiel zeigt auch bereits einen Unterschied zwischen Zeitempfinden und Dauerschätzungen: Während ein Kunde wissen kann, dass eine Wartezeit exakt fünf Minuten betrug, kann er trotzdem das Gefühl äußern, dass ihm die Dauer länger erschien (Wearden et al., 2014, S. 298). Zeitempfinden ist damit nicht nur abstrakter und schwerer zu erfassen als eine Dauerschätzung, sondern es scheint auch näher an dem Erleben einer Situation zu sein. Für das vorliegende Forschungsziel ist dies besonders interessant, da auf Grund der Nähe zum Erleben von einem Bezug zum Nutzererleben ausgegangen werden kann. Aus diesem Grund wird das Zeitempfinden neben Dauerschätzungen als zentrales Element zur Bearbeitung des Forschungsziels verstanden.

Aber wie kann das Zeitempfinden operationalisiert werden? Gorn et al. (2004) nutzten eine Variable namens *wahrgenommene Schnelligkeit* (*perceived quickness*), die mit Hilfe dreier semantischer Differentiale gemessen wurde (slow – fast, not speedy – speedy, not quick – quick). Ähnlich dazu maßen Tanaka und Yotsumoto (2017) die Zeitflussbeurteilung, indem sie fragten, wie schnell die Zeit während einer bestimmten Aktivität auf einer 100-stufigen, visuellen Analogskala vergangen sei. Droit-Volet und Wearden (2016) nutzen die Frage, wie die Zeit im Vergleich zur Uhrzeit vergangen sei, wobei sich die

Skalenendpunkte von viel langsamer zu viel schneller erstreckten. Neben der Operationalisierung als wahrgenommene Schnelligkeit sind jedoch auch Items zu finden, die explizit nach dem Ziehen oder Verfliegen einer erlebten Zeitdauer fragen (Sackett, Meyvis, Nelson, Converse & Sackett, 2010; Sucala, Scheckner & David, 2011). Unabhängig von der Art der Operationalisierung werden solche Maße als *Passage of Time Judgements (PoTJs)* bezeichnet, was mit Zeitflussbeurteilungen übersetzt werden kann. Sie werden in Studien genutzt, in denen Versuchspersonen lediglich warten (z.B. Gorn et al., 2004; Lee et al., 2012) oder aber während des Intervalls einer Tätigkeit nachgehen (z.B. Sackett et al., 2010; Tanaka & Yotsumoto, 2017). Eine einheitliche Operationalisierung von Zeitflussbeurteilungen gibt es nicht. Dies zeigt sich sowohl in der Wortwahl der Items als auch in der den Items zugrundeliegenden Skalen: Sie reichen von fünf- (Sucala et al., 2011) oder sieben-stufigen Likert-Skalen (Droit-Volet & Wearden, 2016; Sackett et al., 2010) bis hin zu neun- (Gorn et al., 2004) oder 100-stufigen (Tanaka & Yotsumoto, 2017) semantischen Differentialen und visuellen Analogskalen.

Ein anderer Zugang zum Zeitempfinden kann in der Bewertung der Dauer gesehen werden. Wenn beispielsweise das Speichern von Änderungen in einem Dokument mehr als 30 Sek. dauert, kann sich diese Dauer als zu lange oder als unangemessen anfühlen. Dies spiegelt anders als PoTJs nicht das Vergehen der Dauer wider, sondern eine abschließende Bewertung. Solche Bewertungen wurden für Wartedauern bereits untersucht (Antonides, Verhoef & van Aalst, 2002; Dellaert & Kahn, 1998; Kurusathianpong & Tangmanee, 2018). Hier werden zur Messung ebenfalls semantische Differentiale oder visuelle Analogskalen mit unterschiedlich benannten Polen (long – short, unacceptable – acceptable, poor – good) bzw. Zustimmungsskalen genutzt. Ähnlich wie bei Zeitflussbewertungen können Dauerbewertungen aber nicht nur nach Wartedauern abgefragt werden, sondern auch nach gefüllten Intervallen, in denen die Versuchsperson einer Tätigkeit nachgeht. Ein Beispiel hierfür sind Mensch-Computer-Interaktionen. Wie in Kapitel 2 diskutiert wurde, kann hier die Dauerbewertung einen neuen Blickwinkel auf das Nutzererleben erlauben, welcher bis jetzt kaum untersucht wurde.

Zusammengefasst lassen sich in Bezug auf das Zeitempfinden zwei mögliche Dimensionen unterscheiden, Dauerbewertungen und Zeitflussbeurteilungen

(PoTJs). Im Gegensatz zu PoTJs beziehen sich Dauerwertungen nicht auf den Zeitfluss an sich, sondern auf die Bewertung der Dauer nach der Beendigung der Aufgabe. Sie stellen damit eine Art summative Evaluation dar. Während Zeitflussbewertungen bereits im Rahmen verschiedener Publikationen und Tätigkeiten untersucht wurden, sind Arbeiten zu Dauerbewertungen wesentlich seltener. Im Bereich Mensch-Computer-Interaktion werden sie teilweise als Operationalisierung genutzt, um die wahrgenommene Wartezeit zu untersuchen (Kurusathianpong & Tangmanee, 2018), aber auch hier gibt es kaum Untersuchungen, die sich systematisch mit der Entstehung und der Bedeutung von Dauerbewertungen im Kontext von UX beschäftigen. Im folgenden werden beide Konstrukte auf Grund ihrer konzeptuellen Trennung unabhängig voneinander diskutiert.

4.1 Dauerbewertungen

Aufgabendauern werden in Usability- und UX-Testungen häufig objektiv gemessen (Albert et al., 2009; Frøkjær et al., 2000; Lesemann et al., 2007), eine subjektive Dauerbewertung wird jedoch in der Regel nicht beachtet. Einige Fragebögen nutzen zwar Items, die zeitliche Bewertungen beinhalten, sie wollen damit aber Effizienz, Erlernbarkeit oder Verzögerungen des Systems messen (vgl. PSSUQ: Lewis, 1992; QUIS: Chin, Diehl & Norman, 1988; UEQ: Laugwitz, Schrepp & Held, 2006, USE: Lund, 2001). Ein Beispiel für ein solches Item stammt aus dem PSSUQ: „I was able to complete the tasks and scenarios quickly using this system” (Lewis, 1992). Liikkanen und Gómez Gómez (2013) argumentieren jedoch, dass die subjektive Bewertung einer erlebten Dauer durch den Nutzer eine wesentlich wichtigere Rolle bei der Evaluation von Systemen spielt als objektiv gemessene Dauern. Solche subjektiven Dauerbewertungen werden durch das Erleben der Zeit, Erwartungen, Vorerfahrungen und den Zustand des Nutzers beeinflusst (Liikkanen & Gómez Gómez, 2013) und sind damit sowohl von Dauerschätzungen als auch von Zeitflussbeurteilungen abzugrenzen. Im Kontext von Mensch-Computer-Interaktionen wurden Dauerbewertungen bis jetzt kaum untersucht. Sie sind allerdings im Rahmen von Wartedauern eine häufig genutzte Messvariable (Antonides et al., 2002; Dellaert & Kahn, 1998; Kurusathianpong & Tangmanee, 2018).

Kurusathianpong und Tangmanee (2018) betrachteten beispielsweise, wie sich die Gestaltung eines Wartebalkens auf die Bewertung der Wartezeit auswirkt. Die Autoren erfassten diese Dauerbewertung über die Items *Boring*, *Unsatisfactory*, *Irritating*, *Too slow* und *Unacceptable*, die auf einer siebenstufigen Zustimmungsskalen bewertet wurden. Hier fällt auf, dass die Operationalisierung von Dauerbewertungen nicht trennscharf von emotionsnahen Konstrukten erfolgte. Zwar beschreiben Items zur Langeweile (*Boring*) oder Verärgerung (*Irritating*) keine Emotionen, sie werden aber im UX-Kontext häufig zur Emotionsmessung verwendet (z.B. im Fragebogen meCUE – “Das Produkt verärgerst mich.”, Minge & Riedel, 2013). Auch andere Autoren (Antonides et al., 2002) nutzen nicht nur dauerbezogene Attribute zur Bewertung (long – short, unacceptable – acceptable), sondern auch emotionsnahe Items (z.B. annoying – pleasant, boring – varied). In der vorliegenden Arbeit soll der Zusammenhang zwischen Dauerbewertungen und Komponenten des Nutzererlebens untersucht werden. Da Emotionen eine wichtige UX-Komponente darstellen, muss eine klare Trennung zwischen den zu messenden Konstrukten in der Operationalisierung erfolgen. Daher werden in dieser Arbeit nur dauerbezogene Items zur Erfassung der Dauerbewertung genutzt.

Trotz der fehlenden Trennschärfe in der Operationalisierung, können die Studien zur Wartezeit einen ersten Einblick geben, in wie weit Dauerbewertungen mit Dauerschätzungen zusammenhängen. Antonides et al. (2002) fanden beispielsweise nur in einer von zwei Studien Evidenz dafür, dass die Dauerschätzung ein signifikanter Prädiktor der Dauerbewertung war: Je länger die Wartezeit geschätzt wurde, umso negativer wurde sie bewertet. Kurusathianpong und Tangmanee (2018) erhoben zwar keine Dauerschätzung als zusätzliches Maß neben der Dauerbewertung, sie fanden aber einen Effekt der Größe des Wartebalkens auf die Dauerbewertung. Genauer gesagt wurde die Wartezeit bei einem größeren Wartebalken als negativer bewertet. Aus anderen Studien ist bekannt, dass die Präsentationsdauer von größeren Objekten meist als länger eingeschätzt wird, als die von kleineren Objekten (Ono & Kawahara, 2007; Rammsayer & Verner, 2014; Xuan, Zhang, He & Chen, 2007). Aus diesen Studien kann also die Hypothese abgeleitet werden, dass sich auch im UX-Kontext ein Zusammenhang zwischen Schätzung und Bewertung einer Interaktionsdauer zeigen sollte. Wie groß und stabil dieser Zusammenhang je-

doch erwartet werden kann, ist unklar, da in den genannten Studien zum einen Dauerbewertungen teilweise über emotionsnahe Konstrukte operationalisiert wurden und zum anderen ausschließlich Wartezeiten betrachtet wurden und keine Interaktionen, in denen der Nutzer aktiv eine Tätigkeit verfolgt. Auch wie sich Dauerbewertungen in das Entstehen von UX einbetten, ist bisher nicht empirisch untersucht worden.

4.2 Zeitflussbeurteilungen - Passage of Time Judgements

Zeitflussbeurteilungen oder auch Passage of Time Judgements (PoTJs) als Konzeptualisierung von Zeitempfinden wurden in den letzten Jahren primär bezüglich zweier Aspekte untersucht. Hierzu gehören zum einen Zusammenhänge zu prospektive Dauerschätzungen und zum anderen Zusammenhänge zu Erlebenskomponenten der Situation, zu der der Zeitfluss beurteilt wird.

Bezüglich des Zusammenhangs zwischen prospektiven Dauerschätzungen und PoTJs gibt es bereits einige empirische Studien sowie erste Modellannahmen. Droit-Volet und Wearden (2016) nutzen beispielsweise eine Experience Sampling Methode, um den Zusammenhang im Feld zu untersuchen. Dazu wurden Probanden acht mal am Tag durch einen Alarm dazu aufgefordert, eine präsentierte Dauer von unter zwei Sekunden verbal oder durch eine Produktion zu schätzen. Danach bewerteten sie, ob die Zeit zum Zeitpunkt des Alarms schnell oder langsam verging. Die Ergebnisse zeigten keine Korrelationen zwischen prospektiver Dauerschätzung und Zeitflussbeurteilung. Dieses Ergebnis wurde von Droit-Volet, Trahanias und Maniadakis (2017) für Dauern bis zu 32 Sek. repliziert. In beiden Studien bezog sich die Zeitflussbeurteilung jedoch auf die Aktivität vor dem Alarm, wohingegen die Dauerschätzung eine Unterbrechung dieser Aktivität darstellte. Daher liegt es nahe anzunehmen, dass sich die Angaben der Probanden an zwei verschiedenen, voneinander unabhängigen Tätigkeiten orientierten, wodurch kein Zusammenhang zu erwarten wäre.

In einer weiteren Studie verlängerten Droit-Volet et al. (2017) das zu schätzende Intervall auf bis zu acht Minuten, welches durch zwei Töne markiert wurde und verbal geschätzt werden sollte. Danach wurde der Zeitfluss beurteilt. In diesem Szenario ist es wahrscheinlich anzunehmen, dass die Probanden

ihre Aktivität zum Zeitpunkt des Alarms auch während des Schätzintervalls fortsetzten, wodurch sich beide Maße auf die gleiche Tätigkeit bezogen. Hier zeigte sich eine signifikante Korrelation zwischen Dauerschätzungen und Zeitflussbeurteilung: Je länger die Dauer eingeschätzt wurde, desto langsamer wurde der Zeitfluss bewertet. Dieser Zusammenhang wurde von anderen Autoren auch für vierminütige Intervalle (Sucala et al., 2011) und zwei- bis 32-minütige Intervalle (Droit-Volet et al., 2018) repliziert und stützt die Annahme, dass Dauerschätzungen und Zeitflussbeurteilungen zumindest bei mehrminütigen Intervallen korrelieren.

Sucala et al. (2011) untersuchten jedoch nicht nur den Zusammenhang zwischen den beiden Maßen. Sie testeten auch, ob beide Maße durch ähnliche Faktoren beeinflusst werden. Dazu gaben die Autoren ihren Probanden eine Leseaufgabe und manipulierten die benötigte Verarbeitungstiefe sowie die zeitliche Relevanz der Aufgabe. Der Faktor Verarbeitungstiefe unterschied sich zwischen oberflächlicher und tiefer Verarbeitung (Detektion von Worten mit dem Anfangsbuchstaben S vs. Detektion und zusätzliches Finden eines Synonyms). Die zeitliche Relevanz wurde über ein Zeitlimit bzw. kein Zeitlimit für die Aufgabe realisiert. Da die Variation der beiden Faktoren sowohl auf Dauerschätzungen als auch auf die Zeitflussbeurteilung den gleichen Einfluss zeigte, schlussfolgerten die Autoren, dass beide Maße durch Prozesse der Aufmerksamkeit beeinflusst wurden (Sucala et al., 2011).

Droit-Volet et al. (2018) stützen diesen Schluss und konkretisieren im Rahmen ihrer *Theory of explicit time judgement of very long durations*, dass das Schätzen der Dauer von mehrminütigen Intervallen zwar auf Prozessen eines internen Taktgebers basiert (vgl. Information Process Model, STT, AGM), hier aber auch Prozesse der exekutiven Kontrolle eine starke Rolle spielen. Sie argumentieren, dass bei der Schätzung mehrminütiger Intervalle eine Unterteilung des Intervalls in mehrere Ereignisse erfolgt. Die Dauer jedes dieser Ereignisse wird mit Hilfe des Taktgebers und des Arbeitsgedächtnisses enkodiert, wobei die geschätzte Dauer eines vergangenen Ereignisses ins Langzeitgedächtnis überführt wird. Mit anderen Worten besteht die Schätzung eines mehrminütigen Intervalls aus vielen kleinen Schätzungen. Die Zeitflussbeurteilung wird erst mit steigender Anzahl an Ereignissen bewusst. Das heißt, je mehr Ereignisse abgespeichert wurden, umso stärker bzw. salienter wird dieses Bewusstsein.

Mit steigendem Bewusstsein wiederum erscheint die Zeit langsamer zu vergehen und es wird mehr Aufmerksamkeit auf die Zeitschätzaufgabe gerichtet, wodurch längere Dauerschätzungen entstehen. Zusammengefasst argumentieren Droit-Volet et al. (2018) also, dass Probanden auf Nachfrage zwar immer eine Aussage über ihre aktuelle Zeitflussbeurteilung geben können, dass der Zeitfluss jedoch vor allem bei mehrminütigen Dauern von alleine bewusst wird und dann einen verlängernden Einfluss auf die Dauerschätzung hat. Durch die beschriebenen Modellannahmen kann die Korrelation bei mehrminütigen Intervallen erklärt werden. Das Modell erklärt jedoch nicht die Entstehung einer Zeitflussbeurteilung, da es dieses lediglich als Einflussfaktor betrachtet.

Eine alternative Erklärung für den Zusammenhang zwischen erlebter Dauer und Zeitflussbeurteilung (PoTJs) liefern Tanaka und Yotsumoto (2017). Sie argumentieren, dass Zeitflussbeurteilungen aus der Diskrepanz zwischen erlebter Zeit und zeitlicher Erwartung resultieren. In einer Reihe von Studien mit n-back Aufgaben und Aufgabendauern zwischen drei und sieben Minuten konnten die Autoren zeigen, dass die Zeitflussbeurteilung systematisch mit der Differenz zwischen erwarteter und erlebter Dauer variiert. Zusammengefasst wird in beiden Erklärungsansätzen erlebte Zeit im Sinne einer Dauerschätzung als ein zentrales Konstrukt für die Zeitflussbeurteilung beschrieben: Laut Droit-Volet et al. (2018) muss eine gewisse Dauer bereits erlebt sein, damit der Zeitfluss bewusst wird. Laut Tanaka und Yotsumoto (2017) entsteht der Zeitfluss unter anderem auf Basis der erlebten Zeit. Die erwartete Dauer, die bei Tanaka und Yotsumoto (2017) als zweite Antezedenz jedoch genannt wird, spielt im Ansatz von Droit-Volet et al. (2018) keine Rolle. Hier sind Droit-Volet et al. (2018) viel unkonkreter, wie Zeitfluss entsteht, da sie mehr auf die Einflussfaktoren von Dauerschätzungen bei mehrminütigen Intervallen eingehen als auf die Erklärung der Entstehung des Zeitflusses. Der Zusammenhang zwischen Zeitflussbeurteilung und Dauerbewertung zeigt zwar, dass Dauern mit dem Zeitfluss verknüpft sind, allerdings bleibt die Frage offen, welche anderen Aspekte einer Situation dafür sorgen, dass eine spezielle Zeitflussbeurteilung entsteht.

Hier können Arbeiten weiterhelfen, die die Frage aufwerfen, welchen Zusammenhang es zwischen Zeitflussbeurteilung und Erleben gibt. In zwei von der Autorin der vorliegenden Arbeit betreuten Masterarbeiten wurden Pro-

banden nach ihrer Zustimmung zur Aussage "Die Zeit verfliegt, wenn man Spaß hat"gefragt (Kim, 2017; Schorling, 2018). Von insgesamt 122 Probanden zeigten 120 mittlere bis hohe Zustimmungswerte, was aufzeigt, dass das Erleben des Zeitflusses mit positiven Aspekten einer Aktivität oder eines Erlebnisses verknüpft wird. Wearden et al. (2014) untersuchten dieses Phänomen, indem sie Probanden den gleichen Filmclip vier mal vorspielten. Während sich die Dauerschätzungen der einzelnen Präsentationen nicht voneinander unterschieden, zeigten die Zeitflussbeurteilungen, dass die Zeit mit jeder Wiederholung langsamer zu vergehen schien und dass jedes erneute Schauen sich länger anfühlte, als das Mal zuvor (keine inferenzstatistische Analyse berichtet). Darüber hinaus ließen positive Erlebenskomponenten (Vergnügen, Spannung und Gefallen) nach, während negative Komponenten (Langeweile und Unmut) zunahmen. Schaut man auf emotionale Aspekte der Situation, gibt es auch Evidenz für einen korrelativen Zusammenhang zwischen Zeitflussbeurteilung und verschiedenen Maßen für Freude/positiven Affekt, Traurigkeit/negativen Aspekt, Arousal und Entspannung (Droit-Volet & Wearden, 2015, 2016; Droit-Volet et al., 2017, 2018).

In einer Serie aus verschiedenen Studien gingen Sackett et al. (2010) noch einen Schritt weiter, indem sie die Zeitflussbeurteilungen manipulierten und Auswirkungen auf verschiedene Erlebenskomponenten untersuchten. In einem ihrer Paradigmen gaben sie Probanden falsche Informationen über eine Aufgabendauer, wodurch die Probanden eine Aufgabendauer von 5 bzw. 20 Minuten erwarteten, jedoch eine Dauer von 10 Minuten erlebten. Sie schufen hier also eine Diskrepanz zwischen erwarteter und erlebter Dauer (vgl. Tanaka & Yotsumoto, 2017). Danach wurden die Zeitflussbeurteilung als Manipulationscheck genutzt und die Aufgabe bezüglich hedonischer Qualitäten bewertet. Die Ergebnisse zeigten, dass Probanden mit einer zu langen Erwartung (20 Minuten) die Zeit als verfliegend, während die anderen (5 Minuten) die Zeit als ziehend bewerteten. Darüber hinaus hatte die Manipulation einen Effekt auf das Aufgabenengagement, auf den Genuss (enjoyability) von Liedern, die während der Aufgabe gespielt wurden, und wie nervend Lärmgeräusche erlebt wurden: Je schneller die Zeit verging, desto positiver waren die Bewertungen. Dieses Befundmuster replizierten Sackett et al. (2010) in mehreren Studien. Darüber hinaus zeigten sie, dass sich die manipulierte Zeitflussbeurteilung nur

dann auf Bewertungen der Situation auswirkte, wenn Probanden an den Zusammenhang „Wenn ich Spaß habe, verfliegt die Zeit.“ glaubten und keine alternative Erklärung für ihr Zeitflussempfinden hatten.

Zusammengefasst zeigen bisherige Studien, dass PoTJs und Dauerschätzungen miteinander korrelieren, wenn sich beide Maße auf die gleiche Aktivität beziehen und diese Aktivität wenige Minuten dauert. In diesen Fällen können Dauern, die kürzer geschätzt werden, mit einem schnelleren Zeitfluss assoziiert werden. Darüber hinaus wird das Erleben einer schnell vergehenden Zeit häufig mit positiven Emotionen oder einem positiveren Erleben verknüpft. An diesen Zusammenhang scheinen Probanden so stark zu glauben, dass sie Situationen teilweise als angenehmer bewerten, wenn Ihnen der Eindruck vermittelt wird, dass die Zeit verflogen sei.

4.3 Zwischenfazit zum Zeitempfinden

Das Forschungsziel der vorliegenden Arbeit liegt in der Untersuchung der Dauerwahrnehmung des Nutzers während einer technischen Interaktion sowie in der Klärung der Bedeutung dieser Wahrnehmung für die Bewertung der Interaktion. Zeitempfinden wird hier als ein Aspekt der Dauerwahrnehmung verstanden. Es kann durch Zeitflussbeurteilungen sowie durch Dauerbewertungen abgebildet werden. Für beide Operationalisierungen stellt sich zum einen die Frage, wie sich zur Wahrnehmung der Interaktionsdauer (Dauerschätzung) verhalten. Die diskutierten Arbeiten haben gezeigt, dass das zugehörige Forschungsfeld noch relativ neu und unvollständig ist. Zu dieser ersten Frage gibt es daher bis jetzt nur Befunde für Zeitflussbeurteilungen. Zum anderen stellt sich für beide Konstrukte die Frage, in welchem Verhältnis sie zum Nutzererleben stehen. Auch hier gibt es zwar Befunde, die Zusammenhänge zwischen Zeitflussbeurteilungen und Erlebenskomponenten zeigen konnten, im Anwendungskontext Mensch-Computer-Interaktion haben beide Konstrukte bis jetzt jedoch noch wenig Beachtung gefunden. Daher werden die folgenden Erwartungen abgeleitet:

- Zeitempfinden kann auf verschiedene Arten **operationalisiert** werden. Im Fokus dieser Arbeit stehen **Zeitflussbeurteilungen** und **Dauerbewertungen**.

- Zeitflussbeurteilungen und Dauerbewertungen **kovariieren** im Rahmen einer technischen Interaktion mit der **Schätzung von Dauern**. Darüber hinaus sollten sie durch ähnliche Charakteristika der Interaktion beeinflusst werden. Eine solcher Befunde würde den Zusammenhang zwischen Zeitflussbeurteilung und Dauerschätzungen weiter stützen und diese Annahmen auch für Dauerbewertungen verallgemeinerbar machen.
- Zeitflussbeurteilungen und Dauerschätzungen **kovariieren** mit verschiedenen Komponenten des **Nutzererlebens**. Für **Zeitflussbeurteilungen** ist bereits bekannt, dass sie mit Erlebenskomponenten der Situation in Verbindung gebracht werden. Ob dies auch für Erlebenskomponenten während einer technischen Interaktion gilt und mit welchen Dimensionen dies der Fall ist, ist eine offene Frage, die im Rahmen dieser Arbeit beantwortet wird. Für **Dauerbewertungen** gibt es bis jetzt weit weniger Studien und Indizien, dass sie mit dem Erleben verknüpft sind. Sie werden jedoch im HCI-Kontext als Maße für Konstrukte genutzt, die die instrumentellen Qualitäten einer Interaktion widerspiegeln. Darüber hinaus werden sie von manchen Autoren als wesentlicher Bestandteil des Erlebens einer Interaktion diskutiert. Daher wird die empirische Basis auch für dieses Konstrukt erweitert.

Kapitel 5

Forschungsfragen und Überleitung zu den empirischen Arbeiten

Das CUE-Modell (Mahlke, 2008; Minge & Thüring, 2018; Thüring & Mahlke, 2007) stellt ein wichtiges Rahmenmodell für UX dar (siehe Kap. 2), welches UX in verschiedene Komponenten zerlegt und Antezedenzen und Konsequenzen benennt und strukturiert. Zu den Antezedenzen gehören das System, der Kontext und der Nutzer. Diese beeinflussen jedoch die Komponenten des Nutzererlebens nicht direkt, sondern führen zu spezifischen Interaktionscharakteristika, die die Grundlage der Wahrnehmung der Komponenten von UX für den Nutzer bilden. Diese Interaktionscharakteristika sind bis jetzt unterspezifiziert. So werden beispielsweise für die Konsequenzen von UX verschiedene Unterkonstrukte genannt, wie die Gesamtbewertung oder das Nutzungsverhalten (Mahlke, 2008), für Interaktionscharakteristika hingegen gibt es keine solche Ausdifferenzierung.

Im Rahmen dieser Arbeit wird daher die Wahrnehmung eines Interaktionscharakteristikums beispielhaft näher untersucht – die Interaktionsdauer. Wie Nutzer Interaktionsdauern wahrnehmen, wann es zu Verzerrungen in ihrer Wahrnehmung kommt und welche Rolle diese Wahrnehmung für das Entstehen des Nutzererlebens spielt, wurde bis jetzt kaum untersucht. Manche Autoren argumentieren jedoch, dass subjektiv erlebte Zeit beim Gestalten technischer Interaktionen berücksichtigt werden sollte, da sie ein wichtiges Kriterium für ein positives Nutzererleben ist (Liikkanen & Gómez Gómez, 2013). Subjektiv erlebte Zeit nach Liikkanen und Gómez Gómez (2013) wird in der vorliegenden

Arbeit mit sogenannten Dauerbewertungen gleichgesetzt, welche in Kapitel 4 beschrieben wurden. Dauerbewertungen stellen jedoch nur eine Operationalisierung für die Wahrnehmung von Interaktionsdauern dar, die wie in Kapiteln 3 und 4 beschrieben auch Dauerschätzungen und Zeitflussbeurteilungen umfassen kann. In Abbildung 5.1 ist der Zusammenhang zwischen diesen Konstrukten veranschaulicht.

Das Ziel dieser Arbeit ist es Interaktionsdauern hinsichtlich dieser verschiedenen Dimensionen näher zu untersuchen und ihre Bedeutung für das Nutzererleben zu klären (vgl. Forschungsziel, Kap.1). Zusammengefasst dient dies der Beantwortung der folgenden Forschungsfragen:

Forschungsfrage 1: In wie weit können Nutzer Interaktionsdauern wahrnehmen?

Forschungsfrage 2: Welchen Zusammenhang gibt es zwischen der Dauerwahrnehmung und dem Nutzererleben?

Die erste Forschungsfrage bezieht sich auf das Verhältnis der Interaktionsdauer zur wahrgenommenen Dauer bezüglich der drei beschriebenen Operationalisierungen. Zur Beantwortung dieser Frage werden in den folgenden Studien sowohl Unterschieds- als auch Zusammenhangshypothesen getestet. Unterschiedshypothesen beziehen sich auf die Effekte verschiedener Aspekte der Interaktion auf die Interaktionsdauer und ihre Wahrnehmung. Hier kann beispielsweise erwartet werden, dass eine Veränderung des Systems (z.B. über Usability) oder eine Zweitaufgabe (z.B. lautes Denken) die Interaktionsdauer beeinflusst. In Studien I und II werden diese Aspekte systematisch variiert, um zu testen, ob ihr Einfluss auf die Interaktionsdauer vergleichbar mit dem



ABBILDUNG 5.1: Operationalisierung der Wahrnehmung von Interaktionsdauern.

auf die Wahrnehmung der Interaktionsdauer ist (siehe Abb. 5.2). Aus den Ergebnissen der beiden ersten Studien leitet sich bezüglich Forschungsfrage 1 Studie IV ab, in der die Einflüsse der Interaktionsdauer gegen die Einflüsse anderer Interaktionscharakteristika auf die Wahrnehmung der Interaktionsdauer abgewägt wird. Darüber hinaus wird durch Zusammenhangshypothesen im Rahmen von Forschungsfrage 1 der korrelative Zusammenhang zwischen Interaktionsdauer und Wahrnehmung sowie zwischen den Operationalisierungen der Dauerwahrnehmung in Studien I, II, III und IV geprüft.

Zur Beantwortung von Forschungsfrage 2 stehen Korrelationen zwischen der

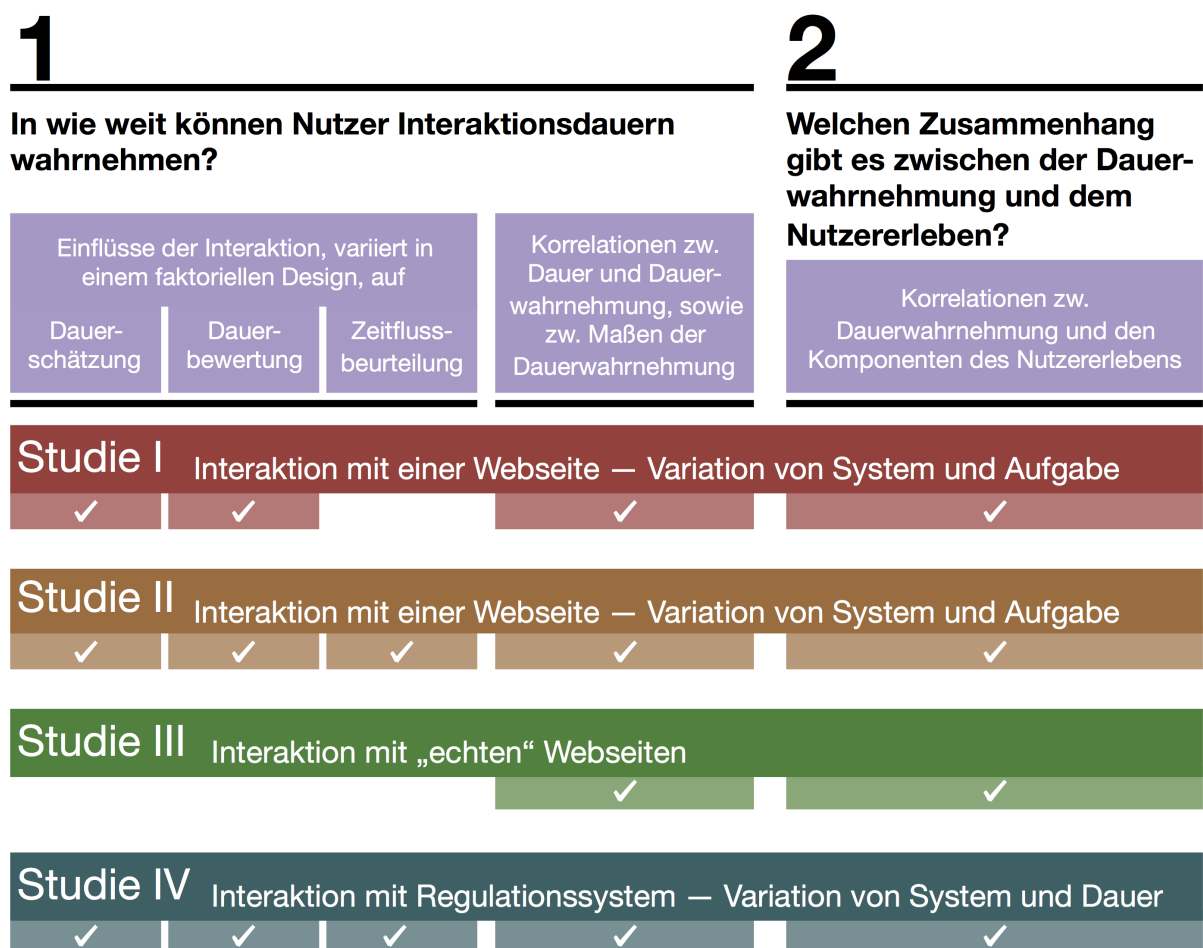


ABBILDUNG 5.2: Übersicht über die Studien I bis IV mit ihrem Bezug zu den Forschungsfragen. Die Häkchen bedeuten, dass die entsprechende Forschungsfrage mit der jeweiligen Operationalisierung in dieser Studie thematisiert wird. Die Farbgestaltung der einzelnen Boxen wird in der Beschreibung der Studien in den folgenden Kapiteln jeweils als Grundfarbe der Ergebnisdarstellung genutzt.

Dauerwahrnehmung mit Komponenten des Nutzererlebens im Fokus, welche in allen Studien untersucht werden. Hier wird betrachtet, welche zeitlichen Variablen mit welchen UX-Komponenten wie stark korrelieren. Ein Aufzeigen und Abgrenzen dieser Zusammenhänge voneinander kann einen Hinweis darauf geben, wie und an welcher Stelle Dauerwahrnehmung bei der Entstehung von UX eine Rolle spielt. Dadurch kann eingegrenzt werden, auf welche Komponenten von UX die Interaktionsdauer und deren Wahrnehmung wirkt. Wie in Abbildung 5.2 dargestellt, dient Studie III als einzige Studie nur der Beantwortung von Forschungsfrage 2. Sie leitet sich aus den Ergebnissen der ersten beiden Studien ab und überträgt diese in den Anwendungskontext. Hier werden Probanden mit “echten Interaktionen” konfrontiert, d.h. sie arbeiten mit real-existierenden Webseiten und lösen typische Aufgaben dieser Webseiten.

In den folgenden Kapiteln erfolgt zum einen die Vorstellung der empirischen Arbeiten und zum anderen die Diskussion der Ergebnisse vor dem Hintergrund der Forschungsfragen. Darüber hinaus enthält der Anhang ein zusätzliches Kapitel zu den statistischen Analysemethoden, die im Rahmen der vorliegenden Studien genutzt werden. Dieses Kapitel B (ab S.221) ist ein Exkurs, welches den statistischen Hintergrund der Themen Linear Mixed-Effects Models (LMMs) und Innersubjektkorrelationen beleuchtet.

Teil II

Empirische Arbeiten

Kapitel 6

Studie I: Effekte von System und Aufgabe auf Dauerwahrnehmung

Das erste Ziel von Studie I besteht darin, Dauerschätzungen und Dauerbewertungen von Nutzern während einer Mensch-Computer-Interaktion näher zu untersuchen (vgl. Abb. 5.2, S. 57). Dazu werden Nutzer mit Aufgaben konfrontiert, die sie mit Hilfe einer Webseite innerhalb weniger Minuten lösen sollen. Um unterschiedliche Interaktionscharakteristika zu erzeugen, werden dabei zwei Faktoren manipuliert: die Usability der Webseite und die mentale Beanspruchung durch eine Zweitaufgabe. Eines der so manipulierten Interaktionscharakteristika ist die Interaktionsdauer. Verschiedene Autoren konnten bereits zeigen, dass eine verminderte Usability zu längeren Interaktionsdauern führt (z.B. Tractinsky et al., 2000; Ziefle & Bay, 2005). Auch die Gabe einer Zweitaufgabe, die hier durch lautes Denken erfolgt, kann Interaktionsdauern verlängern (Hertzum et al., 2009). Mit Hilfe dieser Studie sollen die folgende Teilfrage der Forschungsfrage 1 beantwortet werden:

- a) Können Nutzer die Dauer einer Mensch-Computer-Interaktion adäquat wahrnehmen?
- b) Wie beeinflussen die manipulierten Interaktionscharakteristika die Dauerbewertung?
- c) Welchen Zusammenhang gibt es zwischen der Interaktionsdauer und verschiedenen Operationalisierungen der Dauerwahrnehmung innerhalb von Nutzern?

Eine adäquate Wahrnehmung würde darin bestehen, dass längere Interaktionsdauern auch tatsächlich als länger wahrgenommen werden (vgl. skalare Eigenschaften von prospektiven Dauerschätzungen, z.B. STT nach Gibbon, 1977). Dies würde sich auch darin zeigen, dass sich die Faktoren Usability und Zweitaufgabe auf gleiche Art und Weise auf die tatsächliche und die geschätzte Interaktionsdauer auswirken. Falls der Einfluss unterschiedlich ist, muss es zu Verzerrungen in der Dauerschätzung gekommen sein. In Kapitel 3 wurden Modelle und Befunde zu prospektiven Dauerschätzungen beschrieben, die Ursachen von Verzerrungen beleuchten. Zwei wichtige Einflussfaktoren sind das Arousal-Level als eine Dimension von Emotion und die Verteilung der Aufmerksamkeit. Emotionen können wiederum durch eine Manipulation der Usability beeinflusst werden (Aranyi & van Schaik, 2015; Lavie & Tractinsky, 2004; Thüring & Mahlke, 2007), während die Aufmerksamkeitsverteilung durch eine Zweitaufgabe variiert werden kann (Hertzum & Holmegaard, 2015). Damit erlauben es die Variationen der Faktoren Usability und Zweitaufgabe, Verzerrungen in der Dauerschätzung systematisch anhand von Vorhersagen aus Grundlagenmodellen zu untersuchen.

Neben der Betrachtung von Interaktionsdauern, Dauerschätzungen und Schätzverzerrungen steht in der folgenden Studie auch das Zeitempfinden im Fokus, hier operationalisiert durch eine Dauerbewertung. Bis jetzt gibt es keine Studien, die Dauerbewertungen im Kontext von Mensch-Computer-Interaktionen systematisch untersuchen. Daher stellt sich die Frage, ob sich die Usability und die Zweitaufgabe auch auf die Dauerbewertung auswirken, (Teilfrage b) und ob diese Beeinflussung zu den Effekten auf die Dauerschätzung und die Schätzverzerrungen passt. Darüber hinaus ist unklar, wie die Dauerbewertung sowohl mit der Interaktionsdauer an sich, als auch mit anderen Operationalisierungen von Dauerwahrnehmung zusammenhängt (Teilfrage c).

Das zweite Ziel von Studie I besteht darin, den Zusammenhang zwischen der Dauerwahrnehmung und dem Nutzererleben zu untersuchen (Forschungsfrage 2). Daher wird neben den zeitbezogenen Maßen Dauerschätzung und Dauerbewertung auch verschiedene Komponenten des Nutzererlebens erhoben. Da es sich hierbei primär um Zusammenhänge zwischen Bewertungen handelt, die nach verschiedenen Interaktionen mit dem System erfasst werden, wird hier wie bei Teilfrage c) nach den Zusammenhängen innerhalb von Probanden gefragt:

- d) Welche Zusammenhänge gibt es zwischen den Maßen der Dauerwahrnehmung und den Komponenten des Nutzererlebens innerhalb von Nutzern?

6.1 Methode

Die Vorbereitung und Konzeption von Studie I entstand in Zusammenarbeit mit Nils Backhaus. In diesem Rahmen wurden weitere Daten erhoben, die in der vorliegenden Arbeit nicht vorgestellt werden. Die Auswertung von Teilen dieser Daten kann in der Dissertation von Nils Backhaus (Backhaus, 2017) gefunden werden. Abhängige Variablen, die bereits in Backhaus (2017) genutzt wurden, werden im Folgenden gekennzeichnet. Die Erhebung der Daten erfolgte durch die Verfasserin der vorliegenden Arbeit sowie durch Kevin Nürck und Florian Grieb im Rahmen ihrer Bachelorarbeiten (Grieb, 2015; Nürck, 2015).

6.1.1 Teilnehmer

An Studie I nahmen $N = 65$ Teilnehmer teil. Auf Grund von groben Missverständnissen der Instruktion und einem technischen Ausfall mussten vier Teilnehmer aus der Analyse ausgeschlossen werden. Die so entstandene Stichprobe bestand aus 39 Männern und 22 Frauen ($N = 61$), die im Mittel $M = 24.97$ Jahre ($SD = 4.05$) alt waren. 93.4% der Teilnehmenden waren Studierende der TU Berlin. Die Probanden konnten Versuchspersonenstunden oder 8 € für die Teilnahme erhalten. Die Rekrutierung erfolgte über Lehrveranstaltungen und das Probandenportal des IPAs an der TU Berlin.

6.1.2 Studiendesign

Die Probanden wurden informiert, dass sie an der UX-Testung einer Webseite teilnehmen würden. Die Webseite war so präpariert, dass sie in zwei Versionen vorlag, eine mit niedriger und eine mit hoher Usability. Die Aufgabe der Probanden bestand darin, die Seite zunächst zu explorieren und danach drei Aufgabe mit ihrer Hilfe zu lösen. Darüber hinaus wurden sie gebeten, Aufmerksamkeit auf das Verstreichen der Zeit zu richten, da sie nach jeder Aufgabe eine Schätzung der Aufgabendauer abgeben sollten. Diese Instruktion wurde expli-

zit gegeben, um ein prospektives Schätzparameter zu gewährleisten. Die Hälfte der Probanden wurde darüber hinaus instruiert, während der Interaktion mit der Webseite laut zu denken. Diese Instruktion umfasste die Bitte, möglichst viel und genau zu beschreiben, was sie tun, denken, sehen und fühlen, sowie den Hinweis, ihre Entscheidungen und Ziele während der Interaktion zu verbalisieren. Dies entspricht lautem Denken auf Level 3 der Verbalisierung (Hertzum et al., 2009; Hertzum & Holmegaard, 2013). Für diese Art des lauten Denkens zeigten Hertzum et al. (2009), dass sie im Vergleich zu keinem lauten Denken zu einer Erhöhung der mentalen Beanspruchung führt. Zusammengefasst basierte Studie I auf einem zwei-faktoriellen Design mit den Zwischensubjektfaktoren *Usability* (niedrig vs. hoch) und *lautes Denken* (mit vs. ohne).

6.1.3 Versuchsablauf

Die Probanden wurden einer der vier Gruppen zufällig zu Beginn des Experiments zugeordnet. Alle externen Zeitgeber (z.B. Uhren) wurden aus dem Sichtfeld entfernt und die Probanden unterschrieben eine Einverständniserklärung für die Versuchsteilnahme und die Weiterverarbeitung der Daten¹. Die erste Interaktion der Probanden mit der Webseite bestand in einer freien Exploration, die durch die Probanden jederzeit beendet werden konnte, nach maximal vier Minuten allerdings vom Versuchsleiter abgebrochen wurde. Danach folgten drei Aufgaben, deren Reihenfolge zwischen den Versuchspersonen ausbalanciert wurde (Darstellung der Ausbalancierung siehe Anhang D.1.1). In einer der Aufgaben sollten die Probanden ein Kontaktformular finden und ausfüllen, während sie in den anderen beiden Aufgaben eine spezifische Information auf der Seite suchen sollten. Hierzu standen ihnen maximal sechs Minuten für das Kontaktformular und jeweils maximal vier Minuten für die Informationssuche zur Verfügung. Insgesamt dauerte die Durchführung pro Proband circa 30 Minuten. Abbildung 6.1 fasst den Versuchsablauf graphisch zusammen.

¹Auf Grund einer weiteren Fragestellung, die keine Relevanz für die hier besprochenen Forschungsfragen hat, erfolgte vor Beginn des Experiments eine Baseline-Messung des emotionalen Zustands und die Probanden lasen zwei kurze Texte laut vor. Die Probanden trugen ein Headset während des gesamten Versuchs. Dies wurde mit einer Sprachaufzeichnung während dem lauten Denken bzw. einem späteren Interview begründet, welches nicht stattfand.

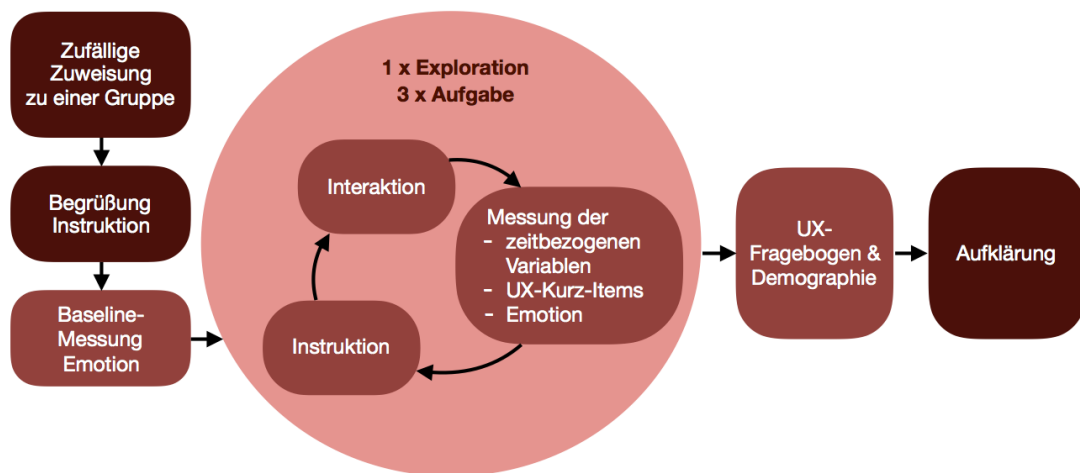


ABBILDUNG 6.1: Übersicht über den Versuchsablauf in Studie I.

6.1.4 Material

Die beiden Versionen der Webseite (siehe Abb. 6.2) entstanden im Rahmen einer Dissertation (Backhaus, 2017) auf Basis einer bestehenden Webseite (cct-ev.de, studentische Unternehmensberatung Berlin). Beide Webseiten unterschieden sich nicht in ihrem Informationsgehalt, aber in der Usability. So war die Webseite mit niedriger Usability schwieriger zu nutzen (zu kleine Schrift, springende Links und häufige, irrelevante Pop-Up-Nachrichten). Der Unterschied in der Usability zeigte sich auch in Bezug auf subjektive Parameter. Backhaus (2017) berichtete, dass sich die Webseiten signifikant in den Dimensionen wahrgenommene Usability, wahrgenommene Nützlichkeit und Gesamtbewertung unterschieden. Darüber hinaus zeigte er, dass die Webseite mit niedriger Usability zu einem größeren Abfall in Valenz und einem größeren Anstieg im Arousal führte.

6.1.5 Operationalisierung der abhängigen Variablen

Die abhängigen Variablen umfassten zeit- und UX-bezogene Variablen. Die zeitbezogenen Variablen wurden nach jeder der vier Aufgaben (Exploration und Aufgabe A bis C) gemessen und bestanden aus der objektiv gemessenen *Aufgabendauer*, einer *Schätzung der Aufgabendauer* in Minuten und Sekunden und der *Dauerbewertung*. Da die Aufgabendauer keine konstante Größe war, sondern durch die Aufgabe, den Probanden und die experimentelle Bedingung beeinflusst werden konnte, muss sie als konfundierende Variable für die



ABBILDUNG 6.2: Webseiten in Studie I. Links hohe Usability, rechts niedrige Usability.

geschätzte Aufgabendauer gesehen werden. Daher wurden *Perceived Time Ratios (PTRs)* als Maß der Schätzverzerrung nach Block et al. (2010) berechnet, um den Schätzfehler in der erlebten Aufgabendauer unabhängig von der objektiven Aufgabendauer betrachten zu können. Hierzu wurde für jede Aufgabe die geschätzte Dauer durch die objektive Dauer geteilt. Ein PTR von eins entsprach damit einer perfekten Schätzung, während ein Wert kleiner eins einer Unter- und ein Wert größer eins einer Überschätzung entsprach. Zur Erfassung der Dauerbewertung wurden drei Items genutzt (*wahrgenommene Länge*, *wahrgenommene Unangemessenheit* und *wahrgenommene Inakzeptabilität der Dauer*). Alle drei Items basierten auf einem semantischen Differential mit 11 Abstufungen, dargestellt auf einer visuellen Analogskala, dessen Pole mit kurz (1) und lang (11)², angemessen (1) und unangemessen (11) sowie akzeptabel (1) und unakzeptabel (11) bezeichnet waren. Ähnliche Items wurden zur Erfassung der Dauerbewertung auch schon von Kurusathianpong und Tangmanee (2018) genutzt. Die Bewertung auf diesen Skalen erfolgte über einen Schieberegler. Die Skalen sind in Anhang D.1.2) dargestellt.

Die UX-bezogenen Variablen unterteilten sich in zwei UX-Kurz-Items³ so-

²Das Item wahrgenommene Länge der Dauer wurde zum besseren Vergleich für die Beschreibung und Analyse gegenüber der Erhebung invertiert.

³Drei weitere Kurz-Items zur Vertrauenswürdigkeit und Gesamtbewertung der Webseite wurden im Rahmen einer zweiten Fragestellung (Backhaus, 2017) ebenfalls gemessen. Diese werden hier nicht diskutiert.

wie ein Emotionsmaß und einen validierten UX-Fragebogen. Die UX-bezogenen Variablen wurden bereits in Backhaus (2017) berichtet. Ähnlich zur Messung der Dauerbewertung wurde für die UX-Kurz-Items ein 11-fach gestuftes semantisches Differential genutzt (Anhang D.1.2). Die Items umfassten die *wahrgenommene Usability* mit den Polen nicht gebrauchstauglich (1) und gebrauchstauglich (11) sowie die *wahrgenommene Attraktivität* mit den Polen unattraktiv (1) und attraktiv (11). Diese Items wurden gewählt, da sie verschiedene Komponenten des Nutzererlebens (wahrgenommene instrumentelle und wahrgenommene nicht-instrumentelle Qualitäten) im Sinne des CUE-Modells repräsentieren (Minge et al., 2016; Mingé & Thüring, 2018; Thüring & Mahlke, 2007). Zur Erfassung der Emotionen wurde das Affect Grid genutzt (Russell, Weiss & Mendelsohn, 1989). Dies ist ein 9x9-Gitter, welches sich über die Dimensionen *Valenz* und *Arousal* aufgespannt. In seiner deutschen Adaption sind die Pole der horizontalen Dimension Valenz mit extrem negativ und extrem positiv sowie die Pole der vertikalen Dimension Arousal mit extrem schläfrig und extrem erregt benannt. Als validierter UX-Fragebogen kam der meCUE zum Einsatz (Minge & Riedel, 2013; Mingé et al., 2016).

Die UX-Kurz-Items und das Emotionsmaß wurden nach jeder Aufgabe zusammen mit den zeitbezogenen Variablen beantwortet, wohingegen der meCUE einmalig nach der letzten Aufgabe ausgefüllt wurde. Diese Aufteilung wurde gewählt, um auf der einen Seite möglichst kurze Items zwischen den einzelnen Aufgaben zu nutzen und auf der anderen Seite die Kurz-Items mit einem validierten Maß abgleichen zu können. Dazu wurden die Bewertungen der Kurz-Items nach der letzten Aufgabe mit den entsprechenden Dimensionen des meCUEs korreliert. Es zeigte sich, dass das Kurz-Item zur wahrgenommenen Usability hoch mit der meCUE-Skala subjektive Usability korrelierte, $r = .87$, $p < .001$. Ähnlich dazu korrelierte die Bewertung der wahrgenommenen Attraktivität signifikant mit der meCUE-Skala visuelle Ästhetik, $r = .64$, $p < .001$. Auf Grund dieser hohen Korrelationen werden die Kurz-Items als valide Maße zur Beantwortung der Forschungsfragen angesehen.

Der Fragebogen und alle Instruktionen wurden digital mit Hilfe von SoSci-Survey (<http://soscisurvey.de>) dargeboten. Eine Ausnahme bildete hierbei die Einverständniserklärung und das Emotionsmaß. Diese wurden den Probanden in Papierform vorgelegt.

6.1.6 Hypothesen

Hypothesen zur Teilfrage a): Einfluss der UVn auf die Dauer, die geschätzte Dauer und die Verzerrung der Schätzung

Da die theoretische Grundlage bzw. die Datenlage im Bereich Zeitwahrnehmung während technischer Interaktionen noch nicht sehr umfangreich ist (Liikkanen & Gómez Gómez, 2013), werden nur Hypothesen zu den Einflüssen der unabhängigen Variablen auf die Aufgabendauer formuliert. Basierend auf Arbeiten, die eine Verlängerung von Aufgabenbearbeitungszeiten zum einen durch lautes Denken (Hertzum et al., 2009) und zum anderen durch niedrige Usability (z.B. Tractinsky et al., 2000; Ziefle & Bay, 2005) zeigen, werden zwei Hypothesen aufgestellt:

Hypothese I.1: Die **Aufgabendauer** ist signifikant länger in der Gruppe mit lautem Denken als in der Gruppe ohne lautes Denken. Dies zeigt sich in einer negativen Steigung: $\beta_{\text{lautesDenken}} < 0$.⁴

Hypothese I.2: Die **Aufgabendauer** ist signifikant kürzer in der Gruppe mit hoher Usability als in der Gruppe mit niedriger Usability. Dies zeigt sich in einer positiven Steigung: $\beta_{\text{Usability}} > 0$.

Teilfrage a) bezieht sich darauf, ob Nutzer die Dauer adäquat wahrnehmen. Eine adäquate Wahrnehmung zeigt sich darin, dass längere Dauern auch länger geschätzt werden. In diesem Fall müssen die geschätzten Dauern durch die gleichen Faktoren beeinflusst werden wie die tatsächliche Aufgabendauer. Es kann jedoch auch zu Verzerrungen in den Dauerschätzungen kommen, die mit Hilfe der PTRs untersucht werden. Solche Verzerrungen können zum Beispiel durch das laute Denken auftreten, wie Arbeiten zum verkürzenden Effekt mentaler Beanspruchung auf Zeitschätzungen (z.B. Hertzum & Holmegaard, 2015) sowie die Vorhersagen des AGMs (Block & Zakay, 1996; Zakay & Block, 1995, 1997) zeigen. Zusätzlich kann auch die Usability eine Verzerrung der Dauerschätzungen hervorrufen, da niedrige Usability zu einer Veränderung im emotionalen Zustand führen kann (Aranyi & van Schaik, 2015; Lavie & Tractinsky,

⁴Die Nullhypothese für Hypothese I.1 lautet: $H_{0I.1}: \beta_{\text{lautesDenken}} \geq 0$. Aus Platzgründen werden die Nullhypothesen der anderen Hypothesen nicht genannt.

2004; Thüring & Mahlke, 2007). In der vorliegenden Studie werden daher die Effekte von lautem Denken und von Usability auf die Dauerschätzung zum einen mit Hilfe der absoluten Werte der Schätzung und zum anderen mit den PTRs-Werten näher betrachtet. Das Ziel besteht darin, die gefundenen Beeinflussungen gegeneinander abzuwägen. Alle Interaktionen werden zusätzlich explorativ analysiert.

Hypothesen zur Teilfrage b): Einfluss der UVn auf die Dauerbewertung

Zur Messung der Dauerbewertung wurden drei Items genutzt (wahrgenommene Länge, wahrgenommene Unangemessenheit und wahrgenommene Inakzeptabilität der Dauer). Da die Items in dieser Form zum ersten Mal zum Einsatz kamen, soll zunächst geprüft werden, ob eine Analyse auf Einzelebene der Items notwendig ist oder eine Zusammenfassung der Items zu einer oder mehreren Komponenten auf Basis der Daten vertretbar ist. Allgemein wird für die Dauerbewertung erwartet, dass sie basierend auf der wahrgenommenen Dauer eingeschätzt wird. Sollte dies der Fall sein, dann müssten Dauern nur dann als eher lang, unangemessen und inakzeptabel beurteilt werden, wenn die Dauer auch als länger wahrgenommen wird. Die Effekte der unabhängigen Variablen auf die Dauerschätzungen würden sich dann in den Effekten auf die Dauerbewertung widerspiegeln. Um dies zu prüfen, soll hier der Einfluss der unabhängigen Variablen auf die Dauerbewertung analysiert werden.

Hypothesen zur Teilfrage c): Innersubjektkorrelation zwischen den zeitbezogenen Maßen

In Teilfrage c) wird nach dem Zusammenhang zwischen Interaktionsdauer und Dauerwahrnehmung gefragt. Hier wird erwartet, dass die objektive Dauer mit der geschätzten Aufgabendauer aber auch der Dauerbewertung⁵ positiv korreliert (Hyp. I.3 (i) und (ii)), d.h. längere Dauern werden länger geschätzt und negativer bewertet. Darüber hinaus wird erwartet, dass die geschätzte Dauer positiv mit der Dauerbewertung korreliert (Hyp. I.4 (i)). Tabelle 6.1 fasst alle

⁵Da die drei Dauerbewertungen nach Möglichkeit auf eine Dimension reduziert werden sollen, wird zu diesen drei Items nur eine Hypothese formuliert.

Korrelationshypothesen zusammen, eine Ausformulierung dieser Hypothesen kann im Anhang C gefunden werden.

Hypothesen zur Teilfrage d): Innersubjektkorrelation zwischen den zeitbezogenen Maßen und dem Nutzererleben

Zusätzlich werden Zusammenhänge innerhalb der Probanden zwischen den zeit- und den UX-bezogenen Variablen erwartet. Hierzu gehören negative Innersubjektkorrelationen zwischen der Dauer bzw. der Dauerwahrnehmung und der wahrgenommenen Usability, der wahrgenommenen Attraktivität und der Valenz (Hyp. I.5 bis I.7 in Tab. 6.1). Je länger eine Dauer ist bzw. je länger sie geschätzt oder je negativer sie bewertet wird, umso negativer sollte das Nutzererleben sein. Diese Erwartung basiert auf empirischen Zusammenhängen zwischen Emotionen und Zeitempfinden (Droit-Volet & Wearden, 2015, 2016) und auf der Relevanz von erlebter Zeit beim Entstehen von UX, die Liikkanen und Gómez Gómez (2013) postulierten. Für das Arousal werden ebenfalls Zusammenhänge zur Dauerwahrnehmung auf Grund von Effekten des Arousals

TABELLE 6.1: Korrelationshypothesen. Erwartet werden Innersubjektkorrelationen der jeweils angegebenen Richtung.

	Aufgaben- dauer	geschätzte Aufgabendauer	Dauer- bewertung
Aufgabendauer			
Hyp. I.3		(i): $r > 0$	(ii): $r > 0$
geschätzte Aufgabendauer			
Hyp. I.4			(i): $r > 0$
wahrg. Usability			
Hyp. I.5	(i): $r < 0$	(ii): $r < 0$	(iii): $r < 0$
wahrg. Attraktivität			
Hyp. I.6	(i): $r < 0$	(ii): $r < 0$	(iii): $r < 0$
Valenz			
Hyp. I.7	(i): $r < 0$	(ii): $r < 0$	(iii): $r < 0$
Arousal			
Hyp. I.8	(i): $r = 0$	(ii): $r = 0$	(iii): $r = 0$

auf die Schätzverzerrungen erwartet. Hier wird keine Richtung angenommen, da sich anders als bei den anderen UX-bezogenen Variablen kein einheitlicher Zusammenhang ableiten lässt.

Zusätzliche Hypothesen: Effekte der UVn auf die UX-bezogenen Variablen

Durch das experimentelle Design werden wahrscheinlich auch die UX-bezogenen Variablen beeinflusst. Daher wird der Effekt der unabhängigen Variablen auch auf die Komponenten des Nutzererlebens betrachtet. Diese Betrachtung kann ggf. die Interpretation der Korrelationen zwischen den zeit- und UX-basierten Maßen erleichtern. Auf Basis von verschiedenen Arbeiten (z.B. Hamborg et al., 2014; Minge & Thüring, 2018; van der Geest & van Dongen, 2009) wird erwartet, dass die Variation der Gebrauchstauglichkeit des Systems einen starken Effekt auf die UX-bezogenen Maße hat. Zu diesen gehört die wahrgenommene Usability als eine instrumentelle Qualität der Interaktion, die wahrgenommene Attraktivität als eine nicht-instrumentelle Qualität und die Emotion mit den Dimensionen Valenz und Arousal. Andere Effekte, wie die des lauten Denkens, werden explorativ betrachtet.

Hyp. I.9 (i) bis (iv): Die (i) **wahrgenommene Usability**, (ii) **wahrgenommene Attraktivität**, (iii) **Valenz** und das (iv) **Arousal** ist signifikant größer in der Gruppe mit hoher Usability als in der mit niedriger Usability. Dies zeigt sich in negativen Steigungen: $\beta_{Usability} 's < 0$.

6.1.7 Datenanalyse

Analyse des Einflusses der UVn auf die AVn

Zur Analyse des Einflusses der zwei unabhängigen Variablen auf die abhängigen Variablen werden LMMs mit einer Maximum Likelihood Schätzung (ML) genutzt (für eine ausführliche Beschreibung siehe Kapitel Abschnitt B.1 im AnhangB). Dazu wird für jede abhängige Variable ein LMM nach dem folgenden Muster aufgestellt:⁶

⁶Die Schreibweise ist an die übliche Schreibweise im lme4-package (Bates, Mächler, Bolker & Walker, 2015) angelehnt. Die Tilde ist zu lesen als „wird vorhergesagt durch“.

$$\begin{aligned}
 AV \sim & 1 + (Usability + lautes\ Denken)^2 + \\
 & Exploration\ vs.\ A + A\ vs.\ B + B\ vs.\ C + \\
 & (1 + Expl.\ vs.\ A + A\ vs.\ B + B\ vs.\ C \parallel Versuchsperson)
 \end{aligned}$$

Wie hier zu sehen ist, werden auf Seiten der Fixed Effects alle möglichen Terme des zwei-faktoriellen Designs (zwei Haupteffekte und eine Zweifach-Interaktionen) integriert. Um den Einfluss der Aufgaben statistisch zu kontrollieren, werden die Aufgaben über einen wiederholten Kontrast (*repeated contrast*, siehe Schad, Hohenstein, Vasisht & Kliegl, 2018 oder Abschnitt B.1.2) in das Modell aufgenommen. Der Kontrast wird einem Random Effect über Aufgabe (1|Aufgabe) vorgezogen, da nur vier Aufgaben inklusive Exploration aus der Grundgesamtheit der Aufgaben gezogen wurden. Diese Anzahl ist für einen Random Effect zu gering (Bolker, 2019), da die Schätzung für die Varianz der Grundgesamtheit nicht zuverlässig ist. Die Effekte der Aufgaben werden nicht weiter interpretiert, da sie der Aufklärung von Fehlervarianz dienen, nicht aber Teil der Hypothesen sind. Alle Effekte auf Seiten der Fixed Effects werden als zentrierte Prädiktoren mit einem Kodierungsabstand von eins zwischen den Stufen in das Modell aufgenommen, um den Intercept des Modells als Gesamtmittelwert (*Grand Mean*) und die β -Gewichte als Unterschiede zwischen den Stufen interpretieren zu können (für eine nähere Beschreibung siehe Abschnitt B.1.2).

Auf Seiten der Random Effects werden ein Random Intercept und drei Random Slopes für die Kontraste der Aufgabe im Modell berücksichtigt. Inhaltlich bedeutet dies, dass das Modell auf Basis der Messwiederholung über die verschiedenen Aufgaben für jede Versuchsperson eine Adjustierung des Intercepts berechnet. Darüber hinaus wird für jede Versuchsperson eine Adjustierung der Aufgabeneffekte berücksichtigt. Die Parameterschätzungen der Korrelationen zwischen Random Effects wird ausgeschlossen, da hierfür zu wenige Datenpunkte vorhanden sind. Entsprechend der Empfehlung von Bates, Kliegl, Vasisht und Baayen (2015) wird die Struktur der Random Effects auf Basis der Daten selektiert, um eine Überparametrisierung des statistischen Modells zu vermeiden (zum Vorgehen siehe Abschnitt B.1.3).

Vor der Betrachtung der LMMs wird für jede abhängige Variable geprüft, ob die Residuen der beschriebenen Modellspezifizierung durch eine Box-Cox-Power-Transformation (Venables & Ripley, 2002) der abhängigen Variable nä-

her an eine Normalverteilung gebracht werden können (vgl. Abschnitt B.1.4). In der folgenden Darstellung der Ergebnisse werden für abhängige Variablen, bei denen eine Transformation der Daten durchgeführt wird, nur die Kennwerte des Modells basierend auf der Transformation berichtet. Die Grafiken werden ebenfalls auf Basis der transformierten Werte erstellt, allerdings sind die Achsen rücktransformiert, so dass eine Interpretation in der Einheit der Originalskalierung möglich ist.

In den präsentierten Analysen wird kein p -Wert berichtet (Begründung siehe Abschnitt B.1.5). Stattdessen werden Effekte mit einem absoluten t -Wert größer zwei ($|t| > 2$) als signifikant angenommen. Zur einfacheren Interpretation wird auf eine Anpassung des Signifikanzniveaus bei gerichteten Hypothesen verzichtet. Zusätzlich wird für jeden Fixed Effekt das 95%-Konfidenzintervalle (Konf.) der Parameterschätzung des β -Gewichts angegeben. Für jedes LMM wird darüber hinaus das marginale $R^2_{GLMM(m)}$, welches die Varianzaufklärung durch die Fixed Effects alleine beschreibt, und das konditionale $R^2_{GLMM(c)}$ berichtet, welches die Aufklärung durch Fixed und Random Effects abbildet (vgl. Abschnitt B.1.5). Abschließend wird für die Effekte der unabhängigen Variablen mit Hilfe des *simr* Paketes (Green & MacLeod, 2016) die erreichte Power für jede unabhängige Variable geschätzt. Für diese Schätzung werden Monte Carlo Simulationen mit jeweils $N=1000$ Simulationen genutzt.

Zusammenhang zwischen den Items zur Dauerbewertung

Um die Dimensionen der Dauerbewertungen zu reduzieren, eignet sich eine Hauptkomponentenanalyse oder *Principle Component Analysis (PCA)*. Dieses Verfahren setzt im Gegensatz zu einer explorativen Faktorenanalyse keine Annahmen über zugrundeliegende Faktoren voraus, da es lediglich eine beschreibende, korrelationsbasierte Reduktionsmethode ist (Bühner, 2011). Durch die Reduktion entstehen „Supervariablen“ oder Komponenten, die die beste Linearkombination der Items darstellen (Bühner, 2011, S. 309).

Die Prüfung der Eignung der Daten für eine Hauptkomponentenanalyse erfolgt nach Bühner (2011) und Field, Miles und Field (2012) mit dem *Bartlett-Test* und dem *Kaiser-Meyer-Olkin-Koeffizienten (KMO)*. Der Bartlett-Test prüft die Alternativhypothese, dass die vorliegende Korrelationsmatrix nicht

der Einheitsmatrix entspricht⁷. Nur wenn er ein signifikantes Ergebnis zeigt, kann eine Hauptkomponentenanalyse durchgeführt werden. Der KMO zeigt an, ob es substantielle Zusammenhänge in der Korrelationsmatrix gibt. Bühner (2011) empfiehlt nur bei einem $KMO > .6$ eine Hauptkomponentenanalyse zu nutzen. Basierend auf verschiedenen Empfehlungen (Bühner, 2011; Courtney & Gordon, 2013; Ledesma & Valero-Mora, 2007) werden zur Extraktion der Anzahl der Komponenten der MAP-Test (*Velicer's Minimum Average Partial*) und die Parallelanalyse (*Horn's Parallel Analysis*) genutzt. Basierend auf der entsprechenden Komponentenanzahl und der Ladungen der Items auf den Komponenten werden Items durch Mittelung zusammengefasst.

Korrelationen

Zur Betrachtung der Zusammenhänge zwischen Variablen werden nach Bland und Altman (1995a) die Korrelationen innerhalb der Probanden (Innersubjektkorrelation) betrachtet (vgl. Abschnitt B.2). Daher beschreiben die berichteten Korrelationskoeffizienten wie zwei Variablen innerhalb einer Person miteinander kovariieren, wenn sie in verschiedenen Situationen gemessen werden. Dieser Ansatz wurde entsprechend der Fragestellung gewählt, da in dieser Arbeit nach den Zusammenhängen innerhalb von Nutzern gefragt wird. In Anlehnung an Cohen (1992) werden Korrelationskoeffizienten von $|r| > .1$ als kleine Effekte, Korrelationskoeffizienten von $|r| > .3$ als mittlere Effekte und Korrelationskoeffizienten von $|r| > .5$ als große Effekte angesehen.

Genutzte Software

Zur Analyse der Daten wurde das Statistik-Programm R (R Core Team, 2018) und die folgenden Pakete genutzt: car (Fox & Weisberg, 2011), corrplot (Wei & Simko, 2017), devtools (Wickham, Hester & Chang, 2018), grid (R Core Team, 2018), lme4 (Bates, Mächler et al., 2015), MASS (Venables & Ripley, 2002), MuMIn (Bartoń, 2018), paramap (O'Connor, 2017), psych (Revelle, 2018), RePsychLing (Baayen, Bates, Kliegl & Vasishth, 2015), reshape (Wickham, 2007), rmcrr (Bakdash & Marusich, 2018), simr (Green & MacLeod, 2016) und tidyverse (Wickham, 2017).

⁷Einheitsmatrix: Alle Korrelationen der Matrix bis auf die Diagonale sind gleich null.

6.2 Ergebnisse

6.2.1 Ergebnisse zur Teilfrage a): Einfluss der UVn auf die Dauer, die geschätzte Dauer und die Verzerrung der Schätzung

Die Ergebnisse zur Aufgabendauer zeigten die erwarteten Effekte des lauten Denkens (Hyp. I.1: $\beta_{\text{lautesDenken}} < 0$) und der Usabilitymanipulation (Hyp. I.2: $\beta_{\text{Usability}} > 0$). Wie anhand der Schätzparameter in Tabelle 6.2 zu sehen ist, waren die Teilnehmer der Gruppe mit lautem Denken im Mittel 37 Sek. langsamer als Teilnehmer in der Gruppe ohne lautes Denken. Teilnehmer der Gruppe mit hoher Usability waren im Mittel 67 Sek. schneller als die der Gruppe mit niedriger Usability. Insgesamt brauchten die Probanden pro Aufgabe im Schnitt 203 Sek.. Eine Visualisierung der Effekte findet sich in Grafik 6.3 a) und b). Es zeigten sich zudem signifikante Unterschiede in der Dauer der verschiedenen Aufgaben, auf die hier nicht weiter eingegangen wird, weil sie nicht hypothesenrelevant sind. Die Interaktion erreichte keine Signifikanz.

Die geschätzte Aufgabendauer wurde auf Grund der Abweichung der Residuen von der Normalverteilung log-transformiert. Die Ergebnisse (siehe Tab. 6.2) zeigten einen Effekt der Usability, der in Abbildung 6.3 c) dargestellt ist. Für einen Effekt des lauten Denkens sowie für die Interaktion wurde keine Evidenz gefunden. Zusammengefasst beeinflusst die Usability auf ähnliche Art und Weise die Dauer und geschätzte Dauer. Das laute Denken hingegen zeigte nur einen Einfluss auf die Dauer, nicht aber auf die Dauerschätzung. Diese Diskrepanz legt nahe, dass es zu Verzerrungen in der Dauerschätzung gekommen ist. Da ein nicht-gefundener Effekt nicht als ein nicht-vorhandener Effekt interpretiert werden kann, wurden die PTRs im Sinne einer Überschätzungsstärke betrachtet. Die Analyse der PTRs wurde ebenfalls auf Grund einer nicht-vorliegenden Normalverteilung mit log-transformierten Werten durchgeführt. Die Ergebnisse der Analyse (Tab. 6.2) zeigten einen Haupteffekt von lautem Denken, nicht aber von Usability. Der Effekt des lauten Denkens ist in Abbildung 6.3 d) dargestellt. Die Interaktion zwischen Usability und lautem Denken hatte weder auf die Dauerschätzung noch auf die PTR-Werte einen signifikanten Einfluss.

KAPITEL 6. STUDIE I: EFFEKTE VON SYSTEM UND AUFGABE AUF DAUERWAHRNEHMUNG

TABELLE 6.2: Effekte der unabhängigen Variablen auf Dauer, transformierte Dauerschätzung und transformierte PRT.

	β -Gewicht	95% Konf.	Std. Fehler	t -Wert	
Effekte auf Aufgabendauer					
Intercept	203.4	[194, 212.89]	4.74	42.90	*
laut. D. mit/ohne	-36.6	[-55.51, -17.74]	9.48	-3.86	*
Usab. hoch/niedrig	66.7	[47.78, 85.55]	9.48	7.03	*
laut. D.:Usab.	19.3	[-18.42, 57.11]	18.97	1.02	
Aufgabe Expl./A	74.9	[53.6, 96.28]	10.83	6.92	*
Aufgabe A/B	-116.1	[-137.42, -94.74]	10.83	-10.72	*
Aufgabe B/C	-30.2	[-51.61, -8.73]	10.88	-2.77	*
Random Effect Struktur: (1 <i>Versuchsperson</i>)					
$R^2_{GLMM(m)}$: 52.3 %, $R^2_{GLMM(c)}$: 57.8 %					
Effekte auf log-transformierte geschätzte Aufgabendauer					
Intercept	5.244	[5.14, 5.34]	0.051	103.518	*
laut. D. mit/ohne	-0.043	[-0.24, 0.16]	0.101	-0.422	
Usab. hoch/niedrig	0.407	[0.21, 0.61]	0.101	4.017	*
laut. D.:Usab.	0.193	[-0.21, 0.6]	0.203	0.951	
Aufgabe Expl./A	0.259	[0.09, 0.43]	0.088	2.942	*
Aufgabe A/B	-0.464	[-0.64, -0.29]	0.088	-5.277	*
Aufgabe B/C	-0.232	[-0.4, -0.06]	0.088	-2.652	*
Random Effect Struktur: (1 <i>Versuchsperson</i>)					
$R^2_{GLMM(m)}$: 24.7 %, $R^2_{GLMM(c)}$: 46.9 %					
Effekte auf log-transformierte PTR					
Intercept	0.071	[-0.01, 0.15]	0.040	1.773	
laut. D. mit/ohne	0.194	[0.04, 0.35]	0.080	2.435	*
Usab. hoch/niedrig	-0.036	[-0.19, 0.12]	0.080	-0.447	
laut. D.:Usab.	-0.049	[-0.37, 0.27]	0.160	-0.310	
Aufgabe Expl./A	-0.044	[-0.13, 0.04]	0.044	-1.015	
Aufgabe A/B	0.140	[0.05, 0.23]	0.044	3.217	*
Aufgabe B/C	0.042	[-0.04, 0.13]	0.043	0.970	
Random Effect Struktur: (1 <i>Versuchsperson</i>)					
$R^2_{GLMM(m)}$: 9.9 %, $R^2_{GLMM(c)}$: 63.4 %					

ANMERKUNG: Die Sterne geben an, ob der Effekt auf einem Signifikanzniveau von $\alpha = .05$ und zweiseitiger Testung signifikant ist.

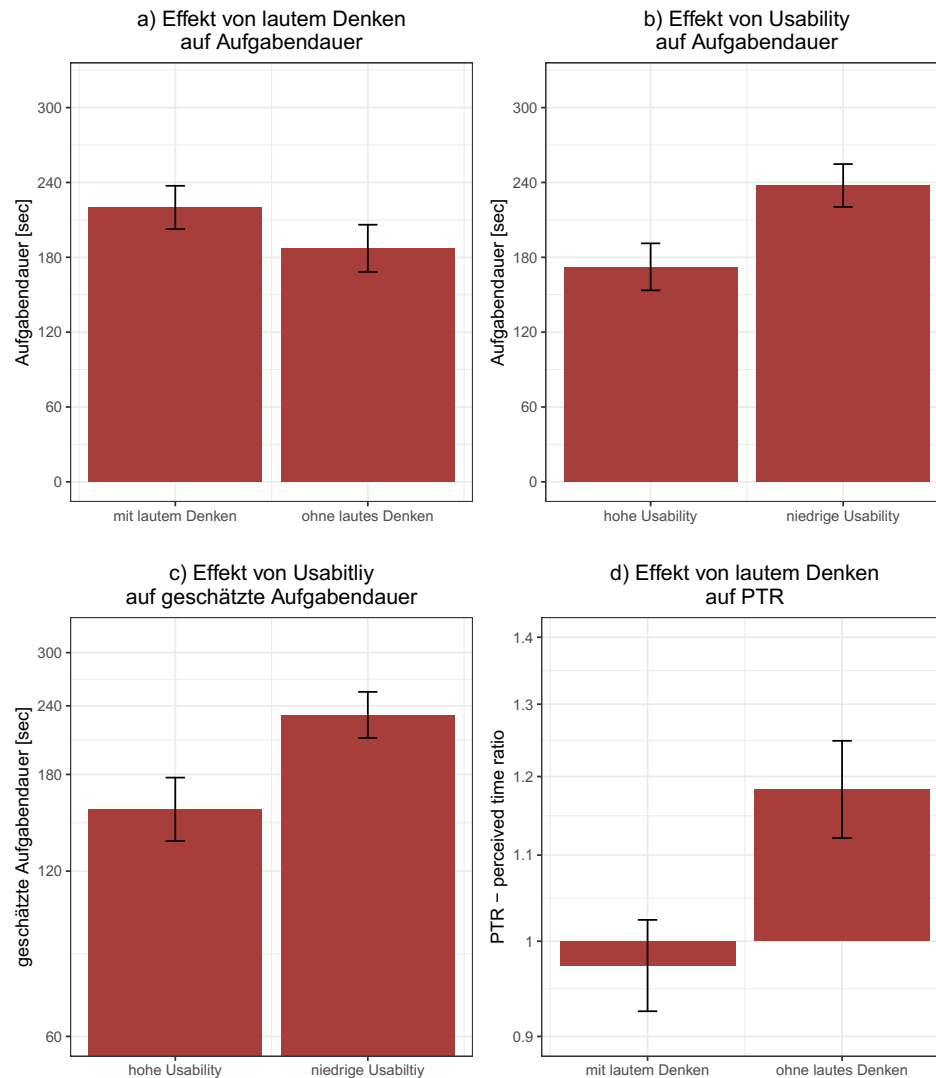


ABBILDUNG 6.3: Signifikante Effekte der unabhängigen Variablen auf Aufgabendauer, log-transformierte geschätzte Aufgabendauer und log-transformierte PTRs. Die Fehlerbalken stellen Konfidenzintervalle der Standardfehler (95 %) korrigiert für Innersubjektfaktoren (Cousineau, 2005; Morey, 2008) dar. Die Beschriftung der y-Achsen für die log-transformierten Daten sind rücktransformiert, um die Interpretation der Werte zu erleichtern. Für die PTRs entspricht ein Wert von 1 einer perfekten Schätzung, Werte über 1 einer Über- und Werte unter 1 einer Unterschätzung.

Die post-hoc Poweranalysen zeigten (Anhang D.2), dass die Effekte der Usability auf die Aufgabendauer und die geschätzte Aufgabendauer sowie der Effekt des lauten Denkens auf die Aufgabendauer auf Basis der hier vorliegenden Ergebnisse als so groß eingeschätzt werden können, dass sie mit der vorliegenden Stichprobengröße auch in einer Replikation sicher zu finden sein sollten ($1 - \beta > 90\%$). Der Effekt des lauten Denkens, wie er hier gefunden wurde, wäre mit der vorliegenden Probandenzahl unterpower ($1 - \beta = 65\%$).

6.2.2 Ergebnisse zur Teilfrage b): Einfluss der UVn auf die Dauerbewertung

Die Hauptkomponentenanalyse über die Items der Dauerbewertung (wahrgenommene Länge, wahrgenommene Unangemessenheit und wahrgenommene Inakzeptabilität der Dauer) wurde zunächst mit den Daten des ersten Messzeitpunktes nach der Exploration durchgeführt. Hier zeigte sich ein signifikanter Bartlett-Test, $\chi^2(3) = 163.96$, $p < .001$, und ein Kaiser-Meyer-Olkin-Koeffizient von $KMO = 0.66$. Auch auf Itemebene waren die KMO-Koeffizienten geeignet, um eine Hauptkomponentenanalyse durchzuführen, $KMOs > 0.60$.

Sowohl der MAP-Test als auch die Parallelanalyse schlugen eine Komponentenstruktur mit nur einer Komponente vor. Die entsprechende Hauptkomponentenanalyse wurde ohne Rotation durchgeführt, da dies bei nur einer Komponente nicht erforderlich ist. Die Ladungen der einzelnen Items auf der Kom-

TABELLE 6.3: Komponentenladungen Studie I beim ersten Messzeitpunkt nach der Exploration.

Item	Komponente Dauerbewertung
wahrgenommene Länge	0.83
wahrgenommene Unangemessenheit	0.95
wahrgenommene Inakzeptabilität	0.96
Eigenwert	2.51
aufgeklärte Varianz	84%

ANMERKUNG: $N = 61$

ponente *Dauerbewertung* und der Eigenwert dieser Komponente sind in Tabelle 6.3 dargestellt. Die Analyse wurde anschließend für den zweiten und dritten Messzeitpunkt sowie für die Aufgaben A bis C durchgeführt. Für die Exploration war der KMO- Koeffizient für die Durchführung einer Hauptkomponentenanalyse zu gering, $KMO = 0.50$. In allen durchgeführten Hauptkomponentenanalysen wurde die gleiche Komponentenstruktur extrahiert (siehe Anhang D.3). Daher wurden die Daten der drei Dauerbewertungsitems pro Messzeitpunkt und Versuchsperson gemittelt. Die neue, gemittelte Variable wird im folgenden unter dem Namen *Dauerbewertung* referenziert.

Für die Variable *Dauerbewertung* zeigte sich ein Haupteffekt der Usability, aber keine weiteren Effekte (siehe Tab. 6.4). Probanden in der Gruppe mit hoher Usability bewerteten die Interaktionsdauern als positiver. Grafik 6.4 zeigt diesen Unterschied. Die post-hoc Poweranalysen zur Teilfrage b) zeigten (Anhang D.2), dass die Auswirkung der Usability auf die *Dauerbewertung* so groß war, dass er mit der vorliegenden Stichprobengröße auch in einer Replikation wieder zu finden sein sollte ($1 - \beta > 99\%$). Die anderen Effekte, die keine Signifikanz erreichten, waren in den vorliegenden Daten so klein, dass sie mit der Stichprobengröße kaum entdeckt werden können ($1 - \beta < 13\%$).

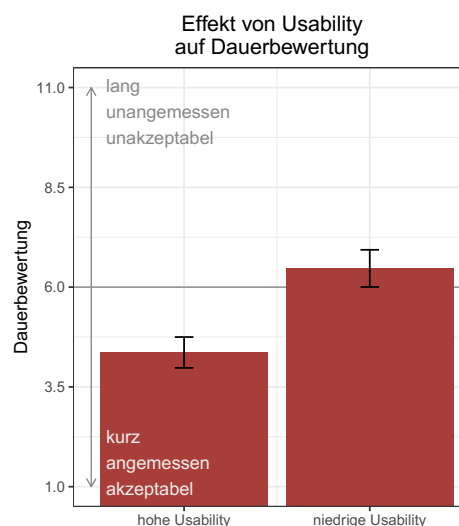


ABBILDUNG 6.4: Effekt der Usability auf die *Dauerbewertung*. Die Fehlerbalken stellen Konfidenzintervalle der Standardfehler (95 %) korrigiert für Innersubjektfaktoren (Cousineau, 2005; Morey, 2008) dar.

TABELLE 6.4: Effekte der unabhängigen Variablen auf die transformierte Dauerbewertung.

	β -Gewicht	95% Konf.	Std. Fehler	t -Wert	
Effekte auf Dauerbewertung					
Intercept	5.430	[5.06, 5.8]	0.188	28.889	*
laut. D. mit/ohne	0.269	[-0.48, 1.02]	0.376	0.715	
Usab. hoch/niedrig	2.080	[1.33, 2.83]	0.376	5.533	*
laut. D.:Usab.	-0.571	[-2.07, 0.93]	0.752	-0.759	
Aufgabe Expl./A	0.186	[-0.54, 0.91]	0.367	0.506	
Aufgabe A/B	0.530	[-0.19, 1.25]	0.367	1.443	
Aufgabe B/C	-0.415	[-1.14, 0.31]	0.367	-1.130	
Random Effect Struktur: (1 <i>Versuchsperson</i>)					
$R^2_{GLMM(m)}$: 18.9 %, $R^2_{GLMM(c)}$: 36.2 %					

ANMERKUNG: Die Sterne geben an, ob der Effekt auf einem Signifikanzniveau von $\alpha = .05$ und zweiseitiger Testung signifikant ist.

6.2.3 Ergebnisse zu den Teilfragen c) und d): Innersubjektkorrelationen

Die Innersubjektkorrelationen zwischen zeitbezogenen und zwischen zeit- und UX-bezogenen Maßen sind in Tabelle 6.5 dargestellt. Die Innersubjektkorrelationen zwischen den UX-bezogenen Variablen können im Anhang D.4 eingesehen werden. Bezüglich der zeitbezogenen Maße zeigte sich, dass Probanden dazu neigten, bei längeren Aufgabendauern auch längere Dauerschätzungen abzugeben. Dies entspricht Hypothese I.3 (i) ($r > 0$). Darüber hinaus vergaben die Probanden in Bezug auf die Dauerbewertung umso größere Werte, je länger die Aufgabendauer war und je länger sie diese schätzten (Hyp. I.3 (ii) und Hyp. I.4 (i): r 's > 0). Ein größerer Wert in der Dauerbewertung entspricht dabei einer negativeren Bewertung als eher lang, unangemessen und unakzeptabel.

Die tatsächliche und die geschätzte Aufgabendauer zeigten die erwarteten negativen Innersubjektkorrelationen nur mit der UX-bezogenen Variable Valenz (Hyp. I.7 (i) und (ii): r 's < 0). Evidenz für einen Zusammenhang zwischen Aufgabendauer und geschätzter Aufgabendauer mit den anderen UX-

TABELLE 6.5: Innersubjektkorrelation zwischen den zeit und UX-bezogenen Variablen.

Hyp. I.		Aufgaben- dauer	geschätzte Aufgaben- dauer	Dauer- bewertung
3 (i)-(ii)	Aufg.dauer	–	.79 ***	.31 **
4 (i)	gesch. Aufg.dauer	–	–	.44 ***
5 (i)-(iii)	wahrg. Usability	-.14, n.s.	-.20, **	-.39 ***
6 (i)-(iii)	wahrg. Attraktivität	-.07, n.s.	-.12, n.s.	-.32 ***
7 (i)-(iii)	Valenz	-.21 *	-.30 *	-.46 ***
8 (i)-(iii)	Arousal	-.04, n.s.	-.02, n.s.	.04, n.s.

ANMERKUNG: * bei $p < .05$, ** bei $p < .01$, *** bei $p < .001$, fett für $|r| \geq .3$

bezogenen Variablen wurde nicht gefunden, für die wahrgenommene Usability (Beibehalten der $H_{0I.5}$ (i) und (ii): $r's \geq 0$), die wahrgenommene Attraktivität (Beibehalten der $H_{0I.6}$ (i) und (ii): $r's \geq 0$) und das Arousal (Beibehalten der $H_{0I.8}$ (i) und (ii): $r's = 0$). Die Dauerbewertung hingegen zeigte negative Innersubjektkorrelationen zu drei Maßen des Nutzerlebens mit mittleren bis starken Effektstärken (Hyp. I.6 (iii), I.6 (iii) und I.7 (iii): $r's < 0$). Hierzu gehörte die wahrgenommene Usability, die wahrgenommene Attraktivität und die Valenz. Je positiver das Nutzererleben bewertet wurde, umso positiver war die Dauerbewertung (kleine Werte). Einzige Ausnahme bildete hier das Arousal (Beibehalten der $H_{0I.8}$ (iii): $r = 0$).

6.2.4 Zusätzliche Ergebnisse: Effekte der UVn auf die UX-bezogenen Variablen

Die Analysen zu den Effekten der unabhängigen Variablen auf die UX-bezogenen Variablen zeigten Haupteffekte der Usability auf die log-transformierte wahrgenommene Usability, die log-transformierte wahrgenommene Attraktivität und die Valenz (Hyp. I.9 (i), (ii) und (iii): $\beta_{Usability} 's < 0$). Sie sind in Abbildung 6.5 a) bis c) dargestellt. Hier ist zu sehen, dass die Webseite mit hoher Usability als gebrauchstauglicher und attraktiver bewertet wurde. Darüber hinaus löste sie eine höhere Valenz aus. Effekte des lauten Denkens

oder der Interaktion gab es nicht. Bezogen auf das Arousal zeigte sich kein Haupteffekt der Usability (Beibehalten der $H_{0I.9}$ (iv): $\beta_{Usability} \geq 0$) oder des lauten Denkens. Die Usability zeigte aber zusammen mit lautem Denken einen Interaktionseffekt, der in Abbildung 6.5 d) dargestellt ist. Unter der Bedingung hohe Usability war das Arousal mit lautem Denken höher als ohne. Dieser Effekt drehte sich in der Bedingung mit niedriger Usability um. Die statistischen Kennwerte dieser Analysen sind im Anhang zu finden (Anhang D.5).

Die post-hoc Poweranalysen zeigten (Anhang D.2), dass der hier geschätzte Effekt der Usability auf die wahrgenommene Usability, die wahrgenommene Attraktivität und die Valenz groß genug war, um ihn mit der vorliegenden Stichprobengröße erneut zu finden ($1 - \beta \geq 79\%$). Für den beobachteten Interaktionseffekt auf das Arousal und den erwarteten Haupteffekt der Usability war die Stichprobengröße jedoch zu gering, um eine akzeptable Power zu erreichen ($40\% \leq 1 - \beta \leq 55\%$).

6.3 Diskussion

In Studie I wurde auf der einen Seite untersucht, in wie weit Nutzer die Dauer einer technischen Interaktion einschätzen können und wie sie diese bewerten, Forschungsfrage 1 mit Teilfragen a) bis c). Auf der anderen Seite wurde untersucht, wie hoch die Zusammenhänge zwischen Dauerwahrnehmung und dem Nutzererleben innerhalb von Nutzern sind, Forschungsfrage 2 mit Teilfrage d). In diesem Rahmen bearbeiteten Nutzer Aufgaben auf einer Webseite, wobei die Usability der Webseite und die mentale Beanspruchung der Aufgabe durch eine Zweitaufgabe (lautes Denken) systematisch variiert wurden.

6.3.1 Zusammenfassung der Ergebnisse und Beantwortung der Forschungsteilfragen

Teilfrage a): Können Nutzer die Dauer einer Mensch-Computer-Interaktion adäquat wahrnehmen?

Die Ergebnisse zeigen, dass Nutzer die Verlängerung der Bearbeitungszeit, die durch schlechte Usability entstand, adäquat wahrnehmen konnten. Dies spie-

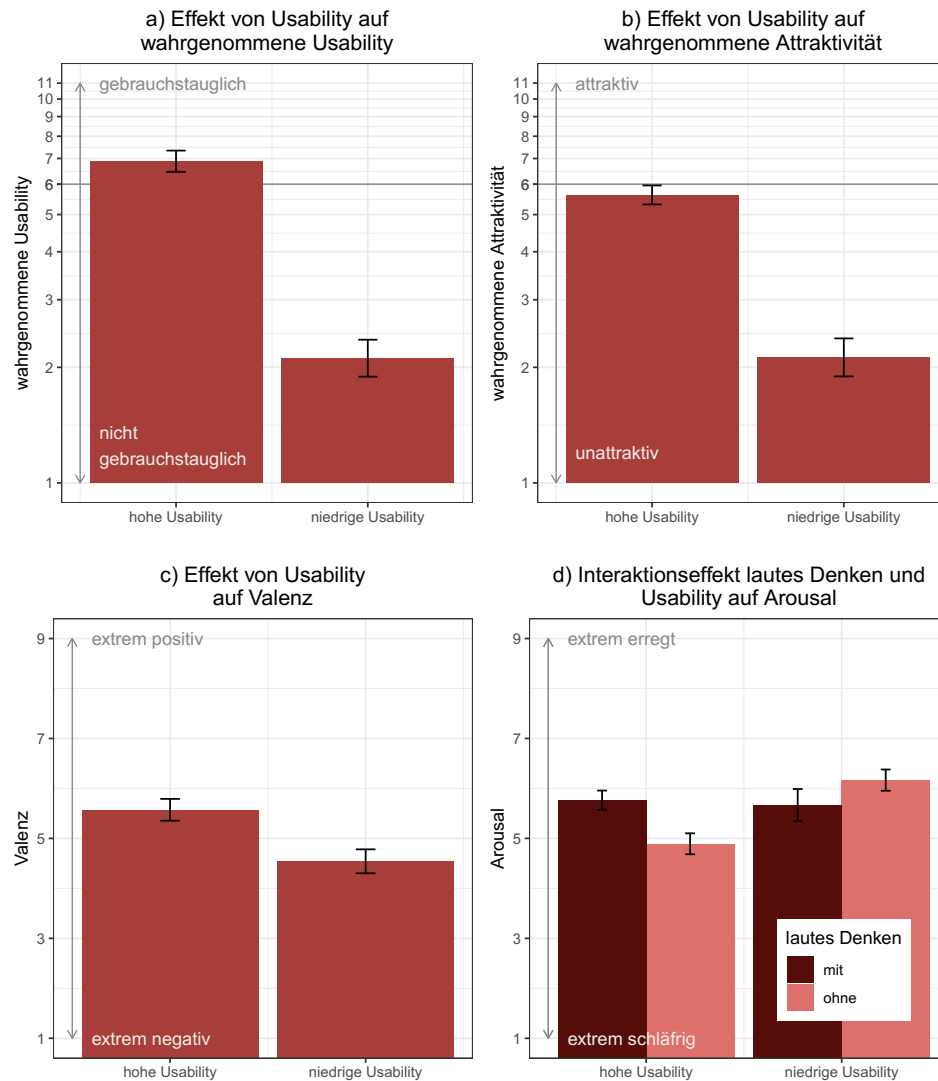


ABBILDUNG 6.5: Effekte der unabhängigen Variablen auf die Komponenten des Nutzererlebens. Die Fehlerbalken stellen Konfidenzintervalle der Standardfehler (95 %) korrigiert für Innersubjektfaktoren (Cousineau, 2005; Morey, 2008) dar. Die Beschriftung der y-Achsen für die log-transformierten Daten sind rücktransformiert, um die Interpretation der Werte zu erleichtern.

gelte sich in größeren Schätzwerten bei längeren Interaktionsdauern wider. Verlängerungen, die durch die Zweitaufgabe lautes Denken entstanden, führten hingegen zu einer systematischen Verzerrung in der Dauerschätzung. Hier wurde die zusätzliche Zeit, die auf Grund des lauten Denkens aufgebracht werden musste, nicht wahrgenommen, wodurch eine kleinere Überschätzung der Interaktionsdauer bei lautem Denken als bei keinem lauten Denken auftrat. Diese Verzerrung wurde auf Grund der Vorhersage des AGM (Block & Zakay, 1996) bezüglich des verkürzenden Effekts von hoher mentaler Beanspruchung bei einer Zweitaufgabe gegenüber keiner Zweitaufgabe erwartet. Sie entspricht dem von Brown (2008) beschriebenen Interferenzeffekt zwischen Schätz- und Zweitaufgabe.

Für einen Effekt des Arousal-Level hingegen gab es keine Evidenz. Zwar unterschied sich der emotionale Zustand der Probanden bezüglich Valenz und in Abhängigkeit von der Zweitaufgabe auch im Arousal-Level zwischen den Usability-Gruppen, dieses Befundmuster spiegelte sich jedoch nicht in den Verzerrungen der Schätzungen wider. Insgesamt legen die Befunde also nahe, dass Nutzer die Interaktionsdauer auch bei unterschiedlicher Usability des Systems adäquat wahrnehmen können. Sobald sie auf Grund einer Zweitaufgabe jedoch weniger Aufmerksamkeit für das Verfolgen der Zeit haben, kommt es zu einer als kürzer eingeschätzten Dauer.

Teilfrage b): Wie beeinflussen die manipulierten Interaktionscharakteristika die Dauerbewertung?

Bei niedriger Usability bewerteten die Probanden die Dauer der Interaktion als negativer als bei hoher Usability. Dies entspricht der tatsächlichen Dauer, die bei niedriger Usability länger als bei hoher Usability war. Da die Usability jedoch ein Zwischensubjektfaktor war, stand den Probanden kein Vergleich zwischen der Dauer mit hoher und niedriger Usability zur Verfügung. Daher kann angenommen werden, dass die Probanden einen internen Vergleichswert hatten, was eine kurze oder eine angemessene Dauer für die beschriebene Aufgabe wäre. Solch ein Vergleich zwischen erlebter und angemessener Dauer könnte der Dauerbewertung zu Grunde liegen. Ein solcher Prozess ist vergleichbar mit dem von Tanaka und Yotsumoto (2017) vorgeschlagenen Prozess zur Entstehung von Zeitflussbeurteilungen. Die Autoren argumentieren auf Basis

mehrerer experimenteller Arbeiten, dass ein Empfinden über den Zeitfluss auf Basis der Diskrepanz zwischen erlebter Zeit und zeitlicher Erwartung entsteht. Für die Entstehung von Dauerbewertungen könnte ein ähnlicher Prozess angenommen werden.

Ein Effekt des lauten Denkens auf die Dauerbewertung wurde nicht gefunden. Das Ausbleiben dieses Effekts trotz eines Unterschieds in der tatsächlichen Dauer könnte verschiedene Ursachen haben und muss mit Vorsicht interpretiert werden, da fehlende Evidenz für einen Effekt nicht als Evidenz für die Nullhypothese gesehen werden darf. Eine mögliche Ursache könnte in dem eben beschriebenen Vergleichsprozess liegen. Zum einen könnten Probanden durch die Zweitaufgabe eine höhere angemessene Dauer veranschlagt haben. Dadurch kann eine längere Dauer mit Zweitaufgabe also genauso kurz, angemessen und akzeptabel wahrgenommen werden wie eine kürzere Dauer ohne Zweitaufgabe. Zum anderen zeigten sich Verzerrungen in der Dauerschätzung zwischen den beiden Gruppen mit und ohne lautes Denken. Es könnte also auch sein, dass die längeren Dauern nicht als länger oder unangemessener bewertet wurden, da die Verlängerung durch das laute Denken gar nicht bemerkt wurde.

Zusammengefasst hatte die Variation des Systems und nicht aber der Aufgabe einen Einfluss auf die Dauerbewertung. Der Einfluss ähnelte den Effekten der unabhängigen Variablen Usability auf die abhängige Variable geschätzte Interaktionsdauer.

Teilfrage c): Welchen Zusammenhang gibt es zwischen der Interaktionsdauer und verschiedenen Maßen der Dauerwahrnehmung innerhalb von Nutzern?

Wie erwartet, zeigte sich ein starker Zusammenhang zwischen der tatsächlichen und der geschätzten Aufgabendauer: Je länger ein Proband für eine Aufgabe brauchte, umso länger schätzte er die Dauer ein. Dies stimmt mit skalaren Eigenschaften von Dauerschätzungen überein, wie sie beispielsweise in der STT (Gibbon, 1977) beschrieben werden. Passend hierzu zeigte sich, dass die Dauerbewertung mit steigender objektiver sowie geschätzter Aufgabendauer negativer wurde. Diese Korrelationen zeigten jedoch eine wesentlich geringere Varianzaufklärung. Daher stellt sich die Frage, welche weiteren Einflüsse auf die Dauerbewertung vorliegen könnten.

Teilfrage d): Welche Zusammenhänge gibt es zwischen den Maßen der Dauerwahrnehmung und den Komponenten des Nutzererlebens innerhalb von Nutzern?

Ein möglicher Ansatzpunkt für die eben genannte Frage liegt in der Beachtung der UX-bezogenen Maße. Die Innersubjektkorrelationen zeigten mittlere bis hohe Korrelationen zwischen der Dauerbewertung und den Skalen wahrgenommene Usability und wahrgenommene Attraktivität: Je negativer die Dauerbewertung war, umso negativer war auch das Nutzererleben. Für diese Skalen zeigte sich jedoch kein signifikanter Zusammenhang zu der tatsächlichen und geschätzten Aufgabendauer. Aus diesen Unterschieden im Befundmuster können zwei Schlüsse gezogen. Erstens, Dauerbewertungen scheinen tatsächlich für Nutzer mit ihrem Nutzererleben verknüpft zu sein, wie es Liikkanen und Gómez Gómez (2013) postuliert haben. Zweitens, Dauerbewertungen sind mehr als nur ein anderes Maß für Dauerschätzungen. Dies wurde beispielsweise bereits von Wearden (2015) argumentiert und kann hier durch weitere Daten gestützt werden.

Die Valenz zeigte den stärksten Zusammenhang zur Dauerbewertung von den UX-bezogenen Maßen. Der Zusammenhang ist nicht überraschend, da auch andere Autoren bereits Evidenz zwischen valenz-ähnlichen Items wie *happiness* und *sadness* zum Zeitempfinden zeigen konnten (Droit-Volet et al., 2018). Diese Autoren nutzten jedoch Zeitflussbeurteilungen als Operationalisierung von Zeitempfinden. Die hier vorliegenden Ergebnisse stützen also die Interpretation, dass das Zeitempfinden mit dem emotionalen Zustand des Nutzers kovariieren und dies für verschiedene Arten der Operationalisierung von Zeitempfinden gilt.

Abschließend kann Teilfrage d) so beantwortet werden, dass die Wahrnehmung einer Interaktionsdauer tatsächlich mit dem Nutzererleben zusammenhängt. Dies gilt vor allem für die Dauerbewertung, da diese Korrelationen zu allen drei Komponenten des Nutzererlebens zeigte. Dies stützt die Annahme, dass die Dauerbewertung als Ergebnis des Vergleichs zwischen erlebter und angemessener Dauer in das Nutzererleben eingebettet wird. Für welche Komponente des Erlebens sie jedoch besonders relevant ist, kann auf Basis dieser Studie nicht beantwortet werden.

6.3.2 Limitationen von Studie I - Ableitung von Studie II

In Studie I arbeiteten Probanden mit lediglich einer Webseite, wodurch unklar ist, wie stark die Befunde verallgemeinert werden können. Darüber hinaus basieren manche gezogenen Schlüsse auf einem Nullbefund, wie beispielsweise die Annahme, dass Probanden längere Interaktionsdauern nicht bemerken, wenn diese durch lautes Denken verlängert werden. In Studie II sollen daher die gefundenen und nicht-gefundenen Effekte mit Hilfe eines anderen Versuchsmaterials repliziert werden, um zum einen durch die Replikation belastbarere Schlüsse ziehen zu können und zum anderen durch das veränderte Versuchsmaterial eine Grundlage für allgemeinere Schlüsse zu schaffen. Darüber hinaus kann das Nicht-Finden mancher Effekte auch an einer zu kleinen Stichprobe gelegen haben, weswegen in Studie II eine Erhöhung der Stichprobengröße erfolgt. Auch der Effekt des lauten Denkens auf die Verzerrung der Schätzung sollte durch einer Vergrößerung der Stichprobengröße zuverlässiger testbar sein.

Eine zweite Limitation von Studie I liegt in der einseitigen Operationalisierung des Zeitempfindens durch die Dauerbewertung. Diese Operationalisierung entspricht der von Liikkanen und Gómez Gómez (2013) vorgeschlagenen Konzeption subjektiv erlebter Zeit. In der Literatur zum Zeitempfinden werden jedoch auch häufig Attribute abgefragt, die sich mehr auf den Zeitfluss während der erlebten Dauer als auf die abschließende Bewertung der Dauer beziehen (vgl. PoTJs, wie z.B. „Die Zeit zog sich“, Sackett et al., 2010; Sucala et al., 2011). Daher soll die Dauerbewertung in Studie II um eine Variable zur Zeitflussbeurteilung erweitert werden. Hier stellen sich ebenfalls die Fragen, in wie weit diese neue Variable in ähnlicher Weise wie die Dauerbewertung von den Interaktionscharakteristika beeinflusst wird und in wie weit sie Varianz mit typischen UX-bezogenen Maßen teilt.

6.3.3 Fazit

Bezogen auf die Frage, in wie weit Nutzer Interaktionsdauern wahrnehmen (Forschungsfrage 1), zeigte Studie I, dass Nutzer während der Interaktion mit einer Webseite dazu in der Lage sind, Dauern adäquat einzuschätzen und sie zu bewerten. Die Schätzung kann jedoch durch eine Zweitaufgabe verkürzt werden

(vgl. AGM Block & Zakay, 1996; Zakay & Block, 1995, 1997). Diese Verzerrung spiegelt sich auch in der Bewertung wider: Systematische Verlängerungen der Dauer durch schlechte Usability werden bemerkt, richtig eingeschätzt und führen zu einer schlechteren Bewertung. Systematische Verlängerungen durch eine zusätzliche Aufgabe werden hingegen nicht bemerkt und bringen keine schlechtere Bewertung mit sich. Bezüglich der zweiten Forschungsfrage zum Zusammenhang zwischen Dauerwahrnehmung und Nutzererleben, kann zusammengefasst werden, dass vor allem die Dauerbewertung und nicht die Dauerschätzung mit den verschiedenen Komponenten des Nutzererlebens korreliert: Je besser die Dauer bewertet wird, desto positiver wird auch das Nutzererleben beurteilt. Insgesamt stützen die Daten also die Annahme, dass die Dauerschätzung zu einer Dauerbewertung führt, die in das Nutzererleben integriert wird.

Kapitel 7

Studie II: Validierung der Effekte von System und Aufgabe auf Dauerwahrnehmung

Studie II stellt eine Fortführung von Studie I mit einem ähnlichen Versuchsdesign dar, in dem Versuchspersonen verschiedene Aufgaben mit Hilfe einer Webseite bearbeiten. Dabei werden zeitbezogene Variablen sowie das Nutzererleben der Versuchspersonen gemessen. In Studie II wird jedoch eine andere Webseite als Versuchsmaterial genutzt, in der sich die Interaktionselemente stärker zwischen den Aufgaben unterscheiden. Darüber hinaus wird die Operationalisierung der Dauerbewertung um Zeitflussbeurteilungen erweitert. Auf Basis dieser Studie werden die gleichen Teilfragen zur Forschungsfrage 1 beantwortet werden, wie in Studie I:

- a) Können Nutzer die Dauer einer Mensch-Computer-Interaktion adäquat wahrnehmen?
- b) Wie beeinflussen die manipulierten Interaktionscharakteristika das Zeitempfinden?
- c) Welchen Zusammenhang gibt es zwischen der Interaktionsdauer und verschiedenen Operationalisierungen der Dauerwahrnehmung innerhalb von Nutzern?

Zu Teilfrage a) wird erwartet, dass Nutzer auf der einen Seite unabhängig von der Usability des Systems die Dauer richtig einschätzen, auf der anderen

Seite aber auf Grund des lauten Denkens eine verzerrte Schätzung abgeben. In dieser Studie soll der Interferenzeffekt mit neuen Aufgaben auf anderen Webseiten gezeigt werden. Bezüglich Teilfrage b) wird auch in Studie II ein starker Einfluss der Usability auf die Dauerbewertung erwartet. Ob es für die Zeitflussbeurteilung als zusätzliche Operationalisierung des Zeitempfindens einen ähnlichen Effekt gibt, kann darüber hinaus geprüft werden. Neben diesen Effekten des faktoriellen Designs werden in Studie II auch wieder die Zusammenhänge zwischen Interaktionsdauer und Dauerwahrnehmung sowie zwischen den Maßen der Dauerwahrnehmung betrachtet (Teilfrage c)). Bezogen auf Forschungsfrage 2 werden zudem die Korrelationen zwischen Dauerwahrnehmung und Nutzererleben geprüft:

- d) Welche Zusammenhänge gibt es zwischen den Maßen der Dauerwahrnehmung und den Komponenten des Nutzererlebens innerhalb von Nutzern?

7.1 Methode

Studie II entstand als Fortführung von Studie I. Studiendesign, Studienvorbereitung, die Analysen und die vorliegenden Beschreibungen stammen von der Verfasserin dieser Arbeit. Die Erhebung der Daten wurde durch die Verfasserin und Lena Mrowetz im Rahmen einer Bachelorarbeit durchgeführt (Mrowetz, 2016).

7.1.1 Teilnehmer

An Studie II nahmen $N = 80$ Teilnehmer teil, wovon die Hälfte männlich und die andere Hälfte weiblich war. Probanden, die bereits an Studie I teilgenommen hatten, wurden von der Rekrutierung ausgeschlossen. Die Probanden waren im Mittel $M = 25.13$ Jahre ($SD = 3.35$) alt. Ähnlich wie in Studie I waren die meisten Teilnehmenden Studierende der TU Berlin (89%). Die Probanden konnten Versuchspersonenstunden oder 8 € für die Teilnahme erhalten. Vor Beginn der Studie wurde das Studiendesign durch die lokale Ethikkommission geprüft. Die Bescheinigung dieser Prüfung ist im Anhang E.2 zu finden.

7.1.2 Studiendesign

Das experimentelle Setting war identisch zu Studie I. Die Webseite, mit der die Probanden interagierten, war aber gegenüber Studie I gegen eine Browser-Oberfläche eines fiktiven Cloud-Services ausgetauscht worden (siehe Abb. 7.1). Auch diese Webseite lag in niedriger und hoher Usability vor. Da die Webseite ein Prototyp war und nicht voll funktionstüchtig, wurde die Exploration der Seite durch eine vierte Aufgabe ersetzt. Studie II basierte wie Studie I auf einem zwei-faktoriellen Design mit den Zwischensubjektfaktoren *Usability* (niedrig vs. hoch) und *lautes Denken* (mit vs. ohne). Die Zuordnung der Probanden zu den vier Gruppen erfolgte randomisiert, so dass in jeder Gruppe 20 Probanden waren. Die Aufteilung der Probanden zu den Gruppen inklusive der Reihenfolge der Aufgaben kann in Anhang E.1.1 eingesehen werden.

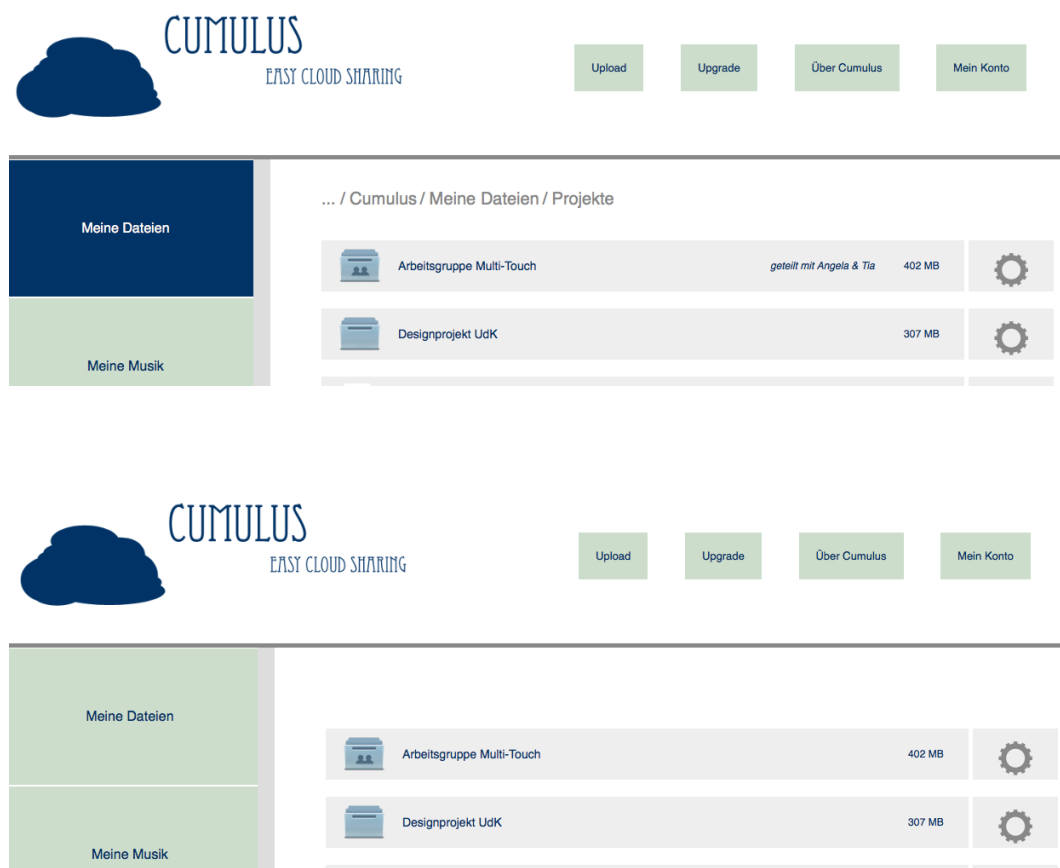


ABBILDUNG 7.1: Webseite aus Studie II für Aufgabe A. Oben hohe Usability, unten niedrige Usability.

7.1.3 Versuchsablauf

Der Versuchsablauf von Studie II entsprach dem von Studie I. Auf Grund der veränderten Webseite wurden die Probanden aber mit anderen Aufgaben konfrontiert. Dazu wurden sie zunächst instruiert, sich vorzustellen, dass sie in einer großen Firma arbeiteten und ein Kollege sie gebeten hatte, ein paar Aufgaben für ihn zu übernehmen. Aus diesem Grund waren sie mit dem Account ihres Kollegen auf der Webseite eines Cloud-Services eingeloggt. Dort sollten sie vier Aufgaben lösen, die in Tabelle 7.1 aufgelistet sind. Für die Version der Webseite mit niedriger Usability wurden für jede Aufgabe unterschiedliche Usability-Probleme integriert, die ebenfalls in Tabelle 7.1 dargestellt sind.

Für jede Aufgabe standen den Probanden jeweils vier Minuten zur Verfügung. Falls ein Proband die Aufgabe nach Ablauf dieser Zeit nicht erledigt hatte, wurde er aufgefordert, den Versuch abubrechen und auf „weiter“ zu klicken. Nach jeder Aufgabe sowie am Ende des Experiments füllten die Probanden einen Fragebogen zur Messung der abhängigen Variablen aus. Die demographischen Variablen wurden zum Ende des Experiments abgefragt. Insgesamt dauerte das Experiment ca. 40 Minuten pro Proband.¹

7.1.4 Material

Die beiden Versionen der Webseite wurde von der Verfasserin in *html* und *css* programmiert. Während der Programmierung fanden iterative Testungen der beiden Versionen mit $N = 12$ Teilnehmern statt, um auf der einen Seite einen Unterschied in der Usability zwischen den beiden Webseiten und auf der anderen Seite die Verständlichkeit der Aufgaben zu garantieren. Hierzu wurden qualitative als auch quantitative Variablen erfasst. Da die Testungen zu verschiedenen Entwicklungsstufen durchgeführt wurden, können keine sinnvollen Mittelwerte der quantitativen Daten der Vortestungen berichtet werden. Vergleicht man aber die Mittelwerte der meCue Skala Gebrauchstauglichkeit aus dem hier präsentierten Datenset, zeigt sich ein signifikanter Unterschied zwi-

¹Auch in diesem Experiment wurden noch weitere Variablen gemessen (z.B. zum Vertrauen), die in dieser Arbeit nicht vorgestellt werden. Zusätzlich wurden die Äußerungen der Probanden mit lautem Denken aufgezeichnet, weswegen alle Probanden während des gesamten Experiments ein Head-Set trugen und zu Beginn des Experiments eine Baseline-Messung der Stimme absolvierten.

schen den Usabilitygruppen, $t(78) = -7.73$, $p < 0.001$, $d = -1.7$. Die Gruppe, die mit der Webseite mit hoher Usability arbeitete, bewertete die Usability also subjektiv besser als die Gruppe, die mit der anderen Webseite arbeitete. Eine genauere Analyse der Effekte des experimentellen Designs auf die UX-bezogenen Maße wird in den Ergebnissen präsentiert.

TABELLE 7.1: Aufgaben und Usability-Probleme in Studie II.

Aufgabe		Usability-Problem bei Webseite mit niedriger Usability
A	Datei finden, in dieser Datei eine Information herausfiltern und auf der Grundlage dieser Information eine andere Datei umbenennen	intransparente Ordnerstruktur, wichtige Informationen sind nicht direkt ersichtlich
B	herausfinden, wer verantwortlich für den Support bei dem Cloud-Service ist	automatisches Weiterleiten zu einer uninformativen Seite, was die Suche erschwert
C	die persönlichen Informationen des Kollegen verändern (z.B. Telefonnummer)	kleine Schriftgröße, uneindeutige Benennung, teilweise vertauschte Position von Links (Abbrechen und Sichern), unpassende Farbgebung (grüner Button für Cancel), unverständliche Fehlermeldung
D	ein Bilderalbum auf die richtige Größe komprimieren, damit das Album geteilt werden kann	ungewöhnliche Position von wichtigen Informationen, unpraktikable Sortierung von Listen, großer Scroll-Aufwand

7.1.5 Operationalisierung der abhängigen Variablen

Die abhängigen Variablen umfassten die gleichen zeit- und UX-bezogene Variablen wie Studie I, die auf die gleiche Art und Weise gemessen wurden. Die Liste der Variablen ist in Tabelle 7.2 zur besseren Übersichtlichkeit dargestellt. Die abhängigen Variablen wurden um die *wahrgenommene Schnelligkeit* zur Erfassung der Zeitflussbeurteilung ergänzt, welches wie die Dauerbewertungen auf einem semantischen Differential mit 11 Abstufungen von schnell (1) bis langsam (11) basierte². Dieses Maß entspricht einer Zeitflussbeurteilung, wie sie von verschiedenen Autoren bereits genutzt wurde (z.B. Droit-Volet & Wearden, 2016; Droit-Volet et al., 2017; Gorn et al., 2004; Tanaka & Yotsumoto, 2017).

Auch in Studie II wurden die genutzten Kurz-Items zur Messung des Nutzererlebens mit Skalen eines validierten Maßes (meCUE, Minge et al., 2016) verglichen, um zu kontrollieren, ob die Kurz-Items valide Maße sind. Dazu wurde der meCUE nach der Bewertung der vierten Aufgabe eingesetzt. Es zeigte sich erneut eine hohe Korrelation zwischen dem Kurz-Item zur wahrgenommenen Usability am vierten Messzeitpunkt mit der meCUE-Skala subjektive Usability, $r = 0.84$, $p < 0.001$, und eine mittlere Korrelation zwischen dem Kurz-Item zur wahrgenommenen Attraktivität am vierten Messzeitpunkt zur meCUE-Skala visuelle Ästhetik, $r = 0.47$, $p < 0.001$. Auch wenn die wahrgenommene Attraktivität in diesem Datenset nur einen mittleren Zusammenhang zur visuellen Ästhetik zeigte, werden die Kurz-Items als ausreichend valide angesehen,

²Das Item wahrgenommene Schnelligkeit wurde zum besseren Vergleich für die Beschreibung und Analyse gegenüber der Erhebung invertiert.

TABELLE 7.2: Übersicht der abhängigen Variablen in Studie II.

zeitbezogenen Variablen	Aufgabendauer, geschätzte Aufgabendauer, PTRs, Dauerbewertung (wahrgenommene Länge, wahrgenommene Unangemessenheit und wahrgenommene Inakzeptabilität), Zeitflussbeurteilung (wahrgenommene Schnelligkeit der Dauer)
UX-bezogene Variablen	wahrgenommene Usability, wahrgenommene Attraktivität, Valenz und Arousal

da keine sehr hohe Korrelation zwischen zwei verwandten aber nicht identischen Konstrukten zu erwarten war.

Alle Fragebögen und Instruktionen wurden mit Hilfe von soscisurvey.de dargeboten, während die Einverständniserklärung und das Emotionsmaß den Probanden in Papierform vorgelegt wurden.

7.1.6 Hypothesen

Die Hypothesen werden zusammen mit der entsprechenden Nummerierung aus Studie I übernommen, um einen direkten Vergleich zu erleichtern. Ähnlich wie in Studie I werden neben den Hypothesen auch andere Haupt- und Interaktionseffekte in den Daten explorativ untersucht, um eine möglichst hohe Vergleichbarkeit zwischen Studie I und Studie II zu gewährleisten.

Hypothesen zur Teilfrage a): Einfluss der UVn auf die Dauer, die geschätzte Dauer und die Verzerrung der Schätzung

Analog zu Studie I werden die Hypothesen II.1 und II.2 zu den Effekten von lautem Denken und Usability auf die Aufgabendauer beibehalten (Hyp. II.1: $\beta_{\text{lautesDenken}} < 0$; Hyp. II.2: $\beta_{\text{Usability}} > 0$). Da in Studie I ein Effekt der Usability auf die geschätzte Aufgabendauer gefunden wurde, wird Hypothese II.10 zum Effekt der Usability neu formuliert. Ebenfalls neu formuliert wird Hypothese II.11 zum Effekt des lauten Denkens auf die Schätzverzerrungen (PTR), der in Studie I gefunden wurde.

Hypothese II.10: Die *geschätzte Aufgabendauer* ist signifikant kürzer in der Gruppe mit hoher Usability als in der Gruppe mit niedriger Usability. Dies zeigt sich in einer positiven Steigung: $\beta_{\text{Usability}} > 0$.

Hypothese II.11: Die *PTRs* sind signifikant kleiner in der Gruppe mit lautem Denken als in der Gruppe ohne lautes Denken. Dies zeigt sich in einer positiven Steigung: $\beta_{\text{lautesDenken}} > 0$.

Hypothesen zur Teilfrage b): Einfluss der UVn auf die Dauerbewertung und die Zeitflussbeurteilung

Ob sich die drei Items der Dauerbewertung auch hier wieder in einer Komponente zusammenfassen lassen, wird erneut geprüft. Auf Basis der ersten Studie wird für diese zusammengefasste Dauerbewertung ein Haupteffekt der Usability erwartet, was zu der neuen Hypothese II.12 (i) führt. Zusätzlich wird ein ähnlicher Effekt auf die Zeitflussbeurteilung erwartet.

Hypothese II.12 (i) bis (ii): *In der Gruppe mit hoher Usability zeigt sich ein signifikant geringerer Wert als in der Gruppe mit niedriger Usability in Bezug auf (i) **die Dauerbewertung**, (ii) **die Zeitflussbeurteilung**. Dies zeigt sich in positiven Steigungen: $\beta_{Usability} 's > 0$.*

Hypothesen zur Teilfrage c): Innersubjektkorrelation zwischen den zeitbezogenen Maßen

Die zeitbezogenen Variablen werden erneut auf ihren Zusammenhang untereinander untersucht. Hier werden positive Korrelationen zwischen Aufgabendauer und Dauerwahrnehmung erwartet (siehe Hyp. II.3 (i) bis (iii) in Tab 7.3). Zusätzlich sollten die Maße der Dauerwahrnehmung positiv miteinander korrelieren (siehe Hyp. II.4 (i) und (ii) sowie II.13 in Tab 7.3).

Hypothesen zur Teilfrage d): Innersubjektkorrelationen zwischen den zeitbezogenen Maßen und dem Nutzererleben

In Anlehnung an Studie I werden für die Zusammenhänge zwischen zeit- und UX-bezogenen Variablen negative bzw. ungerichtete Korrelationen erwartet. Hier werden die Hypothesen II.5 bis II.8 übernommen und für die Zeitflussbeurteilung ergänzt (siehe Hyp. II.5 bis II.8 jeweils (i) bis (iv) in Tab 7.3).

Zusätzliche Hypothesen: Effekte der UVn auf die UX-bezogenen Variablen

Auch in Studie II werden die Effekte der unabhängigen Variablen auf die UX-bezogenen Maße betrachtet, um die Korrelationen zwischen den zeit- und UX-bezogenen Maßen besser einordnen zu können. Dazu werden die Hypothesen

TABELLE 7.3: Korrelationshypothesen. Erwartet werden Innersubjektkorrelationen der jeweils angegebenen Richtung. Gegenüber der entsprechenden Tabelle zu Studie I wurde die Spalte für die Zeitflussbeurteilung ergänzt.

	Aufgaben- dauer	geschätzte Aufg.dauer	Dauer- bewertung	Zeitfluss- beurteilung
Aufgabendauer				
Hyp. II.3		(i): $r > 0$	(ii): $r > 0$	(iii): $r > 0$
geschätzte Aufgabendauer				
Hyp. II.4			(i): $r > 0$	(ii): $r > 0$
Dauerbewertung				
Hyp. II.13				$r > 0$
wahrg. Usability				
Hyp. II.5	(i): $r < 0$	(ii): $r < 0$	(iii): $r < 0$	(iv): $r < 0$
wahrg. Attraktivität				
Hyp. II.6	(i): $r < 0$	(ii): $r < 0$	(iii): $r < 0$	(iv): $r < 0$
Valenz				
Hyp. II.7	(i): $r < 0$	(ii): $r < 0$	(iii): $r < 0$	(iv): $r < 0$
Arousal				
Hyp. II.8	(i): $r = 0$	(ii): $r = 0$	(iii): $r = 0$	(iv): $r = 0$

II.9 (i) bis (iv) ($\beta_{Usability} 's < 0$) zum Effekt der Usability auf die UX-bezogenen Maße (i) wahrgenommene Usability, (ii) wahrgenommene Attraktivität, (iii) Valenz und (iv) Arousal übernommen.

7.1.7 Datenanalyse

Die Datenanalyse erfolgt auf die gleiche Art und Weise wie in Studie I. Es werden die gleichen statistischen Programme und Pakete genutzt. Auf Grund der neuen Aufgabe lautet das gefittete Modell für die Analyse wie folgt:

$$\begin{aligned}
 AV \sim & 1 + (Usability + lautes\ Denken)^2 + \\
 & A\ vs.\ B + B\ vs.\ C + C\ vs.\ D + \\
 & (1 + A\ vs.\ B + B\ vs.\ C + C\ vs.\ D \parallel Versuchsperson)
 \end{aligned}$$

7.2 Ergebnisse

7.2.1 Ergebnisse zur Teilfrage a): Einfluss der UVn auf die Dauer, die geschätzte Dauer und die Verzerrung der Schätzung

Die Analyse zur Aufgabendauer, geschätzten Aufgabendauer und zur Verzerrung der Schätzung in Form der PTRs mussten auf Grund der Verletzung der Normalverteilungsannahme mit log-transformierten Werten durchgeführt werden. Die Ergebnisse sind in Tabelle 7.4 abgebildet.

Im Mittel brauchten die Probanden $M = 143.13$ Sek. ($SD = 63.46$ Sek.) pro Aufgabe. Wie erwartet, zeigten sich signifikante Effekte des lauten Denkens und der Usability auf die Aufgabendauer. Lautes Denken führte im Vergleich zu keinem lauten Denken zu einer durchschnittlichen Verlängerung der Dauer um $M = 35.98$ Sek. (Hyp. II.1: $\beta_{lautesDenken} < 0$) und hohe Usability führte im Vergleich zu niedriger Usability zu einer um $M = 55.45$ Sek. kürzeren Aufgabendauer (Hyp. II.2: $\beta_{Usability} > 0$). Beide Effekte waren zudem Teil einer Interaktion die in Abbildung 7.2 a) abgebildet ist. Hier wird deutlich, dass der verlängernde Effekt des lauten Denkens vor allem bei hoher Usability zum Tragen kam. Darüber hinaus zeigten sich signifikante Unterschiede in der mittleren Aufgabendauer zwischen den Aufgaben.

Mit einer Schätzung von $M = 176.71$ Sek. ($SD = 97.32$ Sek.) überschätzten die Probanden die Dauer im Mittel. In der Analyse zeigte sich der erwartete Effekt der Usability auf die geschätzte Dauer (Hyp. II.10: $\beta_{Usability} > 0$). Probanden, die mit der hohen Usability arbeiteten, schätzten die Aufgabendauer im Mittel als $M = 60.47$ Sek. kürzer ein als diejenigen, die mit niedriger Usability konfrontiert waren. Usability war aber auch Teil einer zweifach Interaktion mit lautem Denken, die in Abbildung 7.2 b) dargestellt ist. In der Abbildung wird deutlich, dass der Unterschied zwischen den Usability-Stufen wesentlich stärker ausgeprägt war für Probanden, die nicht laut dachten. Bezüglich der Überschätzungsstärke (PTRs) zeigte sich der erwartete Haupteffekt des lauten Denkens (Hyp. II.11: $\beta_{lautesDenken} > 0$). Während alle Probanden im Mittel zu einer Überschätzung der Dauern neigten, taten sie dies in der Gruppe ohne lautes Denken stärker, siehe Abb. 7.2 c).

TABELLE 7.4: Effekte der UVn auf log-transformierte Dauer, log-transformierte Dauerschätzung und log-transformierte PRT.

	β -Gewicht	95% Konf.	Std. Fehler	t -Wert	
Effekte auf log-transformierte Aufgabendauer					
Intercept	4.86	[4.81, 4.91]	0.03	178.16	*
laut. D. mit/ohne	-0.30	[-0.41, -0.19]	0.05	-5.44	*
Usab. hoch/niedrig	0.42	[0.31, 0.53]	0.05	7.67	*
laut. D.:Usab.	0.41	[0.2, 0.63]	0.11	3.79	*
Aufgabe A/B	-0.29	[-0.38, -0.2]	0.04	-6.50	*
Aufgabe B/C	0.09	[0.01, 0.18]	0.04	2.11	*
Aufgabe C/D	0.25	[0.17, 0.34]	0.04	5.65	*
Random Effect Struktur: (1 <i>Versuchsperson</i>)					
$R^2_{GLMM(m)}$: 44.7 %, $R^2_{GLMM(c)}$: 62.7 %					
Effekte auf log-transformierte geschätzte Aufgabendauer					
Intercept	5.024	[4.94, 5.11]	0.044	115.393	*
laut. D. mit/ohne	-0.086	[-0.26, 0.09]	0.087	-0.987	
Usab. hoch/niedrig	0.373	[0.2, 0.55]	0.087	4.278	*
laut. D.:Usab.	0.488	[0.14, 0.83]	0.174	2.801	*
Aufgabe A/B	-0.220	[-0.34, -0.1]	0.061	-3.590	*
Aufgabe B/C	0.008	[-0.11, 0.13]	0.061	0.125	
Aufgabe C/D	0.192	[0.07, 0.31]	0.061	3.160	*
Random Effect Struktur: (1 <i>Versuchsperson</i>)					
$R^2_{GLMM(m)}$: 19 %, $R^2_{GLMM(c)}$: 54.6 %					
Effekte auf log-transformierte PTR					
Intercept	0.169	[0.1, 0.24]	0.036	4.642	*
laut. D. mit/ohne	0.228	[0.08, 0.37]	0.073	3.122	*
Usab. hoch/niedrig	-0.047	[-0.19, 0.1]	0.073	-0.643	
laut. D.:Usab.	0.086	[-0.2, 0.38]	0.146	0.589	
Aufgabe A/B	0.088	[-0.01, 0.18]	0.048	1.838	
Aufgabe B/C	-0.096	[-0.19, 0]	0.048	-2.005	*
Aufgabe C/D	-0.069	[-0.16, 0.02]	0.047	-1.461	
Random Effect Struktur: (1 <i>Versuchsperson</i>)					
$R^2_{GLMM(m)}$: 9.1 %, $R^2_{GLMM(c)}$: 53.5 %					

ANMERKUNG: Die Sterne geben an, ob der Effekt auf einem Signifikanzniveau von $\alpha = .05$ und zweiseitiger Testung signifikant ist.

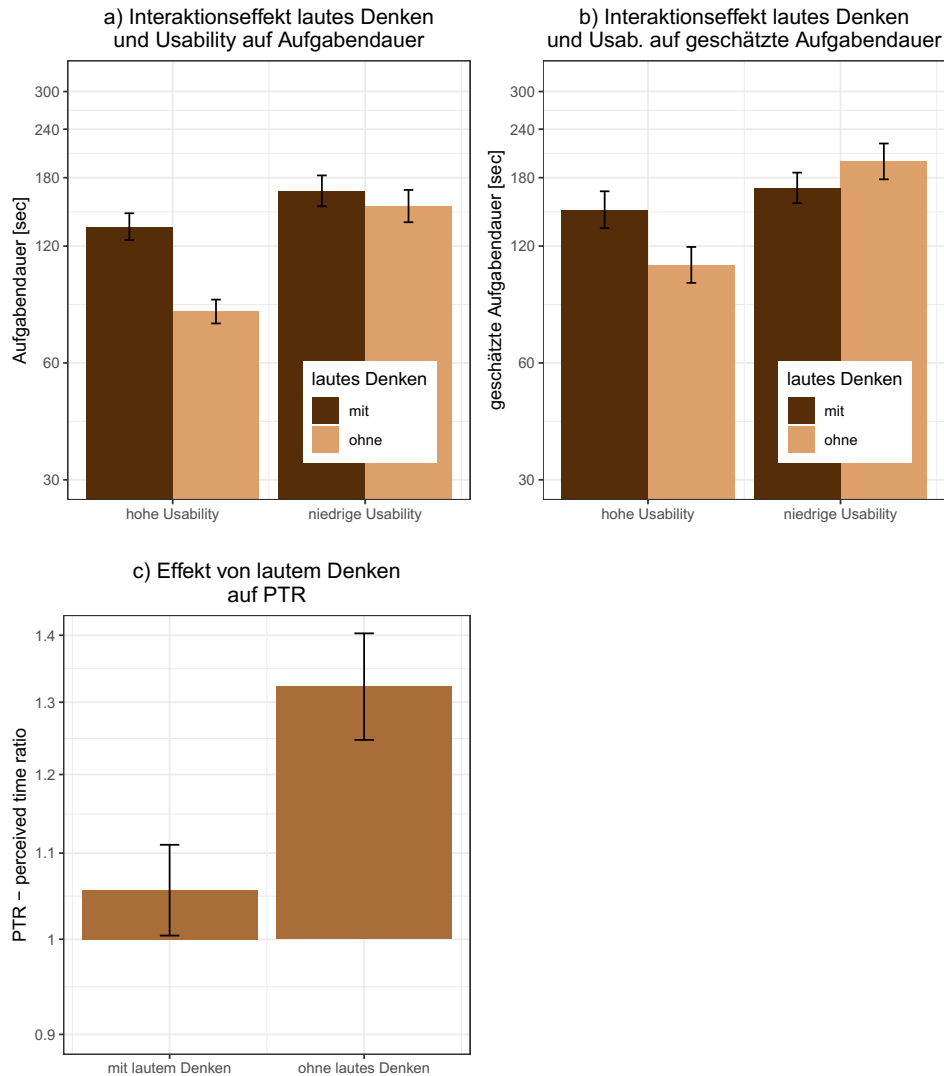


ABBILDUNG 7.2: Effekte der unabhängigen Variablen auf log-transformierte Aufgabendauer, log-transformierte geschätzte Aufgabendauer und log-transformierte PTRs. Die Fehlerbalken stellen Konfidenzintervalle der Standardfehler (95 %) korrigiert für Innersubjektfaktoren (Cousineau, 2005; Morey, 2008) dar. Die Beschriftung der y-Achsen für die log-transformierten Daten sind rücktransformiert, um die Interpretation der Werte zu erleichtern. Für die PTRs entspricht ein Wert von 1 einer perfekten Schätzung.

Die post-hoc Poweranalysen zur Teilfrage a) zeigten (Anhang E.3), dass die erwarteten Effekte der Usability auf die Aufgabendauer und die geschätzte Aufgabendauer sowie die erwarteten Effekte des lauten Denkens auf die Aufgabendauer und die Überschätzungsstärke (PTRs) auf Basis der hier vorliegenden Effektstärken als so groß eingeschätzt werden können, dass sie mit der vorliegenden Stichprobengröße auch in einer Replikation zu finden sein sollten ($1 - \beta \geq 90\%$). Die hier nicht signifikanten Effekte hatten hingegen in den vorliegenden Daten zu kleine Effekte, um detektiert werden zu können ($1 - \beta \leq 16\%$).

7.2.2 Ergebnisse zur Teilfrage b): Einfluss der UVn auf die Dauerbewertung und die Zeitflussbeurteilung

Analog zu Studie I wurde mit einer Hauptkomponentenanalyse geprüft, ob sich die Dauerbewertungsitems wahrgenommene Länge, Unangemessenheit und Inakzeptabilität zu einer Komponente zusammenfassen lassen. Für die Daten des ersten Messzeitpunktes zeigte sich eine ausreichende Eignung der Daten für das Verfahren, Bartlett-Test mit $\chi^2(3) = 161.61$, $p < .001$, Kaiser-Meyer-Olkin-Koeffizient von $KMO = 0.67$ (auf Itemebene KMO-Koeffizienten > 0.60). Auch in dieser Studie schlugen sowohl der MAP-Test als auch die Parallelanalyse eine Komponentenstruktur mit nur einer Komponente vor. Die Analyse wurde zur Validierung dieser Struktur für die anderen drei Messzeitpunkte als auch für jede Aufgabe wiederholt, wobei alle Analysen ebenfalls auf nur eine Komponente hinwiesen (alle Kennwerte siehe Anhang E.4). Auf Grund der Ergebnisse der Hauptkomponentenanalysen wurden für jeden Messzeitpunkt und jede Versuchsperson die Werte der drei Dauerbewertungsitems gemittelt und als neue Variable Dauerbewertung zusammengefasst.

Die Analysen der Dauerbewertung und der Zeitflussbeurteilung sind in Tabelle 7.5 dargestellt. Wie erwartet zeigte sich ein starker Effekt der Usability auf die Dauerbewertung und auf die Zeitflussbeurteilung (Hyp. II.12 (i) und (ii): $\beta_{Usability} 's > 0$). Die Dauern wurden in der Gruppe mit niedriger Usability als länger, unangemessener und unakzeptabler sowie als langsamer vergehend bewertet. Darüber hinaus wurde eine Interaktion zwischen Usability und lautem Denken für die Dauerbewertung gefunden. Die Effekte sind in Ab-

bildung 7.3 visualisiert. Sie zeigt, dass der Effekt der Usability in der Gruppe ohne lautes Denken stärker war als in der Gruppe mit lautem Denken. Dieses Befundmuster entspricht der gefundenen Interaktion von lautem Denken und Usability auf die geschätzte Zeit. Die post-hoc Poweranalysen zur Teilfrage b) zeigten (Anhang E.3), dass der erwartete Effekte der Usability auf die Dauerbewertung und die Zeitflussbeurteilung auf Basis der vorliegenden Daten als so groß eingestuft werden kann, dass er mit gleicher Stichprobengröße auch in einer folgenden Studie erneut gefunden werden sollte ($1 - \beta \geq 92\%$).

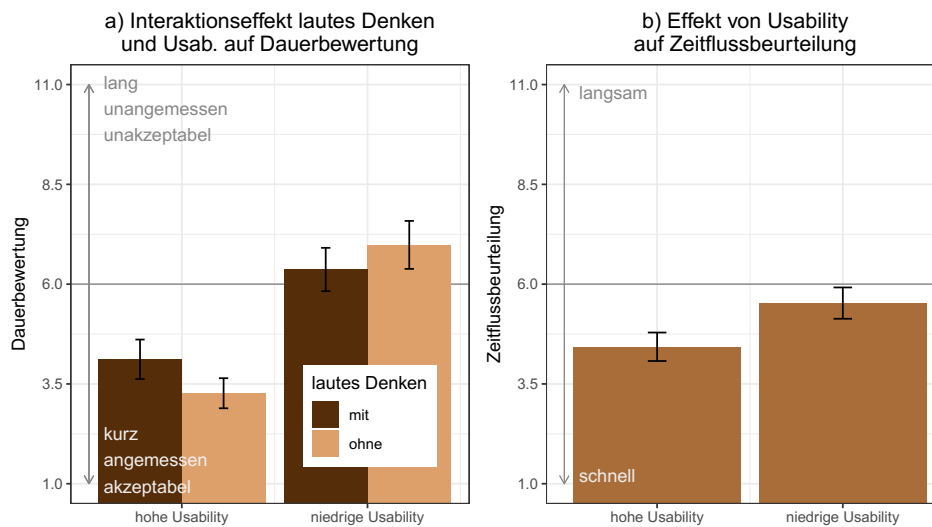


ABBILDUNG 7.3: Effekte der unabhängigen Variablen auf Dauerbewertungen und Zeitflussbeurteilung. Die Fehlerbalken stellen Konfidenzintervalle der Standardfehler (95 %) korrigiert für Innersubjektfaktoren (Cousineau, 2005; Morey, 2008) dar.

7.2.3 Ergebnisse zu den Teilfragen c) und d): Innersubjektkorrelationen

Zur Betrachtung der Korrelationen zwischen den zeit- und zwischen den zeit- und UX-bezogenen Maßen wurden erneut die Innersubjektkorrelationen berechnet. Diese Zusammenhänge sind in Tabelle 7.6 dargestellt. Die Innersubjektkorrelationen zwischen den UX-bezogenen Variablen waren nicht Teil einer Hypothese waren, sind aber zur Vollständigkeit im Anhang E.5 zu finden.

Ähnlich zu Studie I zeigten die Innersubjektkorrelationen zwischen den zeit-bezogenen Maßen, dass längere Dauern auch länger geschätzt und schlechter

TABELLE 7.5: Effekte der unabhängigen Variablen auf Dauerbewertungen und Zeitflussbeurteilung.

	β -Gewicht	95% Konf.	Std. Fehler	t -Wert	
Effekte auf Dauerbewertung					
Intercept	5.183	[4.88, 5.48]	0.151	34.387	*
laut. D. mit/ohne	-0.117	[-0.71, 0.48]	0.301	-0.387	
Usab. hoch/niedrig	2.983	[2.39, 3.58]	0.301	9.896	*
laut. D.:Usab.	1.467	[0.27, 2.66]	0.603	2.433	*
Aufgabe A/B	0.388	[-0.18, 0.96]	0.289	1.343	
Aufgabe B/C	-1.717	[-2.28, -1.15]	0.289	-5.949	*
Aufgabe C/D	0.671	[0.1, 1.24]	0.289	2.325	*
Random Effect Struktur: (1 <i>Versuchsperson</i>)					
$R^2_{GLMM(m)}$: 39.3 %, $R^2_{GLMM(c)}$: 53.2 %					
Effekte auf Zeitflussbeurteilung					
Intercept	4.978	[4.67, 5.29]	0.155	32.014	*
laut. D. mit/ohne	-0.369	[-0.99, 0.25]	0.311	-1.186	
Usab. hoch/niedrig	1.094	[0.48, 1.71]	0.311	3.517	*
laut. D.:Usab.	0.987	[-0.25, 2.22]	0.622	1.588	
Aufgabe A/B	0.400	[-0.23, 1.03]	0.321	1.244	
Aufgabe B/C	-0.950	[-1.58, -0.32]	0.321	-2.955	*
Aufgabe C/D	0.262	[-0.37, 0.9]	0.321	0.817	
Random Effect Struktur: (1 <i>Versuchsperson</i>)					
$R^2_{GLMM(m)}$: 9.4 %, $R^2_{GLMM(c)}$: 25.6 %					

ANMERKUNG: Die Sterne geben an, ob der Effekt auf einem Signifikanzniveau von $\alpha = .05$ und zweiseitiger Testung signifikant ist.

TABELLE 7.6: Innersubjektkorrelationen zwischen den zeit- und UX-bezogenen Variablen.

Hyp. II.		Aufgaben- dauer	geschätzte Aufg.- dauer	Dauer- bewertung	Zeitfluss- beurteilung
3 (i)-(iii)	Aufg.dauer	–	.66 ***	.47 ***	.25 ***
4 (i)-(ii)	gesch. Aufg.dauer	–	–	.48 ***	.27 ***
13	Dauerbewertung	–	–	–	.42 ***
5 (i)-(iv)	wahrg. Usability	-.20 **	-.25 ***	-.45 ***	-.28 ***
6 (i)-(iv)	wahrg. Attrakt.	-.18 **	-.19 **	-.39 ***	-.19 **
7 (i)-(iv)	Valenz	-.39 ***	-.29 ***	-.50 ***	-.30 ***
8 (i)-(iv)	Arousal	-.05, n.s	.06, n.s	-.01, n.s	-.16 *

ANMERKUNG: * bei $p < .05$, ** bei $p < .01$, *** bei $p < .001$, fett für $|r| \geq .3$

bewertet wurden (Annahme der Hyp. II.3 (i) und (ii): $r's > 0$). Darüber hinaus wurden sie mit einem langsameren Zeitfluss assoziiert (Annahme der Hyp. II.3 (iii): $r > 0$). Die geschätzte Dauer zeigte ebenfalls positive Korrelationen zur Dauerbewertung und Zeitflussbeurteilung: Je länger die Aufgabendauer geschätzt wurde, umso negativer waren Dauerbewertung und Zeitflussbeurteilung (Annahme der Hyp. II.4 (i) und (ii): $r's > 0$). Die Korrelationen mit der Zeitflussbeurteilung waren jedoch stets schwächer als die mit der Dauerbewertung. Dahingegen korrelierten die Dauerbewertung und die Zeitflussbeurteilung deutlich miteinander (Annahme der Hyp. II.13: $r > 0$): Je länger, unangemessener und inakzeptabler Probanden die Dauern bewerteten, umso eher befanden sie das Vergehen der Zeit während der Dauer als langsam. Zusammengefasst zeigten die beiden Operationalisierungen von Zeitempfinden die erwarteten Zusammenhänge zueinander, zur Dauer und zur Dauerschätzung.

Auch für die Innersubjektkorrelationen zwischen den zeit- und den UX-bezogenen Maßen zeigten sich ähnliche Ergebnisse wie in Studie I. Die wahrgenommene Usability korrelierte wie erwartet negativ mit allen zeitbezogenen Maßen (Annahme der Hyp. II.5 (i) bis (iv): $r's < 0$). Das gleiche Befundmuster zeigte sich auch für die wahrgenommene Attraktivität und die Valenz (Annahme der Hyp. II.6 und II.7 je (i) bis (iv): $r's < 0$). Die UX war also höher bei kürzeren Dauern, bei als kürzer geschätzten Dauern, als positiver

bewerteten Dauern und bei einem schnelleren Zeitfluss. Für das Arousal bot sich ein anderes Bild. Hier zeigten sich nur schwache Korrelationen zur Zeitflussbeurteilung (Annahme der Hyp. II.8 (iv): $r's \neq 0$, Beibehalten der $H_{0II.8}$ (i) bis (iii): $r's = 0$).

7.2.4 Zusätzliche Ergebnisse: Effekte der UVn auf die UX-bezogenen Variablen

Die Analyse der Effekte der unabhängigen Variablen auf die UX-bezogenen Variablen zeigte, ähnlich wie in der ersten Studie, einen starken Effekt der Usability auf die wahrgenommene Usability, die wahrgenommene Attraktivität und die Valenz (Hyp. II.9 (i), (ii), (iii): $\beta_{Usability}'s < 0$)³. Probanden, die mit der Webseite mit hoher Usability arbeiteten, bewerteten die Seite als gebrauchstauglicher und attraktiver und berichteten eine positivere Valenz (Abb. 7.4). Für das Arousal konnte kein Effekt der Usability gefunden werden (Beibehalten der $H_{0II.9}$ (iv): $\beta_{Usability} \geq 0$). Die Effekte sind in Abbildung 7.4 dargestellt. Die Analysen sind im Anhang E.6 aufgeführt. Ähnlich wie in Studie I waren die hier gefundenen Effekte der Usability auf die wahrgenommene Usability, die wahrgenommene Attraktivität und die Valenz so stark, dass sie mit der vorliegenden Stichprobengröße eine hohe Power erreichten ($1 - \beta \geq 98\%$, Anhang E.3).

³Bei den Variablen wahrgenommene Usability und wahrgenommene Attraktivität lag in den Analysen keine Normalverteilung der Residuen vor. Eine zusätzliche Analyse der Daten als lineares Modell gemittelt über die Aufgaben zeigte bei Erfüllung der Normalverteilungsannahme jedoch die gleichen Effekte.

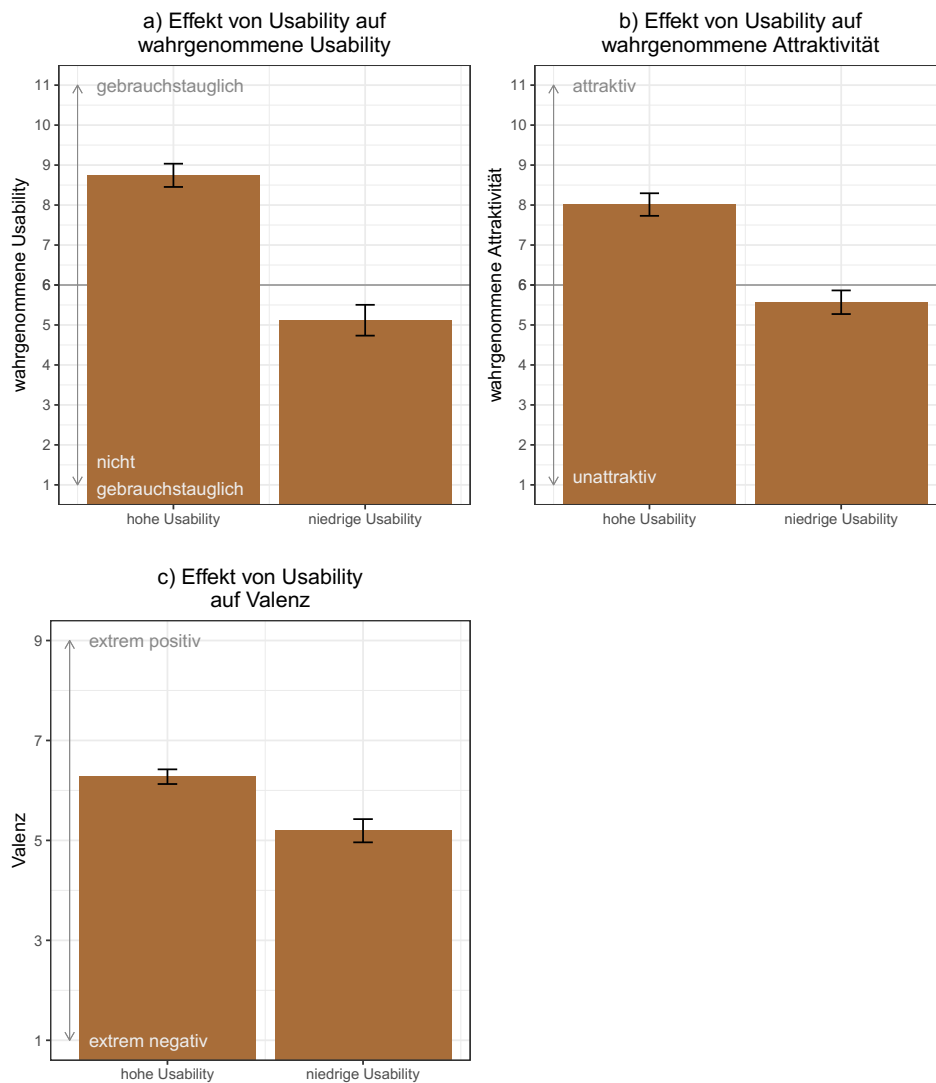


ABBILDUNG 7.4: Effekte der unabhängigen Variablen auf die Komponenten des Nutzererlebens. Die Fehlerbalken stellen Konfidenzintervalle der Standardfehler (95 %) korrigiert für Innersubjektfaktoren (Cousineau, 2005; Morey, 2008) dar.

7.3 Diskussion

In Studie II sollte die Frage beantwortet werden, in wie weit Interaktionsdauern während einer Mensch-Computer-Interaktion adäquat wahrgenommen werden können und wie sich diese Wahrnehmung in das Nutzererleben einbettet. Dazu interagierten Probanden in vier aufeinander folgenden Aufgaben mit einer Webseite. Die Usability der Webseite wurde in zwei Stufen variiert und die Probanden wurden instruiert entweder mit oder ohne der Zweitaufgabe lautes Denken zu arbeiten. Mit diesem Versuchsaufbau sollten verschiedene Interaktionscharakteristika entstehen, wovon die Interaktionsdauer eines ist. Basierend auf dem CUE-Modell (Minge et al., 2016; Minge & Thüring, 2018; Thüring & Mahlke, 2007) wurde angenommen, dass sich solche Interaktionscharakteristika auf das Nutzererleben in einer spezifischen Weise auswirken. Da Interaktionscharakteristika bis jetzt aber kaum systematisch klassifiziert und untersucht wurden, wurde dies in den Studien I und II exemplarisch für die Interaktionsdauer als eine solches Charakteristikum experimentell untersucht.

7.3.1 Zusammenfassung der Ergebnisse aus Studie I und II und Beantwortung der Forschungsteilfragen

Teilfrage a): Können Nutzer die Dauer einer Mensch-Computer-Interaktion adäquat wahrnehmen?

Ähnlich wie in Studie I zeigte sich, dass längere Dauern auch als länger geschätzt wurden, wie es auf Basis grundlagenorientierter Theorien zu erwarten war (vgl. skalare Eigenschaften von Dauerschätzungen, z.B. STT nach Gibbon, 1977). Darüber hinaus zeigten die Ergebnisse, dass die Probanden die Verlängerung der Dauer, die durch niedrige gegenüber hoher Usability entstand, adäquat wahrnahmen und sie in ihrer Dauerschätzung berücksichtigten. Dieser Schluss basiert auf den Befunden, dass sich in Studien I und II sowohl ein Effekt der Usability auf die tatsächliche und die geschätzte Aufgabendauer zeigte sowie kein Effekt der Usability auf die Verzerrung der Schätzungen. Mit anderen Worten scheinen Unterschiede in der Überschätzungsstärke, die durch die Usability des Systems entstehen, so gering zu sein, dass sie in den vorliegenden Studien nicht aufgedeckt werden konnten. Darüber hinaus wur-

de in Studie II genau wie in Studie I eine Verzerrung in der Dauerschätzung durch die Zweitaufgabe lautes Denken gefunden. Eine Gegenüberstellung der Befunde aus beiden Studien kann in Tabelle 7.7 eingesehen werden.

Zusammengefasst konnten in Studie II die Ergebnisse zu Teilfrage a) repliziert werden: Nutzer zeigen eine adäquate Wahrnehmung der Dauer unabhängig von der Usability des Systems. Ihre Dauerwahrnehmung wird jedoch durch eine Zweitaufgabe beeinflusst. In Übereinstimmung mit dem AGM (Block & Zakay, 1996) zeigten Probanden mit hoher mentaler Beanspruchung, ausgelöst durch lautes Denken, eine weniger starke Überschätzung der Dauer als Probanden mit niedriger mentaler Beanspruchung (ohne lautes Denken). Die Replikation dieser Befunde in Studie II trotz unterschiedlichen Aufgaben und Webseiten stützt eine Verallgemeinbarkeit der Ergebnisse auf verschiedene Interaktionen von Personen und Webseiten.

Teilfrage b): Wie beeinflussen die manipulierten Interaktionscharakteristika die Dauerbewertung?

Bezogen auf Teilfrage b) konnte der starke Effekt der Usability auf die Dauerbewertung ebenfalls repliziert werden (vgl. Tab. 7.7): Niedrige Usability, die mit längeren Dauern einherging, wurde bezüglich der Dauer negativer bewertet. Ein Haupteffekt des lauten Denkens, welches ebenfalls die Dauer verlängerte, konnte jedoch auch in Studie II nicht gefunden werden. In Studie II gab es jedoch eine Interaktion zwischen Usability und lautem Denken auf die Dauerbewertung. Diese zeigte sich darin, dass der Unterschied in der Dauerbewertung auf Grund der Usability stärker in der Gruppe ohne lautes Denken ausgeprägt war. Dieser Effekt ähnelt stark der Interaktion zwischen Usability und lautem Denken auf die geschätzte Aufgabendauer. Das Befundmuster legt nahe, dass die Dauerbewertung durch die wahrgenommene Dauer beeinflusst wird. Solche Zusammenhänge wurden bisher zwar für andere Operationalisierungen wie die Zeitflussbeurteilung gezeigt (Droit-Volet et al., 2017, 2018; Sucala et al., 2011), nicht aber für die Dauerbewertung. Die Ergebnisse stützen die Annahme, dass Dauerbewertungen auf einem Vergleichsprozess zwischen der geschätzten und der akzeptablen Dauer basieren, wie er bereits in Studie I diskutiert wurde.

Für das Maß Zeitflussbeurteilung wurde ebenfalls ein Effekt der Usability gefunden: Bei niedriger Usability berichteten die Probanden einen sich stärker

TABELLE 7.7: Zusammenfassung der Ergebnisse zu den Unterschiedshypothesen über die Effekte der UVn auf die AVn in Studie I und Studie II.

AV	Effektor	Studie I	Studie II
Aufgabendauer	lautes Denken	I.1 ✓	II.1 ✓
	Usability	I.2 ✓	II.2 ✓
gesch. Aufgabendauer	lautes Denken	nein	nein
	Usability	ja	II.10 ✓ in Interaktion mit lautem Denken
PTRs	lautes Denken	ja	II.11 ✓
	Usability	nein	nein
Dauerbewertung	Usability	ja	II.12 (i) ✓ in Interaktion mit lautem Denken
Zeitflussbeurteilung	Usability		II.12 (ii) ✓
Komponenten des Nutzererlebens			
wahrgen. Usability	Usability	I.9(i) ✓	II.9(i) ✓
wahrgen. Attraktivität	Usability	I.9(ii) ✓	II.9(ii) ✓
Valenz	Usability	I.9(iii) ✓	II.9(iii) ✓
Arousal	Usability	I.9(iv) –	II.9(iv) –

ANMERKUNG: ✓ für angenommene Hypothesen, ja für explorativ gefundene Ergebnisse und nein für keine Hypothese und kein Effekt.

ziehenden Zeitfluss. Dieser Effekt wurde aber anders als bei der Dauerbewertung nicht durch die Zweitaufgabe lautes Denken moderiert. Die Befunde verdeutlichen, dass die Erweiterung der Dauerbewertung um ein Bewertungsmaß des Zeitflusses sinnvoll war, da diese Variable einen anderen Aspekt des Zeiterlebens darstellt. Dies wurde darüber hinaus durch die geringere Korrelation zwischen geschätzter Aufgabendauer und Zeitflussbeurteilung und einer mittleren Korrelation zwischen Dauerbewertung und Zeitflussbeurteilung gestützt. Vor allem die geringe Korrelation zur geschätzten Aufgabendauer ist dabei mit anderen Studien vergleichbar, in denen ähnliche Korrelationskoeffizienten gefunden wurden (Sucala et al., 2011).

Teilfrage c): Welchen Zusammenhang gibt es zwischen der Interaktionsdauer und verschiedenen Maßen der Dauerwahrnehmung innerhalb von Nutzern?

Wie oben bereits diskutiert, zeigten sich ähnlich hohe Zusammenhänge⁴ zwischen den zeitbezogenen Maßen in Studie II wie in Studie I. Mit steigender Interaktionsdauer wurden längere Dauerschätzungen abgegeben (vgl. skalare Eigenschaften von Dauerschätzungen), die Dauer negativer bewertet und ein langsamer Zeitfluss berichtet. Für die Dauerschätzung war der Zusammenhang zur Dauer jedoch stärker als für die beiden anderen Maße. Darüber hinaus zeigten die verschiedenen Operationalisierungen der Dauerwahrnehmung (Dauerschätzung, Dauerbewertung und Zeitflussbewertung) nur schwache bis mittlere Korrelationen zueinander. Dieses Befundmuster deutet darauf hin, dass sie verschiedene Aspekte der Wahrnehmung einer Interaktionsdauer abbilden. So zeigte sich zwar in Studie II, dass ein schnellerer Zeitfluss mit kürzeren Dauerschätzungen einherging, hier lag aber eine sehr geringe Varianzaufklärung vor. Dieser Befund legt nahe, dass die Schätzung einer mehrminütigen Dauer nicht unbedingt mit dem erlebten Zeitfluss verbunden ist. Wearden et al. (2014) argumentierten bereits, dass Probanden die exakte Dauer eines Intervalls kennen können und trotzdem eine davon unabhängige Zeitflussbeurteilung äußern können. Ein Beispiel hierfür bietet die Studie von Gorn et al. (2004), in der der Effekt der Hintergrundfarbe einer Webseite während Wartezeiten untersucht wurde. Hier zeigte sich kein Einfluss der Farbe auf die geschätzte Dauer, wohl aber auf die Zeitflussbeurteilung. Die Ergebnisse von Studie II sind mit solchen Befunden im Einklang und weisen darauf hin, dass der Zusammenhang zwischen Zeitflussbeurteilung und Dauerschätzungen nicht so einfach ist, dass ein Empfinden von langsam vergehender Zeit automatisch zu einer längeren Dauerschätzung führt.

Zusammengefasst deuten die Befunde darauf hin, dass die geschätzte Interaktionsdauer nicht alleine ausreicht, um das Zeitempfinden vorherzusagen. Dieses scheint durch weitere Interaktionscharakteristika beeinflusst zu werden. Auf Basis des starken Effekts der unabhängigen Variable Usability auf

⁴Die Innersubjektkorrelationen aus Studien I und II sind zusammen mit den Korrelationen aus Studien III und IV in der abschließenden Diskussion zusammengefasst dargestellt, Tab. 10.1, S. 169.

die beiden Konstrukte kann erwartet werden, dass dies Interaktionscharakteristika sind, die durch Merkmale des Systems und nicht durch Merkmale der Aufgabe manipuliert werden.

Teilfrage d): Welche Zusammenhänge gibt es zwischen den Maßen der Dauerwahrnehmung und den Komponenten des Nutzererlebens innerhalb von Nutzern?

Auch für den Zusammenhang zwischen zeit- und UX-bezogenen Variablen zeigten sich ähnliche Ergebnisse zu Studie I. Auch wenn in Studie II mehr Korrelationen Signifikanz erreichten, so sind doch nur spezifische Zusammenhänge auf Grund ihrer Stärke interessant. Genau wie in Studie I sind dies die Zusammenhänge zwischen Dauerbewertung und den Komponenten des Nutzerlebens: Probanden bewerteten die wahrgenommene Usability, die wahrgenommene Attraktivität und die Valenz als negativer, je negativer sie die Dauerbewertung einschätzen. Die Zeitflussbeurteilung zeigte geringere Zusammenhänge zu den UX-bezogenen Variablen, die jedoch die gleiche Richtung aufwiesen. Diese Ergebnisse zeigen auf, dass vor allem die Bewertung der Dauer eine ähnliche Bewertungsdimension zu den Komponenten des Nutzererlebens ist. Dabei stellt sich die Frage, in wie weit die Dauerbewertung durch die Wahrnehmung der Dauer und/oder die Wahrnehmung des eigenen Nutzererlebens beeinflusst wird. Diese Frage kann weder mit Studie I noch II beantwortet werden, da hier die Dauer unter anderem durch die Usability des Systems beeinflusst wurde. Da die Usability aber auch ein wesentlicher Einflussfaktor für das Nutzererleben ist (z.B. Ben-Bassat et al., 2006; Hamborg et al., 2014; Minge & Thüring, 2018; Thüring & Mahlke, 2007), liegt eine Konfundierung vor, die die Abgrenzung der Effekte verhindert.

Bezüglich der Korrelationen zum Arousal-Level zeigten sich in beiden Studien keine signifikanten Zusammenhänge zur Aufgabendauer, geschätzten Aufgabendauer oder Dauerbewertung. Lediglich zur Zeitflussbeurteilung wurde in Studie II ein schwacher Zusammenhang von erhöhtem Arousal mit schnellerem Zeitfluss signifikant, der auch schon von Droit-Volet et al. (2018) berichtet wurde. Insgesamt unterschied sich das Korrelationsmuster zwischen Dauerwahrnehmung und Arousal deutlich vom Muster der Zusammenhänge zwischen Dauerwahrnehmung und den UX-Komponenten wahrgenommene Usa-

bility, wahrgenommene Attraktivität und Valenz. Darüber hinaus korrelierte das Arousal-Level nicht mit den anderen UX-Komponenten, wohingegen diese überwiegend mittlere bis starke Zusammenhänge zueinander zeigten (Tab. D.4 und Tab. E.5 im Anhang). Auf Basis dieser Daten deutet sich an, dass das Arousal keinen eindeutigen Zusammenhang zur Wahrnehmung der instrumentellen und nicht-instrumentellen Qualitäten hat. Dadurch unterscheidet es sich deutlich von der Valenz.

7.3.2 Limitationen von Studien I und II - Ableitung von Studien III und IV

Durch das experimentelle Design in den Studien I und II konnten einige Schlüsse bezüglich der Wahrnehmung von Dauern während einer Mensch-Computer-Interaktionen gezogen werden. Es bestehen jedoch weiterhin offene Forschungsfragen, die im folgenden für ausgewählte Punkte diskutiert werden. Dabei handelt es sich um Effekte des Zusammenspiels der unabhängigen Variablen, um die zweistufige Manipulation der Usability in beiden Studien, um die gleichzeitige Erhebung der korrelierten Maße und um mögliche Konfundierungen durch die variierende Aufgabendauer.

Zusammenspiel der unabhängigen Variablen: Konfundierung zwischen Usability und mentaler Beanspruchung

Ein Kritikpunkt für die beiden ersten Studien liegt in der möglichen Konfundierung der unabhängigen Variablen Usability und lautes Denken als Zweit-aufgabe. So kann die mentale Beanspruchung nicht nur durch das laute Denken beeinflusst worden sein, sondern auch durch die Usability: Eine Aufgabe bei schlechter Usability zu lösen, erfordert oft längeres Suchen nach der richtigen Handlung oder ein stärkeres Hinterfragen von ungünstig gewählten Begrifflichkeiten. Hinweise auf Zusammenhänge zwischen Usability und mentaler Beanspruchung sind auch in der Definition von Usability nach der DIN-Norm (DIN, EN ISO 9241-11, 2018) sowie in der Norm zur Dialoggestaltung (DIN, EN ISO 9241-110, 2008) zu finden. Hier wird die Usability über Effizienz definiert, die wiederum über aufgebrauchte Ressourcen bezüglich geistigem und physischem Aufwand beschrieben wird.

Da die mentale Beanspruchung nicht als Kontrollvariable erhoben wurde, kann diese Konfundierung nicht direkt geprüft werden. Ein schwaches Indiz, welches jedoch gegen eine solche Konfundierung spricht, ist der fehlende Effekt der Usability auf die Verzerrungen in der Zeitschätzung. Hätten die beiden Usabilitystufen einen starken Unterschied in der mentalen Beanspruchung ausgelöst, so hätte dies zu einer Verzerrung in der Zeitschätzung führen müssen (vgl. Interferenzeffekt, Brown, 1997). Da dies weder in Studie I noch in Studie II der Fall war, kann davon ausgegangen werden, dass die Usability zumindest keinen großen Effekt auf die mentale Beanspruchung hatte, welche primär durch die Zweitaufgabe lautes Denken beeinflusst wurde.

Eine weitere mögliche Konfundierung kann aber auch in die andere Richtung vorliegen: In der Gruppe mit lautem Denken können durch das Verbalisieren der Usabilityprobleme diese stärker wahrgenommen worden sein als in der Gruppe ohne lautes Denken. Dies würde dazu führen, dass der Unterschied zwischen den Usabilitygruppen in der Gruppe mit lautem Denken größer wäre als in der Gruppe ohne lautes Denken. Um dies näher zu untersuchen, können die Ergebnisse der zusätzlichen Analysen zum Effekt der unabhängigen Variablen auf die Komponenten des Nutzerlebens betrachtet werden. Hier zeigte sich in Studien I und II keine Interaktion zwischen lautem Denken und Usability auf die wahrgenommene Usability, die wahrgenommene Attraktivität und die Valenz. Zusammenfassend scheint eine Konfundierung der manipulierten Usability und Zweitaufgabe in Studien I und II nicht in einer relevanten Stärke vorgelegen zu haben.

Zweistufige, experimentelle Manipulation der Usability des Systems

Sowohl in Studie I als auch in Studie II wurden Innersubjektkorrelationen betrachtet, das heißt es wurde der Zusammenhang zwischen zeitbezogenen und zwischen zeit- und UX-bezogenen Maßen innerhalb von Probanden betrachtet. Jeder Proband erlebte jedoch immer nur eine Usability-Stufe. Dadurch war die Varianz im Erleben innerhalb eines Probanden stark eingegrenzt, wodurch korrelative Zusammenhänge zwischen zeit- und UX-bezogenen Variablen unterschätzt worden sein könnten. In realen Situationen haben Nutzer einen großen Erfahrungsschatz an bereits erlebten Situationen, die sich bezüglich der Systeme und deren Usability stark unterscheiden. In dieses Vorwissen können

Nutzer eine gerade erlebte Nutzungssituation einbetten und den Vergleich zur Bewertung und Einschätzung ihres Erlebens nutzen. Daher stellt sich die Frage, wie hoch die betrachteten Zusammenhänge sind, wenn ein Proband nicht nur mit einer Usability-Stufe konfrontiert ist, sondern mehrere Abstufungen in verschiedenen Aufgaben und einem realistischeren Setting erlebt. Um diese Folgefrage aus Studien I und II mit Hilfe einer höheren Varianz an Systemen zu beantworten, wird Studie III durchgeführt.

Gleichzeitige Erhebung der korrelierten Maße: Common-Method-Bias

Zusätzlich soll in Studie III für einen weiteren Schwachpunkt der ersten beiden Studien kontrolliert werden: Hier wurden die korrelierten Maße immer zum gleichen Messzeitpunkt erhoben. Eine solche Erhebungsart birgt die Gefahr des *Common-Method-Bias* (Podsakoff, MacKenzie, Lee & Podsakoff, 2003) oder zu deutsch einer Einheitsmethodenvarianz. Der Common-Method-Bias beschreibt die gegenseitige Beeinflussung von Messwerten, wenn Variablen gleichzeitig gemessen werden. Gerade bei stark subjektiv beeinflussten Maßen, wie hier den Bewertungen, kann alleine durch die direkt aufeinanderfolgende Messung von zwei verschiedenen Konstrukten eine Ähnlichkeit zwischen den Messwerten entstehen. Diese ist jedoch nicht auf einen grundlegenden Zusammenhang zwischen den Konstrukten zurückzuführen, sondern auf die zeitliche Nähe der Messung (Podsakoff, MacKenzie & Podsakoff, 2012). Zusätzlich kann die Itemgestaltung (*Item Characteristic Effects*) sowie die Ähnlichkeit in der Gestaltung verschiedener Items (*Item Context Effects*) die Entstehung eines Common-Method-Bias begünstigen und Probanden dazu ermutigen, ihre eigenen Bewertungen zu verfälschen (Podsakoff et al., 2003). Eine mögliche Konsequenz ist eine Über- oder aber auch eine Unterschätzung von Zusammenhängen. Um die mögliche Konfundierung der Ergebnisse zu verringern, wird in Studie III zum einen eine zeitliche Lücke zwischen der Erhebung verschiedener Konstrukte gelegt und zum anderen Veränderungen der Operationalisierung vorgenommen.

Variierende Aufgabendauer

In Studien I und II kam es zu variierenden Aufgabendauern, da die Probanden die Aufgaben in ihrem eigenen Tempo erledigten und selbstständig nach

erfolgreichem Abschluss oder nach Aufgeben beendeten. Die Aufgabendauern unterlagen dabei systematischen Schwankungen, wie den Effekten von lautem Denken, Usability und den Aufgaben an sich. Es gab jedoch auch zufällige Schwankungen, die durch die jeweilige Versuchsperson entstanden. Diese Ausgangslage muss bei der Interpretation der Dauerbewertung beachtet werden. In beiden Studien wurden die Aufgabendauern bei niedriger Usability negativer bewertet als bei hoher Usability. Tatsächlich waren die Aufgabendauern bei niedriger Usability aber auch länger, weswegen eine negativere Bewertung nicht verwunderlich ist. Die Verlängerung der Aufgabendauern in der Gruppe mit lautem Denken hingegen, spiegelte sich in beiden Studien nicht in den Dauerbewertungen wieder.

Eine mögliche Erklärung für dieses Ergebnis liegt in der Annahme, dass Dauerbewertung eine Bewertung der erlebten Dauer und nicht der tatsächlichen Dauer darstellen. Diese Annahme ist im Einklang mit den Daten, wie bereits weiter oben diskutiert wurde. Eine alternative Erklärung liegt im Zusammenhang zwischen Dauerbewertungen und den Komponenten des Nutzererlebens. Die Bewertung der Dauer kann weniger auf der wahrgenommenen Dauer, als auf anderen wahrgenommenen Erlebenskomponenten beruhen. Beispielsweise kann das Gefühl einer problemlosen Lösung der Aufgaben dazu führen, dass Probanden auch die Dauer positiver bewerteten. Da die längere Aufgabendauer hier mit der niedrigen Gebrauchstauglichkeit des Systems und damit mit einem negativeren Erleben konfundiert ist, kann keine der beiden Erklärungsansätze ausgeschlossen werden. Daher werden beide Ansätze in Studie IV gegeneinander getestet werden, wobei eine Konfundierung durch die variierende Bearbeitungsdauern ausgeschlossen wird.

7.3.3 Fazit Studien I und II

Zusammengefasst haben Studien I und II gezeigt, dass Nutzer die Interaktionsdauer mit einer Webseite adäquat einschätzen können, es aber durch Variationen in der mentalen Beanspruchung zu Fehleinschätzungen kommen kann. Diese stimmen mit Vorhersagen aus Grundlagenarbeiten zum Interferenzeffekt (z.B. Brown, 2008) und Modellen wie dem AGM (z.B. Zakay & Block, 1995) überein. Darüber hinaus konnten Nutzer in beiden Studien die Dauer nicht nur einschätzen, sondern auch bewerten sowie ihr Erleben des Zeitflusses angeben.

Diese Bewertungen scheinen teilweise auf der geschätzten Dauer zu beruhen, zeigen aber auch überlappende Varianz mit Komponenten des Nutzererlebens.

Studien I und II haben zwei Fragen aufgeworfen: Erstens, wie groß sind die Zusammenhänge zwischen zeit- und UX-bezogenen Variablen, wenn die Probanden realistischere Aufgaben an echten Webseiten mit mehr Varianz bearbeiten? Diese Frage wird mit Hilfe von Studie III beantwortet. Zweitens, worauf basiert das Zeitempfinden, wird es stärker von der erlebten Zeitdauer oder von anderen Erlebenskomponenten wie den Komponenten des Nutzererlebens beeinflusst? Diese Frage soll zusammen mit einer erneuten Betrachtung der Korrelationen in Studie IV beantwortet werden, in der der die Bearbeitungsdauern konstant gehalten werden.

Kapitel 8

Studie III: Korrelationen zwischen Dauerwahrnehmung und UX bei „echten Interaktionen“

In den vorangegangenen Studien wurden Zusammenhänge zwischen zeit- und UX-bezogenen Maßen immer nach Interaktionen mit einem System mit spezifischer Bedienbarkeit betrachtet. Dabei zeigten sich überwiegend mittlere Zusammenhänge zwischen der Dauerbewertung und verschiedenen Komponenten des Nutzererlebens. In Studie III wird geprüft, ob sich die Stärke dieser Inner-subjektkorrelationen ändert, wenn jeder Proband eine größere Variation an Systemen erlebt. Dazu sollen Aufgaben auf verschiedenen, echten Webseiten genutzt werden, die in ihrer Bedienbarkeit variieren. Mit Hilfe dieser Betrachtung lässt sich prüfen, wie Nutzer die Wahrnehmung von Dauern in ihr Nutzererleben integrieren (Forschungsfrage 2). Die in dieser Studie relevante Teilfrage lautet daher:

- a) Wie stark sind die Zusammenhänge zwischen der Dauerwahrnehmung und den Komponenten des Nutzererlebens, wenn Probanden realistische Aufgaben an echten Webseiten bearbeiten?

Darüber hinaus soll das Erleben und das Bewerten der Dauer bzw. des Zeitflusses einer Komponente des Nutzererlebens zugeordnet werden. Hierfür kommen jedoch nur bestimmte Komponenten in Frage. Nach dem CUE-Modell (Mahlke, 2008; Minge & Thüring, 2018; Thüring & Mahlke, 2007) entsteht Nutzererleben

auf Basis der Wahrnehmung der Interaktionscharakteristika. Nutzer nehmen hierbei die instrumentellen Qualitäten, die nicht-instrumentellen Qualitäten und ihre eigene Emotion wahr. Die Wahrnehmung der Zeit kann auf Basis dieser theoretischen Grundannahmen nicht jeder Komponente zugeordnet werden. So ist es konzeptuell nicht plausibel anzunehmen, dass die Wahrnehmung der Interaktionsdauer Teil der Komponente Emotion ist, da Emotionen ein klar beschriebenes Konstrukt sind (Scherer, 2005), zu dem die Wahrnehmung von Dauern nicht gehört.

Daher stellt sich die Frage, ob die Dauerwahrnehmung eher als Teil der Komponente instrumentelle oder nicht-instrumentelle Qualitäten gesehen werden kann. Inhaltlich scheint das schnelle Erreichen eines Interaktionsziels sehr wohl aufgabenrelevant, weswegen es zu den instrumentellen Qualitäten gezählt werden könnte. Diese Zuordnung wird auch empirisch durch die vorangegangenen Studien gestützt. Hier korrelierten die zeitbezogenen Maße stärker mit den instrumentellen als mit den nicht-instrumentellen Qualitäten. Auch wenn die Unterschiede nicht groß waren, können die Befunde als Indizien dafür gesehen werden, dass die Wahrnehmung der Interaktionsdauer primär als eine Wahrnehmung instrumenteller Qualitäten verstanden werden kann.

Auf Grund des Studiendesigns in Studien I und II sollten diese Indizien jedoch nicht überinterpretiert werden. Bei den beiden vorangegangenen Studien bestand auf Grund der gleichzeitigen Messung der korrelierten Maße die Gefahr eines Common-Method-Bias (Podsakoff et al., 2003). In Studie III soll diesem vorgebeugt werden, um die Korrelationen zwischen Dauerwahrnehmung und Komponenten des Nutzererlebens präziser abzubilden. Damit werden drei Erwartungen getestet. Erstens, auf Grund des CUE-Modells (Mahlke, 2008; Minge & Thüring, 2018; Thüring & Mahlke, 2007) wird angenommen, dass die Wahrnehmung der Usability in direktem Zusammenhang zur wahrgenommenen Attraktivität steht. Zweitens, Auf Basis der vorangegangenen Studien wird zudem die Hypothese aufgestellt, dass die wahrgenommene Usability mit den zeitbezogenen Maßen verknüpft ist. Drittens wird vermutet, dass Zusammenhänge zwischen wahrgenommener Attraktivität und zeitbezogenen Maßen in den vorangegangenen Studien über die wahrgenommene Usability vermittelt wurden. Daher wird im Studiendesign eine zeitliche Lücke zwischen der Erhebung der wahrgenommenen Usability und der Erfassung der anderen UX-

Konstrukte erzeugt. Zusätzlich soll ein größerer Unterschied in der Itemgestaltung für die Maße wahrgenommene Usability und zeitbezogene Variablen geschaffen werden. Diese Anpassungen der Messmethode dienen dazu, die folgende Teilfrage (Forschungsfrage 2) zu beantworten:

- b) Wie verändern sich die Zusammenhänge zwischen zeit- und UX-bezogenen Maßen im Vergleich zu den vorangegangenen Studien, wenn sich die Messung der wahrgenommenen Usability von der Messung der anderen Konstrukte zeitlich und von der Darstellung absetzt?

Zusammenfassend sollen in Studie III die betrachteten Korrelationen durch das neue Versuchsmaterial valider und durch das Ausschließen eines Common-Method-Bias für das Konstrukt wahrgenommene Usability reliabler als in den vorangegangenen Studien geschätzt werden¹.

8.1 Methode

Studie III wurde im Rahmen der Masterarbeit von Anna Steinbach getestet. Die Idee zur Studie und die experimentelle Planung stammten von der Verfasserin der vorliegenden Arbeit. Im Rahmen der Betreuung der Masterarbeit durch die Verfasserin hat Anna Steinbach (Steinbach, 2017) die Aufgaben für die Probanden ausgesucht und sie zusammen mit den Fragebögen für die Studie aufbereitet. Steinbach (2017) hat zudem die Daten erhoben. Alle hier präsentierten Analysen und Beschreibungen wurden von der Verfasserin der vorliegenden Arbeit ausgeführt und beschrieben.

8.1.1 Teilnehmer

An Studie III nahmen $N = 33$ Probanden teil, wovon 20 weiblich und 13 männlich waren. Im Mittel waren die Probanden $M = 27,5$ Jahre ($SD = 3.9$) alt. Da in einigen Aufgaben mit einer englisch-sprachigen Webseite interagiert werden sollte, wurden die Probanden gebeten einzuschätzen, wie gut

¹Die externe Validität im Sinne einer Übereinstimmung von tatsächlichem und intendiertem Untersuchungsgegenstand wird durch ein realistischeres Versuchsmaterial erhöht. Die Reliabilität, also die Verlässlichkeit der Messung, wird durch das Ausschließen einer Ursache für Messfehler gesteigert.

ihre Englischkenntnisse waren. Alle Probanden gaben an, dass sie mindestens Grundkenntnisse hätten.

Die Probanden erhielten für die Teilnahme am Experiment eine Versuchspersonenstunde. Vor der Studie wurde im Rahmen des Fast-Track-Verfahrens der Ethikkommission des Instituts für Psychologie und Arbeitswissenschaft an der TU Berlin das Forschungsvorhaben geprüft (Stellungnahme der Ethikkommission siehe Anhang F.2).

8.1.2 Studiendesign

Die Teilnehmer wurden zu Beginn der Studie darüber informiert, dass sie an einer Studie zum Zeitempfinden bei der Nutzung von Webseiten teilnehmen würden. In diesem Rahmen wurden ihnen nacheinander vier verschiedene Aufgaben auf je einer real-existierenden Webseite gegeben, die sie zunächst bearbeiten und danach bewerten sollten. Jede Aufgabe war dabei spezifisch für eine Webseite. Da beides miteinander konfundiert war, werden die Begriffe Webseite und Aufgabe im Folgenden synonym zueinander verwendet.

Ein browser-basierter Fragebogen führte die Probanden durch den Prozess der Bearbeitung. Bei jeder Aufgabe gab es eine kurze Instruktion und einen Link, mit dem die Aufgabe gestartet werden konnte. Sobald der Proband die Aufgabe gelöst hatte, aufgeben wollte oder das Zeitlimit von fünf Minuten erreicht hatte, sollte er zum Fragebogen zurückkehren und dort die entsprechenden Fragen beantworten. Die vier Aufgaben wurden auf Basis einer Vorstudie (Steinbach, 2017) nach den Kriterien Lösbarkeit und Unterschiede in der subjektiven Usability ausgewählt. Das Zeitlimit von fünf Minuten für jede Aufgabe wurde ebenfalls auf Grund der Vorstudie gewählt. Unterschiede in der subjektiven Usability zwischen den Aufgaben waren gewünscht, um eine Variation von verschiedenen Usability-Stufen zu realisieren. Die Präsentation der vier verschiedenen Aufgaben ist somit die einzige experimentelle Manipulation. Sie wird hier jedoch nicht als eine klassische unabhängige Variable gesehen, da sich die Fragestellung der vorliegenden Studie auf korrelative Zusammenhänge zwischen verschiedenen Bewertungsmaßen bezieht und nicht auf Unterschiede zwischen Stufen einer experimentellen Variation. Diese Unterschiede sind lediglich im Rahmen des Manipulationschecks relevant.

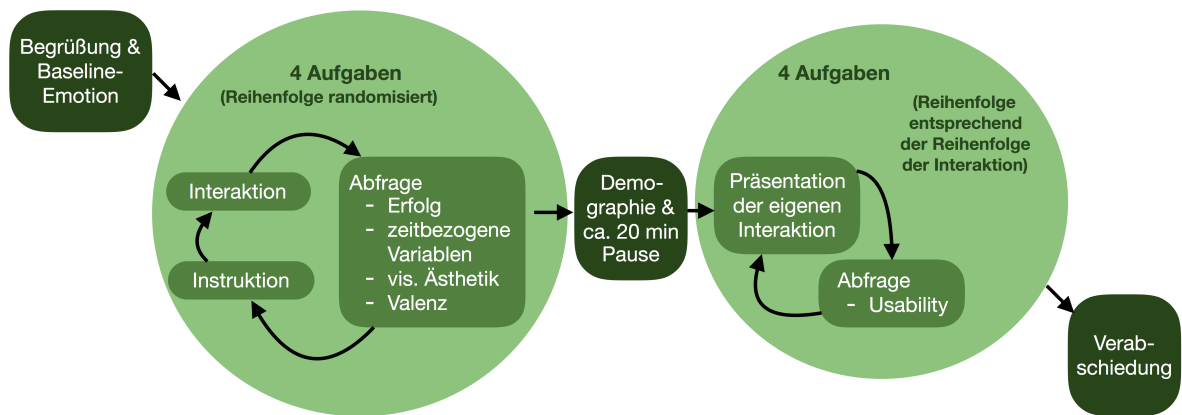


ABBILDUNG 8.1: Übersicht über den Versuchsablauf in Studie III.

8.1.3 Versuchsablauf und Material

Die Probanden wurden begrüßt und erhielten eine Instruktion und die Einverständniserklärung zur Teilnahme an der Studie sowie zur Weiterverarbeitung der Daten. Danach durchliefen sie in randomisierter Reihenfolge die vier Aufgaben (Randomisierung siehe Anhang F.1.1), wobei die Bearbeitung jeder Aufgabe mit Hilfe eines Screen-Recorders aufgenommen wurde. In Tabelle 8.1 sind die vier Aufgaben beschrieben. Nach jeder Aufgabe gaben die Probanden an, ob sie die Aufgabe lösen konnten und es erfolgte die Erhebung der zeitbasierten Maße sowie von Teilen der UX-bezogenen Maße. Nach der Bearbeitung aller Aufgaben wurden die demographischen Variablen erhoben. Nach einer mindestens 20-minütigen Pause, in denen die Probanden zwei Online-Fragebögen zu anderen Forschungsvorhaben ausfüllten, wurde für jede Aufgabe ein Video der eigenen Interaktion vorgespielt und die Usability der jeweiligen Webseite auf einem Papierfragebogen bewertet. Dieses Vorgehen wurde gewählt, um einen zeitlichen Abstand zwischen der Bewertung der wahrgenommenen Usability und den anderen Maßen zu erzeugen und damit dem Common-Method-Bias vorzubeugen. Durch das Vorspielen der eigenen Interaktion sollte eine retrospektive Bewertung der Usability ermöglicht werden (Burmester, Jäger, Mast, Peissner & Sproll, 2010), in dem erlebte Schwierigkeiten erneut visuell beobachtet werden konnten. Der Versuchsablauf der Studie ist in Abbildung 8.1 dargestellt.

TABELLE 8.1: Inhalt der vier Aufgaben in Studie III.

Aufgabe A	Lingscars - Telefonnummer https://www.lingscars.com Bitte Suchen Sie die Telefonnummer des Eigentümers. Bitte benutzen Sie keine Suchfunktion!
Aufgabe B	Universität Groningen - Sekretariat für Organisationspsychologie https://www.rug.nl Bitte Suchen Sie die Telefonnummer des Sekretariats für Organisationspsychologie der Universität Groningen (Englische Webseite). Bitte nutzen Sie nicht die Suchfunktion!
Aufgabe C	TK - Verwaltungsrat https://www.tk.de/techniker Wie viele Mitglieder sind im Verwaltungsrat der Techniker Krankenkasse (TK)? Bitte nutzen Sie keine Suchfunktion!
Aufgabe D	Lieferando - Pasta Carbonara https://www.lieferando.de Wie viel kostet ein Pastagericht Carbonara von Vapiano im Gebiet der Machstraße, Berlin Charlottenburg (PLZ 10587) mit Penne Nudeln und extra Artischocken, Grana Padano und getrockneten Tomaten? Bitte suchen Sie den Gesamtpreis mit Versandkosten! Bitte nutzen Sie nur die Suchfunktion auf der Startseite um die Postleitzahl einzugeben! Sie müssen nichts bestellen oder irgendwo andere Daten als die Postleitzahl angeben! Sollte der Laden noch nicht geöffnet haben können Sie das Gericht "vorbestellen". Ihnen entstehen dadurch keine Kosten, wenn Sie nicht auf Bestellen klicken!

8.1.4 Operationalisierung der abhängigen Variablen

Die abhängigen Variablen umfassten vergleichbar zu den vorangegangenen Studien zeitbasierte und UX-basierte Maße. Die zeitbasierten Maße Aufgabendauer, geschätzte Aufgabendauer, Dauerbewertung und Zeitflussbeurteilung wurden wie in Studie II beschrieben operationalisiert. Um ein ausgeglicheneres Itemverhältnis bei der Messung der Dauerbewertung und der Zeitflussbeurteilung zu erhalten, wurden hier jedoch zwei Veränderungen durchgeführt. Zum einen wurde das Item wahrgenommene Unangemessenheit der Dauer nicht mehr erhoben, da es sehr hohe Ähnlichkeiten zum Item Inakzeptabilität der Dauer zeigte ($r_{Studie\ 1} = .81$ und $r_{Studie\ 2} = .84$) und dadurch eine redundante Messung darstellte. Zum anderen wurde die Zeitflussbeurteilung um ein zweites Item (*wahrgenommenes Ziehen der Zeit*) ergänzt, welches auf einem semantischen Differential mit 11 Abstufungen von „Die Zeit ist verfliegen.“ (1) zu „Die Zeit hat sich gezogen.“ (11) basierte². Dieses Maß entspricht einer Operationalisierung, wie sie von anderen Autoren bereits genutzt wurde (z.B. Sackett et al., 2010; Sucala et al., 2011). Die Fragebogenitems zu den zeit-basierten Maßen können im Anhang F.1.2) eingesehen werden.

Die UX-basierten Maße umfassten die wahrgenommene Usability, die wahrgenommene visuelle Attraktivität und die Valenz. Die wahrgenommene visuelle Attraktivität und die Valenz wurden nach jeder Aufgabe mit Hilfe des VisAWi-Shorts (Moshagen & Thielsch, 2013) und des Affect Grids (Russell et al., 1989) gemessen. Die wahrgenommene Usability wurde nicht nach jeder Aufgabe erhoben, sondern retrospektiv nach einer Pause von mindestens 20 Minuten mit Hilfe von Videos der eigenen Interaktion. Hierzu wurde Modul 1 des meCUEs (Minge et al., 2016) genutzt. Dieses Modul umfasst Items zur wahrgenommenen Usability und zur wahrgenommenen Nützlichkeit. Im Folgenden werden jedoch nur die Daten der wahrgenommenen Usability berichtet, da nur diese auch in den anderen Studien erhoben wurde.

Zur Kontrolle der Aufgabenerfüllung wurden die Probanden nach jeder Aufgabe gebeten eine Antwort zur Aufgabe abzugeben (*objektiver Erfolg*) sowie anzugeben, ob sie die Aufgabe lösen konnten (*selbstberichteter Erfolg*). Darüber hinaus wurde die Bekanntheit der Webseite abgefragt. Die Einver-

²Das Item wahrgenommenes Ziehen der Zeit wurde zum besseren Vergleich für die Beschreibung und Analyse gegenüber der Erhebung invertiert.

ständniserklärung und die Bewertung der wahrgenommenen Usability lagen in Papierform vor, während der restliche Fragebogen online mit Hilfe von SoSci-Survey (<http://soscisurvey.de>) dargeboten und erhoben wurde.

8.1.5 Hypothesen

Manipulationscheck: Effekte der Aufgaben

Die vier Aufgaben mit den zugehörigen Webseiten wurden in einer Vorstudie ausgewählt, da sie sich in ihrer wahrgenommenen Usability unterschieden. Ob dies auch in dieser Stichprobe so wahrgenommen wurde, soll mit Hilfe eines Manipulationschecks geprüft werden. Auf Basis der Vorstudie wird dazu die folgende Hypothese als Manipulationscheck aufgestellt, die einem wiederholten Kontrast (siehe Abschnitt B.1.2) entspricht:

Hypothese III.MC (i) bis (iii): Die wahrgenommene Usability ist geringer (i) für Aufgabe A (LingsCars) im Vergleich zu Aufgabe B (Uni Groningen), (ii) für Aufgabe B (Uni Groningen) im Vergleich zu Aufgabe C (TK) und (iii) für Aufgabe C (TK) im Vergleich zu Aufgabe D (Lieferando)
(i) $\text{wahrg. Usab.}_{\text{Aufgabe A}} < \text{wahrg. Usab.}_{\text{Aufgabe B}}$
(ii) $\text{wahrg. Usab.}_{\text{Aufgabe B}} < \text{wahrg. Usab.}_{\text{Aufgabe C}}$
(iii) $\text{wahrg. Usab.}_{\text{Aufgabe C}} < \text{wahrg. Usab.}_{\text{Aufgabe D}}$
Dies zeigt sich in positiven Steigungen: $\beta_{\text{Kontrast}} 's > 0$.

Zusätzlich soll im Kontext des Manipulationschecks betrachtet werden, ob es Unterschiede zwischen den Aufgaben in Bezug auf die Aufgabendauer und die Schätzverzerrungen der Aufgabendauer gibt. Dazu werden die Variablen Aufgabendauer sowie Perceived Time Ratios (PTRs) als Verhältnis zwischen geschätzter und tatsächlicher Dauer genutzt. Darüber hinaus wird der Bekanntheitsgrad der Webseiten hinsichtlich der Frage betrachtet, ob sich die Webseiten unterscheiden und ob dies einen Einfluss auf die wahrgenommene Usability hat. Ein solcher Einfluss könnte die Interpretierbarkeit der Ergebnisse erschweren. Abschließend wird der objektive und der selbstberichtete Erfolg der Probanden bei der Aufgabenerledigung betrachtet.

Hypothesen zu den Teilfragen a) und b): Innersubjektkorrelationen zwischen zeit- und UX-bezogenen Maßen

Da in dieser Studie die korrelativen Zusammenhänge im Mittelpunkt stehen, werden in Anlehnung an die Studien I und II die Hypothesen zu den Innersubjektkorrelationen zwischen den zeitbezogenen Maßen aus den vorangegangenen Studien übernommen. Diese sind in Tabelle 7.3 im vorangegangenen Kapitel zusammengefasst. Sie umfassen auf der einen Seite Zusammenhänge zwischen den zeitbezogenen Maßen Aufgabendauer, geschätzte Aufgabendauer, Dauerbewertung und Zeitflussbeurteilung. Auf der anderen Seite beinhalten sie Hypothesen zu den Zusammenhängen zwischen diesen zeitbezogenen und den UX-bezogenen Variablen wahrgenommene Usability, wahrgenommene Attraktivität, Valenz und Arousal.

8.1.6 Datenanalyse

Der Effekt der Aufgabe auf die wahrgenommene Usability, die Aufgabendauer und die Verzerrung der Schätzungen (PTRs) werden mit Hilfe eines wiederholten Kontrasts (*repeated contrast*, Schad et al., 2018) in einem LMM mit Maximum Likelihood Schätzung (ML) getestet. Dabei werden auf Seiten der Fixed Effects der Intercept als Gesamtmittelwert und die drei Kontraste entsprechend der Vergleiche in Hypothese III.MC (i) bis (iii) kodiert. Auf Seiten der Random Effects werden ein Random Intercept sowie drei Random Slopes entsprechend der Kontraste spezifiziert. Die Random Effect Struktur wird wieder nach dem in Abschnitt B.1.3 beschriebenen Vorgehen selektiert (Bates, Kliegl et al., 2015). Das Prüfen der Korrelationshypothesen erfolgt erneut durch die Berechnung von Innersubjektkorrelationen. Diese Werte geben an, wie stark die jeweils betrachteten Messwerte eines Probandens miteinander kovariieren, wenn der Proband in variierenden Situationen (verschiedene Aufgaben auf unterschiedlichen Webseiten) Messwerte liefert, Teilfrage a).

Genutzte Software

Zur Analyse der Daten wurde das Statistik-Programm R (R Core Team, 2018) und die folgenden Pakete genutzt: car (Fox & Weisberg, 2011), corrplot (Wei & Simko, 2017), devtools (Wickham et al., 2018), grid (R Core Team, 2018),

lme4 (Bates, Mächler et al., 2015), MASS (Venables & Ripley, 2002), MuMIn (Bartoń, 2018), psych (Revelle, 2018), RePsychLing (Baayen et al., 2015), reshape (Wickham, 2007), rmcrr (Bakdash & Marusich, 2018), simr (Green & MacLeod, 2016) und tidyverse (Wickham, 2017).

8.2 Ergebnisse

8.2.1 Manipulationscheck: Effekte der Aufgaben

Das LMM zum Effekt der Aufgaben auf die wahrgenommene Usability zeigte die erwarteten Unterschiede zwischen den Aufgaben (siehe Abb. 8.2 a) und Tab. 8.2). So wurde die wahrgenommene Usability für Aufgabe A im Mittel schlechter bewertet als die für Aufgabe B, Aufgabe B im Mittel geringer bewertet als Aufgabe C und Aufgabe C im Mittel negativer bewertet als Aufgabe D. Die β -Werte aus Tabelle 8.2 können dabei jeweils als Unterschiede zwischen den jeweiligen Stufen interpretiert werden bzw. im Falle des Intercepts als Gesamtmittelwert.

Ein LMM zur Aufgabendauer zeigte ebenfalls Effekte der Aufgaben, die in Abbildung 8.2 b) und Tabelle 8.2 dargestellt sind. Hier zeigte sich, dass Aufgabe B im Mittel die längsten und Aufgabe C die kürzesten Bearbeitungszeiten hatten. Die hier gefundenen Effekte zeigen jedoch kein entgegengesetztes Muster zum Effekt der Aufgaben auf die wahrgenommene Usability. Hier war die am schlechtesten bewertete Aufgabe nicht die längste, wodurch eine systematische Verzerrung der wahrgenommenen Usability durch die Aufgabendauer ausgeschlossen werden kann.

Für die Schätzverzerrungen zeigten sich erneut Überschätzungen ($PTR - Werte > 1$), siehe Abb. 8.2 c). Die Aufgaben wurden jedoch alle ähnlich stark überschätzt. Es gab lediglich einen signifikanten Unterschied zwischen Aufgabe C und D (siehe Tab. 8.2). Hier ist jedoch zu beachten, dass die Varianzaufklärung durch die Fixed Effects (*marginales R^2*) viel geringer ist als durch die Fixed und Random Effects zusammen (*conditionales R^2*), was auf wenig systematische Effekte der Aufgaben, aber starke systematische Effekte der einzelnen Versuchspersonen hindeutet.

Zu den Effekten der Aufgabe wurden ähnlich wie in den vorangegange-

nen Studien Poweranalysen berechnet (siehe Anhang F.3). Besonders interessant sind hier jedoch die Poweranalysen für die Manipulationshypothesen zur wahrgenommenen Usability. Auf Basis der vorliegenden Daten können die Unterschiede zwischen den Aufgaben A und B sowie B und C in der wahrgenommenen Usability als so groß eingeschätzt werden, dass sie mit gleicher Stichprobengröße erneut gefunden werden sollten ($1 - \beta \geq 88\%$). Für den Unterschied zwischen Aufgaben C und D war die Power jedoch geringer ($1 - \beta = 67\%$), da der Unterschied in den vorliegenden Daten kleiner war.

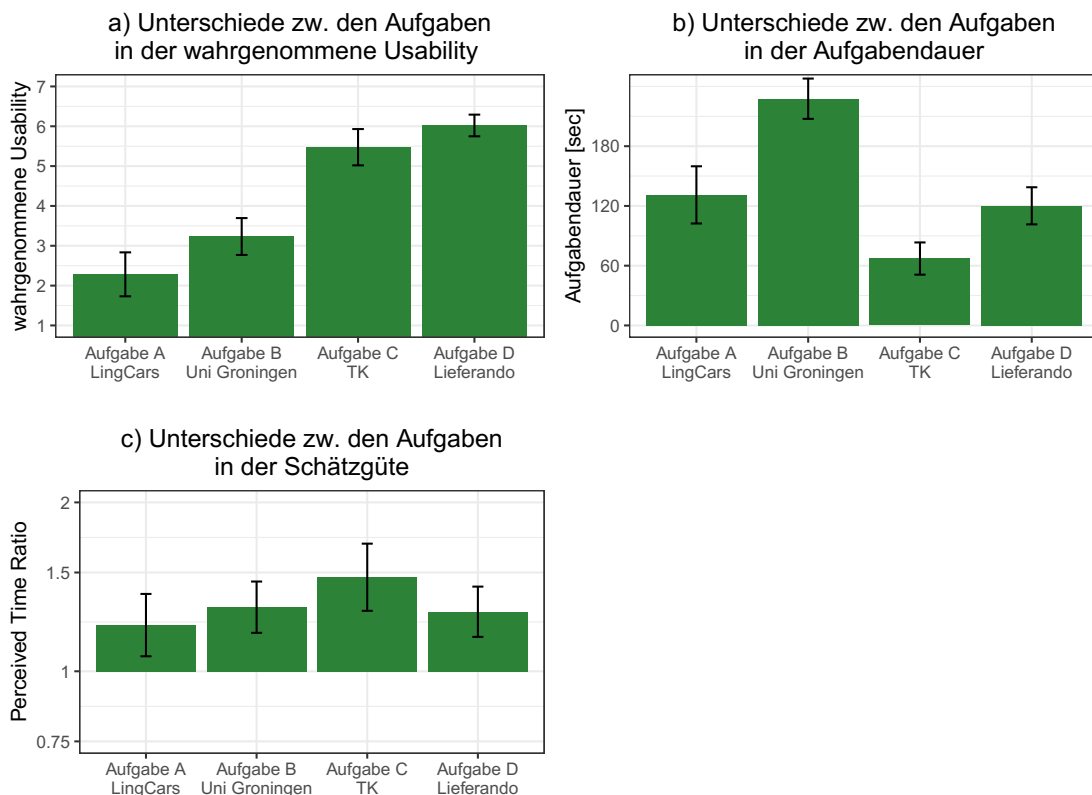


ABBILDUNG 8.2: Effekte der Aufgaben auf die wahrgenommene Usability, die Aufgabendauer und die log-transformierten PTR-Werte. Die Fehlerbalken stellen Konfidenzintervalle der Standardfehler (95 %) korrigiert für Innersubjektfaktoren (Cousineau, 2005; Morey, 2008) dar. Für die PTRs ist die y-Achse rücktransformiert. Ein Wert von 1 entspricht einer perfekten Schätzung.

KAPITEL 8. STUDIE III: KORRELATIONEN ZWISCHEN DAUERWAHRNEHMUNG UND UX BEI „ECHTEN INTERAKTIONEN“

TABELLE 8.2: Effekt der Aufgaben auf die wahrgenommene Usability, die Aufgabendauer und die log-transformierten PTR-Werte.

	β -Gewicht	95% Konf.	Std. Fehler	t -Wert	
Effekte auf wahrgenommene Usability					
Intercept	4.25	[4.08, 4.42]	0.08	50.07	*
Aufgabe A/B	0.95	[0.35, 1.55]	0.30	3.15	*
Aufgabe B/C	2.24	[1.77, 2.72]	0.24	9.33	*
Aufgabe C/D	0.55	[0.07, 1.02]	0.24	2.27	*
Random Effect Struktur: (0 + <i>Aufg. A vs. B</i> <i>Versuchsperson</i>)					
$R^2_{GLMM(m)}$: 67.5 %, $R^2_{GLMM(c)}$: 73.2 %					
Effekte auf Aufgabendauer					
Intercept	136.5	[124.56, 148.49]	5.93	23.03	*
Aufgabe A/B	96.6	[66.44, 126.71]	15.04	6.42	*
Aufgabe B/C	-160.5	[-180.19, -140.82]	9.90	-16.22	*
Aufgabe C/D	53.0	[33.28, 72.64]	9.90	5.35	*
Random Effect Struktur: (1 + <i>Aufg. A vs. B</i> <i>Versuchsperson</i>)					
$R^2_{GLMM(m)}$: 51.6 %, $R^2_{GLMM(c)}$: 75.3 %					
Effekte auf log-transformierte PTR-Werte					
Intercept	0.27	[0.14, 0.4]	0.06	4.19	*
Aufgabe A/B	0.07	[-0.07, 0.21]	0.07	1.04	
Aufgabe B/C	0.12	[-0.02, 0.26]	0.07	1.73	
Aufgabe C/D	-0.14	[-0.28, 0]	0.07	-2.00	*
Random Effect Struktur: (1 <i>Versuchsperson</i>)					
$R^2_{GLMM(m)}$: 2.5 %, $R^2_{GLMM(c)}$: 59.6 %					

ANMERKUNG: Die Sterne geben an, ob der Effekt auf einem Signifikanzniveau von $\alpha = .05$ und zweiseitiger Testung signifikant ist.

Die Bekanntheit der genutzten Webseiten wurde nach jeder Aufgabe geprüft. Hier zeigte sich ein einheitliches Bild für die Webseiten der Aufgaben A, B und C: 94% (A), 94% (B) bzw. 73% (C) der Probanden gaben an, die Webseite nicht zu kennen. Bei Aufgabe D war die Webseite jedoch für nur 40% unbekannt. Da dies auch die Webseite mit der höchsten wahrgenommenen Usability war, wurde post-hoc ein Welch-Test durchgeführt, um zu prüfen, ob die Bekanntheit einen Einfluss auf die wahrgenommene Usability hatte. Es zeigte

sich, dass die Gruppe, die die Webseite kannte, sie im Mittel nicht signifikant anders bewertete ($M = 6.00$, $SD = 0.52$) als die Gruppe, die sie nicht kannte ($M = 6.05$, $SD = 0.76$), $t(19.3) = -0.21$, $p > .05$.

Beim selbstberichteten Erfolg stach Aufgabe B heraus: Während 88% der Probanden bei Aufgabe A, 97% bei Aufgabe C und 100% bei Aufgabe D angaben, die Aufgabe gelöst zu haben, waren es bei Aufgabe B nur 61%. Ein post-hoc t-Test zeigte jedoch auch hier, dass sich die Gruppe mit Erfolg ($M = 3.50$, $SD = 1.16$) nicht signifikant von der Gruppe ohne Erfolg ($M = 2.82$, $SD = 1.18$) in ihrer Usabilitybewertung der Aufgabe B unterschied, $t(31) = -1.63$, $p > .05$. Für die Analyse des objektiven Erfolgs ergab sich das folgende Bild: 82% bei Aufgabe A, 33% bei Aufgabe B, 79% bei Aufgabe C und 58% bei Aufgabe D konnten nach der entsprechenden Aufgabe die richtige Lösung geben. Auch hier zeigten sich keine Unterschiede in der Usabilitybewertung für die Aufgaben mit einer geringeren objektiven Erfolgsrate, Aufgabe B: $t(31) = -1.06$, $p > .05$, Aufgabe D: $t(31) = -0.93$, $p > .05$. Die Hypothese III.MC (i) bis (iii) (β 's > 0) können ohne Einschränkung angenommen werden, da die gefundenen Ergebnisse nicht auf eine Konfundierung der Dauer, der Bekanntheit der Webseite oder des Erfolgs zurückgeführt werden können.

8.2.2 Ergebnisse zu den Teilfragen a) und b): Innersubjektkorrelationen zwischen den zeit- und UX-bezogenen Maßen

In Tabelle 8.3 sind die Innersubjektkorrelationen zu den Hypothesen zusammengefasst. Die Korrelationen zwischen den UX-basierten Maßen sind nicht hypothesenrelevant, können aber im Anhang G.4 eingesehen werden. Eine inhaltliche Beschreibung aller Zusammenhänge erfolgt zusammen mit der Interpretation in der Diskussion, um Doppelungen in der Beschreibung zu vermeiden.

Es zeigten sich wie erwartet positive Zusammenhänge zwischen Aufgabendauer, geschätzter Aufgabendauer, Dauerbewertung und Zeitflussbeurteilung. Auf Basis dieser Ergebnisse wurden die Hypothesen zu den Zusammenhängen zwischen den zeitbasierten Maßen angenommen (Hyp. III.3 (i) bis (iii), III.4

TABELLE 8.3: Innersubjektkorrelationen zwischen den zeit- und den UX-bezogenen Variablen.

Hyp. III.		Aufgaben- dauer	geschätzte Aufg.- dauer	Dauer- bewertung	Zeitfluss- beurteilung
3 (i)-(iii)	Aufg.dauer	–	.84 ***	.82 ***	.49 ***
4 (i)-(ii)	gesch. Aufg.dauer	–	–	.73 ***	.45 ***
13	Dauerbewertung	–	–	–	.54 ***
5 (i)-(iv)	wahrg. Usability	-.48 ***	-.43 ***	-.65 ***	-.36 ***
6 (i)-(iv)	wahrg. Attrakt.	-.13, n.s.	-.10, n.s.	-.28 **	-.19, n.s.
7 (i)-(iv)	Valenz	-.50 ***	-.53 ***	-.52 ***	-.56 ***
8 (i)-(iv)	Arousal	-.27 **	-.22 *	-.21 *	-.32 **

ANMERKUNG: * bei $p < .05$, ** bei $p < .01$, *** bei $p < .001$, fett für $|r| \geq .3$

(i) bis (ii) und III.13: $r's > 0$). In Bezug auf die Zusammenhänge zwischen den zeit- und den UX-basierten Maßen werden die Hypothesen zur wahrgenommenen Usability (Hyp. III.5 (i) bis (iv): $r's < 0$), zur Valenz (Hyp. III.7 (i) bis (iv): $r's < 0$) und zum Arousal (Hyp. III.8 (i) bis (iv): $r's < 0$) angenommen. Für die wahrgenommene Attraktivität wurde nur Evidenz für den Zusammenhang zur Dauerbewertung gefunden (Hyp. III.6 (iii): $r < 0$), nicht aber für die anderen Zusammenhänge (Beibehalten der $H_{0III.6}$ (i), (ii) und (iv): $r's \geq 0$). Eine inhaltliche Interpretation erfolgt in der Zusammenfassung der Ergebnisse in der Diskussion.

8.3 Diskussion

In Studie III lösten Probanden vier Aufgaben auf vier verschiedenen Webseiten. Die zu lösenden Aufgaben bestanden darin, eine bestimmte Information auf der Webseite ausfindig zu machen und waren damit recht ähnlich zueinander. Sie waren jedoch mit Hilfe verschiedener Webseiten zu lösen, welche sich in ihrer wahrgenommenen Usability stark voneinander unterschieden (siehe Manipulationscheck). Mit diesem Versuchsdesign wurde die Kovariation von verschiedenen zeit- und UX-bezogenen Bewertungen untersucht, wenn Probanden sehr verschiedene Interaktionen erleben und diese jeweils bewerten. Darüber

hinaus sollten Verzerrungen in den Zusammenhängen, die durch eine gleichzeitige Messung von ähnlichen Maßen entstehen können (Common-Method-Bias, Podsakoff et al., 2012), für einen Teil der gemessenen Konstrukte (hier wahrgenommene Usability) reduziert werden.

8.3.1 Zusammenfassung und Beantwortung der Forschungs- teilstudien

Innersubjektkorrelationen zwischen den zeitbezogenen Maßen

Studie III zeigte zunächst durch die hohe Korrelation zwischen der tatsächlichen und der geschätzten Aufgabendauer die typischen skalaren Eigenschaften von prospektiven Dauerschätzungen (vgl. STT, Gibbon, 1977). Dieser Zusammenhang stützt im Einklang mit den vorangegangenen Studien die Annahme, dass prospektive Dauerschätzungen im Minutenbereich auch während einer Mensch-Computer-Interaktion auf einem internen Taktgeber beruhen (vgl. Informations-Prozess-Modell der Dauerwahrnehmung nach Treisman, 1963; STT nach Gibbon et al., 1984 und AGM nach z.B. Zakay & Block, 1995).

Für die Korrelationen zwischen den Operationalisierungen der Dauerwahrnehmung zeigten sich überwiegend die erwarteten Zusammenhänge: Je größer die geschätzte Aufgabendauer war, umso schlechter wurde die Dauer als auch der Zeitfluss bewertet. Auch die Dauerbewertung und die Zeitflussbeurteilung zeigten den erwarteten Zusammenhang. So wurden Dauern besser bewertet, wenn die Zeit während der Dauer schnell zu vergehen schien. Diese Befunde decken sich mit der vorangegangenen Studie, weisen aber auf Grund ihrer teilweise nur mittleren Stärke erneut darauf hin, dass die verschiedenen Operationalisierungen nicht deckungsgleich sind. Darüber hinaus stützen sie die Argumentation aus der Diskussion zu Studie II, dass der Zusammenhang zwischen Zeitempfinden und Dauerschätzungen nicht so einfach ist, dass ein Empfinden von langsam vergehender Zeit automatisch zu einer längeren Dauerschätzung führt.

Verglichen mit den Ergebnissen der vorangegangenen Studien waren alle gefundenen Zusammenhänge stärker. Dies legt nahe, dass die größere Variation von Aufgaben wie intendiert zu einem größeren Spektrum an Erfahrungen innerhalb jeder Versuchsperson geführt hat. Dieses Spektrum konnten die Pro-

banden bei der Bewertung der zeitbezogenen Maße nutzen und Vergleiche über die verschiedenen Aufgaben und Systeme hinweg anstellen.

Innersubjektkorrelationen zwischen den zeitbezogenen Maßen und der wahrgenommenen Usability

In Studie III wurde zusätzlich zur größeren Variation der präsentierten Systeme auch die Wahrscheinlichkeit der Entstehung eines Common-Method-Bias für das Konstrukt wahrgenommenen Usability verringert. Dies wurde auf der einen Seite durch eine zeitliche Pause zwischen der Messung der wahrgenommenen Usability und den anderen Konstrukten und auf der anderen Seite durch größere Unterschiede in der Itemgestaltung erzeugt (Podsakoff et al., 2012). Die Unterschiede in der Gestaltung bestanden in der Messung der wahrgenommenen Usability und der visuellen Attraktivität auf einer Likert-Zustimmungsskala anstelle einer bipolaren Analogskala. Darüber hinaus wechselte das Medium der Erfassung der wahrgenommenen Usability von digital zu Papier.

Vor diesem Hintergrund sind die hohen Korrelationen zwischen Dauerwahrnehmung und wahrgenommener Usability besonders stark zu gewichten. So war die geschätzte Aufgabendauer umso länger und die Dauerbewertung und die Zeitflussbeurteilung umso negativer, je schlechter die wahrgenommene Usability bewertet wurde. Diese Zusammenhänge deuten darauf hin, dass Probanden die Konstrukte der Dauerwahrnehmung inhaltlich mit der Bewertung der Usability verknüpfen, auch wenn sie zu unterschiedlichen Zeitpunkten und auf unterschiedliche Art und Weise erfragt werden. Damit liefert Studie III empirische Evidenz für einen starken Zusammenhang zwischen Dauerbewertungen und wahrgenommener Usability, wie er in verschiedenen Fragebögen bereits vermutet und zur Messung von wahrgenommener Usability genutzt wird (vgl. PSSUQ: Lewis, 1992; QUIS: Chin et al., 1988; UEQ: Laugwitz et al., 2006, USE: Lund, 2001).

Innersubjektkorrelationen zwischen den zeitbezogenen Maßen und der wahrgenommenen Attraktivität

Auch die geringen Korrelationen zwischen Dauerwahrnehmung und wahrgenommener Attraktivität sind vor diesem Hintergrund interessant: Trotz gleichzeitiger Messung waren diese Zusammenhänge größtenteils geringer als in vor-

angegangenen Studien. Ein Grund für diesen Unterschied könnte einer der Veränderungen im Versuchsdesign zugeschrieben werden, also (i) der größeren Variation der präsentierten Systeme, (ii) dem anderen Fragebogen für die Attraktivität, (iii) der größeren Unterscheidbarkeit in der Erfassung zwischen wahrgenommener Attraktivität und Zeitempfinden oder (iv) der Unterdrückung des Common-Method-Biases für die wahrgenommene Usability.

Die größere Variation der präsentierten Systeme (i) auf der einen Seite kann dazu geführt haben, dass die verschiedenen Usability-Abstufungen nicht mehr so eindeutig mit einer bestimmten Attraktivität verknüpft waren, wie es beim Arbeiten mit einer konstanten Usability-Abstufungen über alle Aufgaben hinweg der Fall ist. Gegen diesen Erklärungsansatz sprechen jedoch die gefundenen Korrelationen zwischen wahrgenommener Usability und wahrgenommener Attraktivität, die in allen drei Studien ähnlich hoch waren (Studie I mit $r = .63$, Studie II mit $r = .70$ und Studie III mit $r = .75$). Dies zeigt, dass Probanden auch in dieser Studie hohe Usability mit hoher Attraktivität verknüpften.

In Studie III änderte sich jedoch die Messung der wahrgenommenen Attraktivität (ii). In Studien I und II wurde ein Single-Item zur Attraktivität genutzt, wobei der Begriff Attraktivität nicht weiter spezifiziert wurde. Dieses Item kann von Probanden sehr allgemein interpretiert worden sein und eine Art Gesamtbewertung miteingeschlossen haben. In Studie III wurde hingegen ein validierter Fragebogen genutzt, der die Attraktivität bezüglich der visuellen Ästhetik abfragt. Das in Studie III erfasste Konstrukt sollte daher trennschärfer eine rein nicht-instrumentelle Qualität erfassen. Vor diesem Hintergrund kann die geringe Korrelation zwischen wahrgenommener Attraktivität und Dauerwahrnehmung als eine bessere Abbildung der Zusammenhänge zwischen nicht-instrumentellen Qualitäten und zeitbezogenen Maßen angesehen werden.

Die größere Unterscheidbarkeit in der Erfassung (iii) sowie die Unterdrückung des Common-Method-Biases (iv) kann zusätzlich dazu geführt haben, dass Probanden nicht mehr versuchten, konsistent in ihrer Bewertung der Dauerwahrnehmung und der wahrgenommenen Attraktivität zu sein. Eine solche Verzerrung des Antwortverhaltens wird auch Konsistenzmotiv (*Consistency Motif*, Podsakoff et al., 2003) genannt. Sie entsteht, wenn Pro-

banden einen Zusammenhang sowohl der Dauerwahrnehmung als auch der nicht-instrumentellen Qualitäten zu den instrumentellen Qualitäten sehen. Damit beide Zusammenhänge konsistent in den Bewertungen widerspiegelt werden können, erfolgt eine Anpassung der Konstrukte Dauerwahrnehmung und nicht-instrumentelle Qualitäten, wodurch der tatsächliche Zusammenhang überschätzt wird.

Die vorliegenden Ergebnisse stützen die Befunde aus Studien I und II, dass es sowohl reliable Zusammenhänge zwischen Dauerwahrnehmung und wahrgenommener Usability, als auch zwischen wahrgenommener Usability und wahrgenommener Attraktivität gibt (Aranyi & van Schaik, 2015; Lavie & Tractinsky, 2004; Minge, 2008). Die Verringerung der Korrelationen zwischen Dauerwahrnehmung und wahrgenommener Attraktivität gegenüber den ersten beiden Studien stützt zudem die Argumentation, dass Dauerwahrnehmung primär mit der Wahrnehmung instrumenteller Qualitäten und nicht mit der Wahrnehmung nicht-instrumenteller Qualitäten zusammenhängt, was in den vorangegangenen Studien auf Grund der gleichzeitigen und sich ähnelnder Messung nicht dissoziiert werden konnte.

Innersubjektkorrelationen zwischen den zeitbezogenen Maßen und der Emotion

Auch zwischen den Konstrukten der Dauerwahrnehmung und der Valenz zeigten sich starke Korrelationen, was darauf hindeutet, dass Probanden ihren emotionalen Zustand inhaltlich mit der Dauerwahrnehmung verknüpfen. Dies steht im Einklang mit Befunden von anderen Autoren, die einen Zusammenhang zwischen Emotion und Zeitempfinden bzw. Dauerschätzungen bei mehrminütigen Intervallen berichten (Droit-Volet & Wearden, 2015, 2016; Droit-Volet et al., 2017, 2018; Wearden et al., 2014). Auf Grund des Versuchsdesigns kann in der vorliegenden Studie jedoch nicht der Einfluss eines Common-Method-Bias als Grund für den Zusammenhang ausgeschlossen werden, da Valenz und Dauerwahrnehmung gleichzeitig gemessen wurden.

Im CUE-Modell wird angenommen (Mahlke, 2008; Minge & Thüring, 2018; Thüring & Mahlke, 2007), dass die Wahrnehmung der Interaktionscharakteristika (hier Dauer) keinen direkten Einfluss auf die Emotionen hat, sondern über die Wahrnehmung instrumenteller oder nicht-instrumenteller Qualitäten

vermittelt wird. Der Zusammenhang zwischen Dauerwahrnehmung und Valenz kann jedoch als Indiz dafür gesehen werden, dass die Wahrnehmung der Interaktionscharakteristika die Emotionen direkt beeinflusst. Dieses Indiz ist allerdings schwach, da der Zusammenhang auch über die wahrgenommene Usability vermittelt worden sein kann, ohne dass diese zum gleichen Zeitpunkt explizit erfragt wurde.

Das Arousal zeigte in Studie III ähnliche, aber schwächer ausgeprägte Zusammenhänge zu den zeitbezogenen Maßen als die Valenz. So ging ein erhöhtes Arousal-Level mit kürzeren tatsächlichen und geschätzten Dauern, sowie mit einer positiveren Bewertung der Dauer und des Zeitflusses einher. Der stärkste Zusammenhang war ähnlich zu Studie II der zur Zeitflussbeurteilung, wie er auch bereits von Droit-Volet et al. (2018) berichtet wurde.

Beantwortung der Teilfragen a) und b) von Studie III

Zusammengefasst stützen die Ergebnisse die Zusammenhänge zwischen Dauerwahrnehmung und den UX-Komponenten Wahrnehmung instrumenteller Qualitäten und Emotion. Hier scheint vor allem die Dauerbewertung mit den UX-Komponenten verknüpft zu sein, wohingegen der Zeitfluss einen geringeren Zusammenhang aufweist. Darüber hinaus deuten die Ergebnisse darauf hin, dass die Dauerwahrnehmung als Teil der instrumentellen Qualitäten gesehen werden kann.

8.3.2 Limitationen von Studie III

Studie III war eine Fortführung der ersten beiden Studien, in der mehr Varianz in den erlebten Interaktionen erzeugt und eine Beeinflussung der Zusammenhänge mit dem Konstrukt wahrgenommene Usability durch den Common-Method-Bias reduziert wurde. Beide Veränderungen im Versuchsdesign gehen jedoch mit Limitationen für die Interpretation einher.

Zum einen stellten die verschiedenen Webseiten mit ihren Aufgaben nur eine kleine Auswahl dar. Hier hätten andere oder mehr Aufgaben genutzt werden können, die gegebenenfalls zu einem anderen Ergebnis geführt hätten. Darüber hinaus stellten auch sie nur einen Ausschnitt auf einem Kontinuum von möglichst schlechter zu möglichst guter Usability dar. Hier spiegelt sich

jedoch ein Problem bei der Vorauswahl verschiedener Usabilitystufen wieder: Wahrgenommene Usability entsteht in der Interaktion mit einem Probanden, da ein wesentlicher Teil von Usability die Zufriedenheit des Nutzers ist (vgl. Definition Usability, DIN, EN ISO 9241-11, 2018). Daher kann sie nicht präzise vorhergesagt oder so gestaltet werden, dass die gleiche wahrgenommene Usability für jeden Nutzer erzeugt wird. Zum anderen wurde der Common-Method-Bias nur für ein Konstrukt reduziert, die wahrgenommene Usability. Diese Auswahl erfolgte auf Basis der Vorarbeiten und Vorüberlegungen, hätte aber auch für andere Konstrukte angewendet werden können. So bleibt die Frage offen, in wie weit die zeitbasierten Variablen miteinander korrelieren, wenn ihre Messungen getrennt voneinander erfolgen.

Ein weiterer Kritikpunkt kann in der Wahl der gemessenen Konstrukte gesehen werden. Auf Basis der schwachen Korrelationen zwischen wahrgenommener Attraktivität und Dauerwahrnehmung wurde argumentiert, dass die Ergebnisse den Schluss stützen, dass Dauerwahrnehmung primär mit der Wahrnehmung instrumenteller Qualitäten und nicht mit der Wahrnehmung nicht-instrumenteller Qualitäten zusammenhängt. Dies kann jedoch als Übergeneralisierung gesehen werden, da sowohl instrumentelle Qualitäten als auch nicht-instrumentelle Qualitäten durch jeweils nur ein Konstrukt gemessen wurden. Im CUE-Modell werden weitere Konstrukte unter diesen Überkategorien genannt, wie zum Beispiel die wahrgenommene Nützlichkeit, der Status oder die Identifikation mit dem Produkt (Mahlke, 2008; Minge & Thüring, 2018; Thüring & Mahlke, 2007). Hier fehlen weitere Studien, um die Generalisierbarkeit der gemachten Aussagen zu prüfen.

8.3.3 Fazit Studie III

Studie III zeigte, dass Dauerbewertung und wahrgenommene Usability bzw. Valenz bei einer großen Varianz der betrachteten Aufgaben bzw. Webseiten stark miteinander kovariieren. Dies deutet darauf hin, dass Probanden in der Bewertung dieser drei Konstrukte eine große inhaltliche Nähe sehen. Die Zeitflussbeurteilung hingegen zeigte außer zur Valenz schwächere Zusammenhänge. Dies weist darauf hin, wie wichtig eine Diskriminierung zwischen verschiedenen Operationalisierungen von Zeitempfinden ist.

Kapitel 9

Studie IV: Effekte von System und Dauer auf Dauerbewertung und Zeitflussbeurteilung

In der Diskussion zu Studien I und II wurde zu Forschungsfrage 1 die folgende Teilfrage aufgeworfen:

- a) Wird die Dauerbewertung und die Zeitflussbeurteilung primär von der Dauerschätzung oder vom Nutzererleben beeinflusst?

Die theoretische Modellannahme zu dieser Frage ist in Abbildung 9.1 in einem Arbeitsmodell dargestellt. In Anlehnung an das CUE-Modell (Mahlke, 2008; Minge & Thüring, 2018; Thüring & Mahlke, 2007) ist hier visualisiert, dass verschiedene Interaktionscharakteristika auf Basis des Nutzers, des Systems und des Kontextes entstehen. Ein solches Charakteristikum ist die Dauer der Interaktion. Diese kann vom Nutzer wahrgenommen und geschätzt werden, wobei es durch beispielsweise mentale Beanspruchung während der Interaktion zu Verzerrungen kommen kann (vgl. Effekt von lautem Denken auf PTRs in Studien I und II oder bei Hertzum & Holmegaard, 2015). Basierend auf der Wahrnehmung der Interaktionscharakteristika folgt auch das Nutzererleben, welches aus der Wahrnehmung von instrumentellen und nicht-instrumentellen Qualitäten sowie der erlebten Emotion besteht.

Im Arbeitsmodell wird angenommen, dass die Dauerbewertung und die Zeitflussbeurteilung auf der einen Seite davon beeinflusst werden können, wie lange

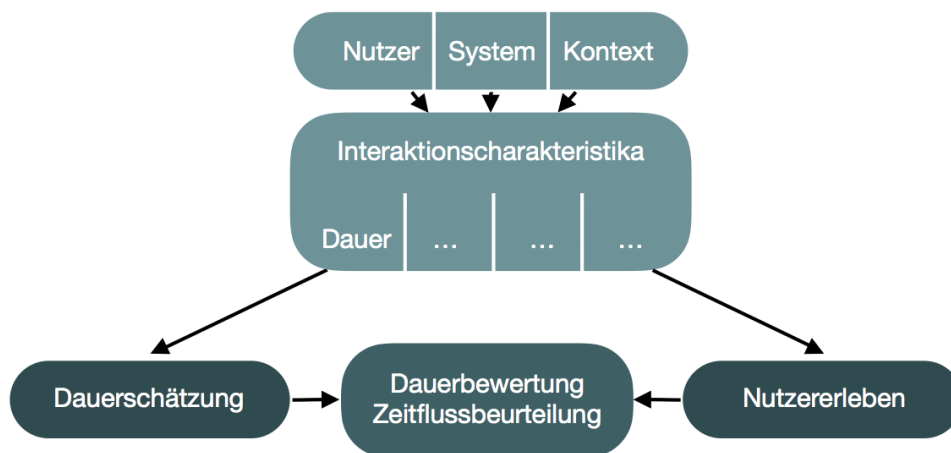


ABBILDUNG 9.1: Annahmen zu den Einflüssen von Dauerschätzung und Nutzererleben auf das Zeitempfinden.

die Dauer der Interaktion erlebt wurde (Dauerschätzung). Evidenz für diese Annahme lieferten die Ergebnisse der Studien I, II und III, in denen Korrelationen zwischen Dauerschätzung und Dauerbewertung bzw. Zeitflussbeurteilung innerhalb der Subjekte gezeigt wurden. Zusätzlich wurden diese drei Konstrukte in den Studien I und II auf die gleiche Art und Weise von der experimentell manipulierten Usability beeinflusst. Auf der anderen Seite könnten Dauerbewertung und Zeitflussbeurteilung auch durch das Nutzererleben geprägt sein. Auch hier gibt es Evidenz durch die entsprechenden Korrelationen und die ähnliche Beeinflussung durch Usability.

Die beiden Pfade konnten bis jetzt auf Grund einer Konfundierung zwischen der tatsächlichen Dauer und der Usability-Stufe in Studien I und II nicht voneinander dissoziiert werden. In den beiden Studien waren die Interaktionsdauern bei niedriger Usability länger als bei hoher Usability. Dies führte zu einem Unterschied in der Dauerschätzung zwischen den beiden Usability-Gruppen. Grundsätzlich kann aber davon ausgegangen werden, dass dieser Unterschied alleine durch die Dauern entstanden ist und nicht durch die Usability-Manipulation, da die Verzerrung der Schätzung bezüglich Über- oder Unterschätzungen in Studien I und II nicht von der Usability beeinflusst wurde.

In Studie IV wird der Einfluss von Usability vom Effekt der Dauer auf das Zeitempfinden experimentell getrennt, um die beiden Wirkungspfade aus Abbildung 9.1 gegeneinander abzuwägen. Dazu werden in einem zweifaktoriellen Design die Faktoren Usability und Dauer systematisch variiert bzw. kon-

trolliert. Sollte die Dauerschätzung die primäre Quelle für das Entstehen von Dauerbewertungen und Zeitflussbeurteilungen sein, so sollten diese zwei Konstrukte von der Dauer der Interaktion und nicht von der Usability beeinflusst werden. Sollte hingegen das Nutzererleben als Grundlage für Dauerbewertungen und Zeitflussbeurteilungen dienen, so müsste sich der Effekt der Usability auf das Nutzererleben auch bei Dauerbewertung und Zeitflussbeurteilung widerspiegeln.

In Studie IV wird ein ähnliches Design wie in den vorangegangenen Studien gewählt: Probanden interagieren mit einem technischen System, welches in zwei Versionen mit unterschiedlicher Usability vorliegt. Um die Aufgabendauer in Studie IV anders als in den vorangegangenen Studien kontrollieren zu können, muss die Aufgabe mit dem technischen System jedoch so gestaltet sein, dass sie kein natürliches Ende hat. Daher soll als Versuchsmaterial keine Webseite, sondern die Simulation eines Systems genutzt werden, welches kontinuierlich überwacht und reguliert werden muss. Dieses experimentelle Design erlaubt auf der einen Seite einen Vergleich mit Studien I und II und eine Übertragung der Ergebnisse in einen neuen Kontext.

Abschließend werden im Rahmen von Forschungsfrage 2 die Korrelationen zwischen den Maßen der Dauerwahrnehmung und zwischen der Dauerwahrnehmung und dem Nutzererleben in Studie IV erneut betrachtet werden, um die gefundenen Zusammenhänge im Rahmen eines anderen Systems zu validieren und damit weiter verallgemeinern zu können. Teilfrage b) lautet daher:

- b) Wie hängen die Operationalisierungen der Dauerwahrnehmung untereinander und mit den Komponenten des Nutzererlebens bei konstanten Dauern zusammen?

9.1 Methode

Studie IV wurde im Rahmen zweier Masterarbeiten (Balashova, 2018; Ullrich, 2018) getestet. Die Idee zur Studie und die experimentelle Planung der beiden Masterarbeiten stammen von der Verfasserin der vorliegenden Arbeit. Im Rahmen der Masterarbeiten haben die beiden Studentinnen die genutzte Simulationsumgebung entworfen, programmiert und vorgetestet. Zudem haben sie die Fragebögen für die Studie aufbereitet und die Daten erhoben. Alle hier

präsentierten Analysen und Beschreibungen wurden von der Verfasserin der vorliegenden Arbeit ausgeführt und beschrieben.

9.1.1 Teilnehmer

An Studie IV nahmen $N = 60$ Teilnehmer teil. 20 von ihnen waren männlich, 35 weiblich und 5 machten dazu keine Angaben. Die Probanden waren im Mittel $M = 27.28$ Jahre ($SD = 4.22$) alt. Die Probanden konnten Versuchspersonenstunden für die Teilnahme erhalten. Die Studie wurde durch die Ethikkommission des Instituts für Psychologie und Arbeitswissenschaften an der TU Berlin begutachtet (Stellungnahme der Ethikkommission siehe Anhang G.2).

9.1.2 Simulationsumgebung

Für Studie IV musste eine Prüfaufgabe gefunden werden, bei der die Gebrauchstauglichkeit des genutzten Systems unabhängig von der Aufgabendauer variiert werden konnte. Dazu wurde eine Simulation eines Heizungssystems gewählt, welches von Balashova (2018) programmiert wurde. Diese Simulation stellt ein System bestehend aus einem Heizkessel und zwei zu heizenden Räumen dar. Die Aufgabe des Probanden bestand darin, die Temperatur der beiden Räume zwischen 21 und 25 Grad Celsius zu halten und zu verhindern, dass sich der Heizkessel auf Grund zu hohen Drucks ausschaltete. Dazu konnten sie in den Räumen die Heizzufuhr einstellen und ein Fenster öffnen oder schließen. Der Druck des Heizkessels erhöhte sich beim Heizen automatisch und konnte von den Probanden über ein Ausgleichsventil heruntergesetzt werden. Wie schnell die Temperatur in den Räumen und der Druck des Heizkessels anstieg, hing nicht nur von den Einstellungen des Systems, sondern auch von der fiktiven und simulierten Außentemperatur ab. Diese änderte sich während der Interaktion, was in einem Informationsfenster dargestellt wurde.

Das Interface der Simulation lag in zwei Versionen mit verschiedenen Usability-Stufen (hoch und niedrig) vor (siehe Abb. 9.2). In der Version mit niedriger Usability waren die nötigen Handlungen zur Regulierung des Systems umständlicher als bei der anderen Version. So mussten Eingaben, wie das Ändern der Stufe des Heizkörpers, nur bei niedriger Usability zusätzlich

in einem Pop-Up-Fenster bestätigt werden. Darüber hinaus waren die Wettervorhersagen bei hoher Usability leichter verständlich und die Zustände der Heizung (an- vs. ausgeschaltet) oder des Fensters (offen vs. geschlossen) ließen sich bei niedriger Usability weniger gut einsehen. Zusätzlich war die Lesbarkeit bei niedriger Usability auf Grund geringerer Kontraste eingeschränkt.

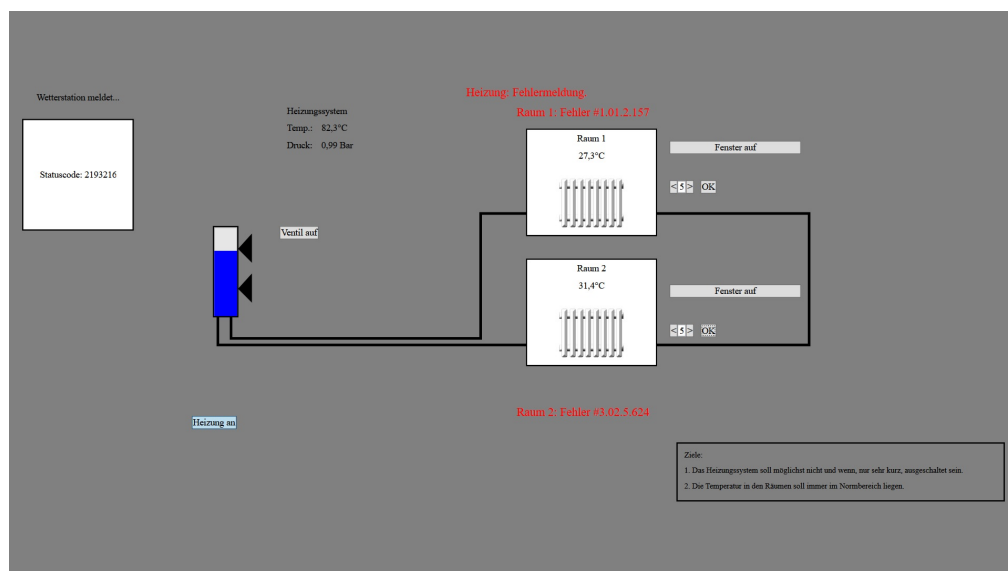
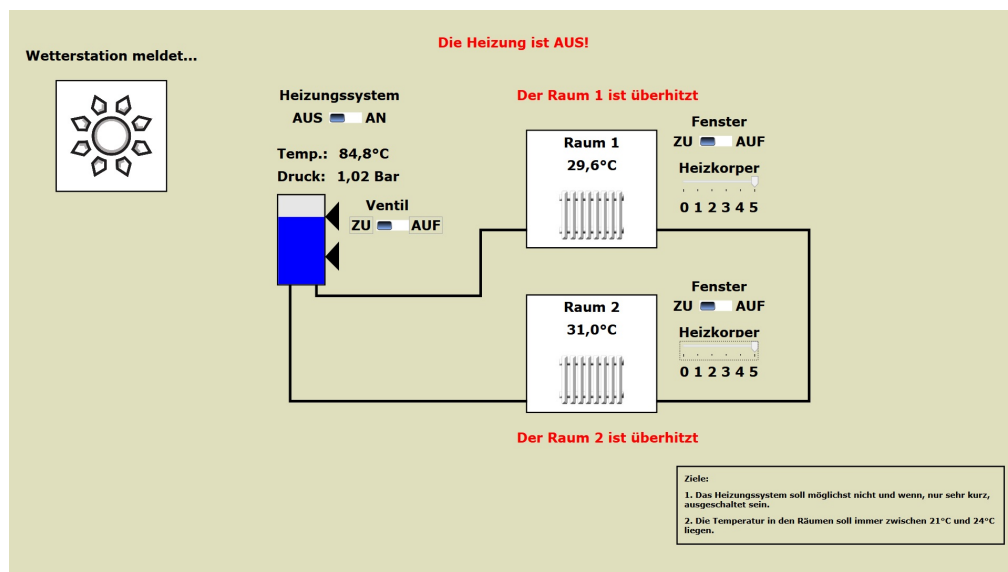


ABBILDUNG 9.2: Interface des Heizungssystems in Studie IV. Oben mit hoher Usability, unten mit niedriger Usability.

9.1.3 Studiendesign

Studie IV basierte auf einem zwei-faktoriellen, gemischten Design mit dem Zwischensubjektfaktor *Usability* (niedrig vs. hoch) und dem Innersubjektfaktor *Dauer* (3 vs. 6 Min.). Die Zuordnung der Probanden zu den beiden Usability-Gruppen erfolgte randomisiert. Die Reihenfolge der Aufgabendauern war für alle Probanden gleich, da dies für den zweiten Teil der Studie wichtig war, der hier nicht näher beschrieben wird. Alle externen Zeitgeber (z.B. Wand- oder Armbanduhren und Handys) wurden entfernt.

9.1.4 Versuchsablauf und Material

Das Experiment begann mit einer allgemeinen Instruktion und der Erhebung der Demographie. Danach erhielten die Probanden eine schriftliche Instruktion zu der Simulation und die Möglichkeit, die Simulation in einer viereinhalb-minütigen Übungsphase kennenzulernen. In der Übung interagierten sie mit der Version entsprechend ihrer Usability-Gruppe. Danach wurden die Probanden gebeten, die Simulation zu bewerten (u.a. bezüglich Zeitschätzung, Usability, Emotion und mentale Beanspruchung). Diese Messung diente lediglich dazu, die Probanden mit dem Ablauf und den Erhebungsinstrumenten vertraut zu machen. Die Daten werden daher nicht berichtet.

Danach folgten zwei Durchgänge, in denen die Probanden mit der Simulation arbeiteten, wobei der erste Durchgang drei und der zweite sechs Minuten dauerte. Nach jedem Durchgang füllten die Probanden einen Fragebogen aus, welcher die Grundlage für die vorgestellte Analyse lieferte. Danach begann der zweite Teil des Experiments, welches für die hier vorgestellte Fragestellung keine Relevanz hat, aber Teil einer der Masterarbeiten war (Balashova, 2018). Insgesamt dauerte das Experiment ca. 50 Minuten pro Proband. Der Versuchsablauf von Studie IV ist in Abbildung 9.3 visualisiert.

Die Erhebung wurde an einem Computer-Arbeitsplatz mit zwei Bildschirmen durchgeführt. Auf dem linken wurde die Simulation gezeigt und auf dem rechten der Fragebogen dargeboten. Dieser wurde mit Hilfe eines Online-Tools SoSciSurvey (<http://soscisurvey.de>) dargestellt und gespeichert. Die Einverständniserklärung lag in Papierform vor.

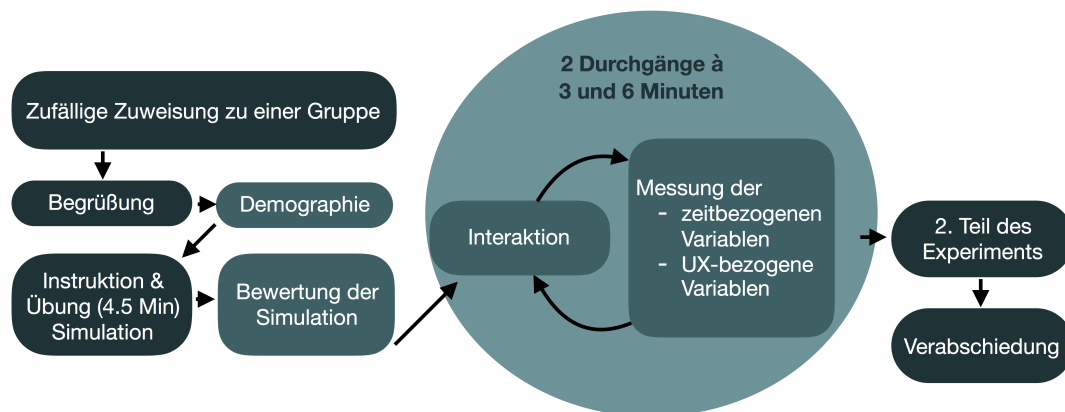


ABBILDUNG 9.3: Übersicht über den Versuchsablauf in Studie IV.

9.1.5 Operationalisierung der abhängigen Variablen

Die zeitbezogenen Variablen umfassten erneut die geschätzte Aufgabendauer, die Dauerbewertung und die Zeitflussbeurteilung. Wie in Studie III wurde die Dauerbewertung mit den Items wahrgenommene Länge der Dauer und wahrgenommene Inakzeptabilität der Dauer abgebildet. Die Zeitflussbeurteilung bestand aus den Items wahrgenommene Schnelligkeit der Dauer und wahrgenommenes Ziehen der Zeit. Diese vier Items wurden jedoch im Vergleich zu Studie III in ihrer Skalierung verfeinert. Während sie in den vorangegangenen Studien auf einer Skala mit 11 Abstufungen gemessen wurden, wurden sie nun in Anlehnung an die Operationalisierung der Passage of Time Judgements von Tanaka und Yotsumoto (2017) auf einer Skala mit 101 Abstufungen von 1 bis 101 gemessen. Dadurch können Probanden die einzelnen Stufen nicht mehr visuell voneinander diskriminieren und gleichzeitig gemessene Konstrukte schwieriger auf die exakt gleiche Stufe gesetzt werden. Die Item können im Anhang G.1.2 eingesehen werden.

Die UX-bezogenen Variablen wurden erneut mit den Kurz-Items aus Studien I und II gemessen. Für diese wurde ebenfalls die Skala weiter verfeinert und zu einer Skalierung mit 101 Abstufungen geändert. Zusätzlich wurden die Single-Items für die wahrgenommene Usability und die wahrgenommene Attraktivität um je ein Item erweitert. So wurde das Item wahrgenommene Gebrauchstauglichkeit um ein zweites, bipolares Item (*Bedienbarkeit*) erweitert, welches von (1) schwierig zu bedienen zu (101) einfach zu bedienen

reichte¹. Das Item wahrgenommene Attraktivität wurde um das bipolare Item *Schönheit* an Anlehnung an Minge (2011) erweitert, welches von (1) hässlich zu (101) schön reichte und einen klareren Bezug zur visuellen Ästhetik der Simulation herstellen sollte². Die Items wurden jeweils gemittelt, um eine zuverlässigere Schätzung der wahrgenommenen Usability und der wahrgenommenen Attraktivität zu gewährleisten. Sie sind in Anhang G.1.2 dargestellt.

Die Validität der gemittelten UX-Kurzitems über die jeweils zwei Items wurde mit Hilfe des VisAWi-Short (Thielsch & Moshagen, 2011; Moshagen & Thielsch, 2013) und dem Modul 1 des meCUEs (Minge et al., 2016) geprüft, die beide nach der Übung gemessen wurden. Die gemittelten Kurzitems zur wahrgenommenen Usability gemessen nach der ersten Interaktion zeigten hohe Korrelationen mit der Skala wahrgenommene Usability des Modul 1 des meCUEs nach der Übung, $r = .64$, $p < .001$. Ähnlich hoch war die Korrelation zwischen der gemittelten wahrgenommenen Attraktivität und der Skala Attraktivität des AttrakDiffs, $r = .66$, $p < .001$.

9.1.6 Hypothesen und Datenanalyse

Hypothesen zur Replikation: Effekte der UVn auf die Dauerschätzung und die Komponenten des Nutzererlebens

Basierend auf Studien I und II wird erwartet, dass der Faktor Usability einen Effekt auf das Nutzererleben hat, nicht aber zu Unterschieden in der geschätzten Dauer führt. Die Manipulation der Dauer unabhängig von der Usability hingegen sollte einen Effekt auf die geschätzte Dauer zeigen, nicht aber auf das Nutzererleben. Diese Effekte erneut zu finden, würde der Replikation einer Teilmenge an Ergebnissen der ersten beiden Studien entsprechen und ist eine Voraussetzung für das Abgrenzen der Wirkung der Dauerschätzung und des Nutzererlebens auf die Dauerbewertung und die Zeitflussbeurteilung.

Ein Einfluss der Usability-Manipulation auf die UX-Komponenten wahrgenommene Usability, wahrgenommene Attraktivität und Valenz wurde bereits in Studien I und II gezeigt und kann im Rahmen der Hypothesen IV.9 (i)

¹Das Item Bedienbarkeit wurde zum besseren Vergleich für die Beschreibung und Analyse gegenüber der Erhebung invertiert.

²Das Item Schönheit wurde zum besseren Vergleich für die Beschreibung und Analyse gegenüber der Erhebung invertiert.

bis (iii) getestet werden. Diese Hypothese wird daher aus den vorangegangenen Studien übernommen. Ein Effekt der Dauer auf die Dauerschätzung wurde bis jetzt nur korrelativ betrachtet, kann aber auch als Unterschiedshypothese formuliert werden:

Hypothese IV.14: *Die **Dauerschätzung** ist signifikant kleiner in den Durchgängen mit kürzeren objektiven Dauern als in den Durchgängen mit längeren objektiven Dauern. Dies zeigt sich in positiven Steigungen: $\beta_{Dauer} > 0$.*

Bezüglich des Einflusses der Usability auf die Dauerschätzung und des Einflusses der Dauer auf das Nutzererleben werden keine signifikanten Effekte erwartet. Nullhypothesen können grundsätzlich nicht mit Hilfe der klassischen Inferenzstatistik getestet werden (Bortz, 2005). Das Ziel der vorliegenden Studie ist es jedoch nicht zu zeigen, dass die betroffenen Effekte nicht existieren, sondern dass die betroffenen Effekte in der vorliegenden Studie keine relevante Stärke haben. Daher wird für die folgende Analyse angenommen, dass der Einfluss der Usability auf die Dauerschätzung und der Einfluss der Dauer auf das Nutzererleben vernachlässigt werden kann, wenn sie keine Signifikanz erreichen.

Zusätzlich wird neben den hier formulierten Haupteffekten der beiden unabhängigen Variablen auch immer die Interaktion explorativ betrachtet. Dazu werden LMMs der folgenden Form mit Maximum Likelihood Schätzung (ML) aufgestellt:

$$\begin{aligned} \text{abhängige Variable} &\sim 1 + \text{Dauer} + \text{Usabilitygruppe} + \\ &\quad \text{Dauer} : \text{Usabilitygruppe} + (1|\text{Versuchsperson}) \end{aligned}$$

Hypothesen zur Teilfrage a): Abwägen der Erklärungsansätze durch Mediatoranalysen

Anschließend wird untersucht, ob primär die Wahrnehmung der Aufgabendauer oder die Wahrnehmung des Nutzererlebens die Bewertung der Dauer und die Beurteilung des Zeitflusses bestimmt, siehe Teilfrage a). Dazu werden Mediatoranalysen nach Baron und Kenny (1986) durchgeführt. Baron und Kenny (1986) argumentieren, dass für eine vollständige Mediation drei Bedingungen

erfüllt sein müssen. Am Beispiel des Zeitempfinden können diese wie folgt zusammengefasst werden:

1. Die unabhängige Variable ist ein signifikanter Prädiktor für den potentiellen Mediator (Pfad A in Abb. 9.4).
2. Der mögliche Mediator ist ein signifikanter Prädiktor für das Zeitempfinden (Pfad B).
3. Der Effekt der unabhängigen Variable auf das Zeitempfinden (Pfad C) verschwindet, wenn für Pfade A und B in einem gemeinsamen Modell kontrolliert wird. Reduziert sich Pfad C lediglich, kann nicht von einer vollständigen, sondern nur von einer partiellen Mediation ausgegangen werden.

Um dies zu testen, muss zunächst in zwei verschiedenen Modellen geprüft werden, ob es Evidenz für die Pfade A und C gibt. Pfad A soll bereits im Rahmen der Hypothesen IV.9 (i) bis (iii) sowie IV.14 zur Replikation betrachtet werden (siehe vorangegangener Abschnitt). Daher wird Pfad C für die Konstrukte Dauerbewertung und Zeitflussbeurteilung jeweils mit der gleichen Analysestrategie wie im vorangegangenen Abschnitt überprüft. Für diejenigen Kombinationen aus unabhängiger Variable und möglichem Prädiktor, für die

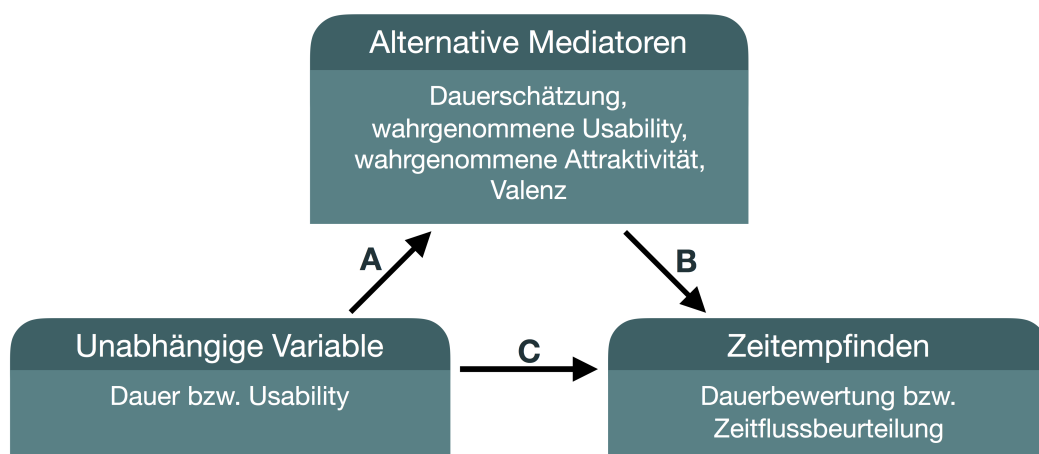


ABBILDUNG 9.4: Veranschaulichung der Mediatoranalyse nach Baron und Kenny (1986) in Studie IV. Die alternativen Mediatoren entsprechen den verschiedenen Pfaden der Beeinflussung aus Abb. 9.1.

sowohl Evidenz für Pfad A als auch für Pfad C gefunden wird, kann dann die Mediation in einem dritten Modell der folgenden Form geprüft werden:

$$\text{Zeitempfinden} \sim 1 + \text{unabhängige Variable} + \text{möglicher} \\ \text{Prädiktor} + (1 \mid \text{Versuchsperson})$$

Mit diesem Modell werden Pfad B sowie die Reduktion des Effekts von Pfad C entsprechend der Punkte 2. und 3. getestet.

Hypothesen zur Teilfrage b): Korrelationen

Ähnlich wie in den vorangegangenen Studien sollen die Korrelationen zwischen den Maßen der Dauerwahrnehmung und zwischen der Dauerwahrnehmung und dem Nutzererleben erneut betrachtet werden. Dazu werden die entsprechenden Korrelationshypothesen aus den vorangegangenen Studien übernommen (für eine Übersicht siehe Tab. 7.3, S.97). Diese betreffen auf der einen Seite die Zusammenhänge zwischen den Operationalisierungen der Dauerbewertung (Hyp. IV.4 (i) und (ii) sowie IV.13) und auf der anderen Seite die Korrelationen zwischen Dauerbewertung und UX-Komponenten (Hyp. IV.5 bis 8 jeweils (ii) bis (iv)). Die Zusammenhänge zur objektiven Dauer werden in diesem Rahmen nicht betrachtet, da diese als unabhängige Variable variiert wurde.

Wie in den vorausgegangenen Studien werden Innersubjektkorrelationen berechnet. Da im vorliegenden Versuchsdesign die Unterschiede zwischen den Messzeitpunkten durch die identische Aufgabe mit demselben System jedoch sehr gering zu erwarten sind und dadurch die Zusammenhänge innerhalb der Probanden sehr gering sein können, werden zusätzlich klassische Korrelationen präsentiert. Hierfür werden über die Messwiederholung hinweg gemittelt. Zusammenhänge zur geschätzten Aufgabendauer werden hier allerdings nicht betrachtet, da sich im Bezug auf die Dauer die beiden Messzeitpunkte zu stark unterscheiden (3 min vs. 6 min).

Genutzte Software

Zur Analyse der Daten wurde das Statistik-Programm R (R Core Team, 2018) und die folgenden Pakete genutzt: car (Fox & Weisberg, 2011), corpcor (Schafer et al., 2017), corrplot (Wei & Simko, 2017), devtools (Wickham et al., 2018), grid (R Core Team, 2018), lattice (Sarkar, 2008), lme4 (Bates,

Mächler et al., 2015), MASS (Venables & Ripley, 2002), MuMIn (Bartoń, 2018), plyr (Wickham, 2011), psych (Revelle, 2018), RePsychLing (Baayen et al., 2015), reshape2 (Wickham, 2007), rmcrr (Bakdash & Marusich, 2018), simr (Green & MacLeod, 2016), tidyverse (Wickham, 2017) und WSR2 (Mair & Wilcox, 2018).

9.2 Ergebnisse

9.2.1 Hypothesen zur Replikation: Effekte der UVn auf die Dauerschätzung und die Komponenten des Nutzererlebens

Die Ergebnisse der Analysen zum Effekt der unabhängigen Variablen Dauer und Usability auf die Dauerschätzung und die Komponenten des Nutzererlebens sind in Tabelle 9.1 zusammengefasst. Diese Analysen prüfen Pfad A der Mediatoranalysen (Abb. 9.4). Für die Dauerschätzung zeigte sich der erwartete Effekt der Dauer³: mit jeder zusätzlichen Minute schätzten die Probanden die Dauer im Mittel um $M = 57.4$ Sek. ($SD = 32.9$ Sek.) länger. Einen Haupteffekt der Usability gab es wie erwartet nicht. Die Interaktion erreichte ebenfalls Signifikanz. Der Unterschied in der Dauerschätzung war für Probanden in der Gruppe mit niedriger Usability zwischen den Dauern größer, siehe Abbildung 9.5 a). Daher muss bei der Annahme der Hypothese IV.14 ($\beta_{Dauer} > 0$) beachtet werden, dass der sehr starke Effekt der Dauer auf die geschätzte Dauer durch die Usabilitygruppe moderiert wurde.

Für die wahrgenommene Usability zeigte sich wie erwartet der Effekt der Usability-Gruppe und kein Effekt der Dauer oder der Interaktion⁴. Der Unterschied zwischen den Usability-Gruppen ist in Abbildung 9.5 b) dargestellt und

³Für die Dauerschätzung wurden vor der Analyse zwei Werte durch not availables (NAs) ersetzt, da hier Probanden eine Dauer von drei Minuten auf 99 Min. 59 Sek. schätzten. Diese Werte waren so weit von der tatsächlichen Dauer entfernt, dass sie nicht als valide angesehen wurden.

⁴Für diese Analyse lag keine Normalverteilung der Residuen vor. Diese konnte auch nicht durch eine Transformation der abhängigen Variable erreicht werden. Ein nicht-parametrisches Testverfahren (`bwtrim`) aus dem *WSR2*-Paket (Mair & Wilcox, 2018) für *R* zeigte jedoch die gleichen Effekte wie die hier berichtete Analyse.

TABELLE 9.1: Effekte der unabhängigen Variablen auf transformierte Dauerschätzung, wahrgenommene Usability, wahrgenommene Attraktivität und Valenz. Analysen entsprechen der Prüfung des Pfads A der Mediatoranalyse.

	β -Gewicht	95% Konf.	Std. Fehler	t -Wert	
Effekte auf log-transformierte Dauerschätzung					
Intercept	5.44	[5.35, 5.52]	0.04	124.90	*
Dauer [min]	0.23	[0.2, 0.26]	0.01	15.31	*
Usab. hoch/niedrig	-0.13	[-0.3, 0.05]	0.09	-1.44	
Dauer:Usab.	0.09	[0.03, 0.15]	0.03	2.84	*
Random Effect Struktur: (1 <i>Versuchsperson</i>)					
$R^2_{GLMM(m)}$: 47.2 %, $R^2_{GLMM(c)}$: 78.1 %					
Effekte auf wahrgenommene Usability					
Intercept	56.90	[52.04, 61.76]	2.44	23.34	*
Dauer [Min.]	-0.23	[-1.7, 1.25]	0.74	-0.31	
Usab. hoch/niedrig	-13.83	[-23.54, -4.12]	4.88	-2.84	*
Dauer:Usab.	0.30	[-2.65, 3.25]	1.48	0.20	
Random Effect Struktur: (1 <i>Versuchsperson</i>)					
$R^2_{GLMM(m)}$: 10.1 %, $R^2_{GLMM(c)}$: 69.1 %					
Effekte auf wahrgenommene Attraktivität					
Intercept	42.11	[38.2, 46.01]	1.96	21.47	*
Dauer [Min.]	-0.95	[-1.98, 0.08]	0.52	-1.84	
Usab. hoch/niedrig	-6.83	[-14.65, 0.98]	3.92	-1.74	
Dauer:Usab.	0.80	[-1.26, 2.86]	1.03	0.77	
Random Effect Struktur: (1 <i>Versuchsperson</i>)					
$R^2_{GLMM(m)}$: 5 %, $R^2_{GLMM(c)}$: 74.3 %					
Effekte auf Valenz					
Intercept	5.62	[5.26, 5.99]	0.19	30.33	*
Dauer [Min.]	-0.15	[-0.28, -0.02]	0.06	-2.31	*
Usab. hoch/niedrig	-1.18	[-1.92, -0.44]	0.37	-3.19	*
Dauer:Usab.	-0.06	[-0.31, 0.2]	0.13	-0.43	
Random Effect Struktur: (1 <i>Versuchsperson</i>)					
$R^2_{GLMM(m)}$: 13.4 %, $R^2_{GLMM(c)}$: 62.6 %					

ANMERKUNG: Die Sterne geben an, ob der Effekt auf einem Signifikanzniveau von $\alpha = .05$ und zweiseitiger Testung signifikant ist.

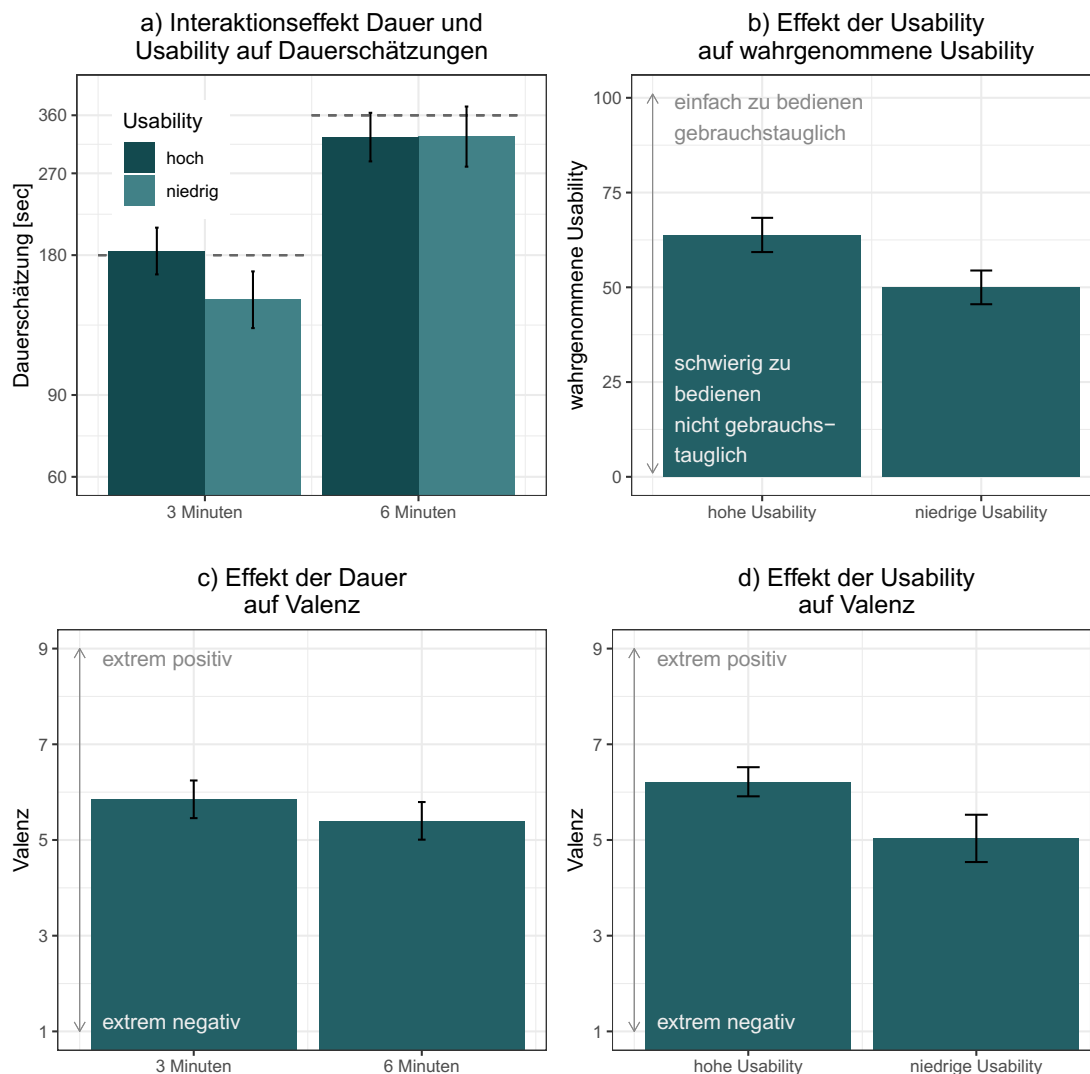


ABBILDUNG 9.5: Effekte der unabhängigen Variablen auf log-transformierte geschätzte Aufgabendauer und Komponenten des Nutzererlebens. Die Fehlerbalken stellen Konfidenzintervalle der Standardfehler (95 %) korrigiert für Innersubjektfaktoren (Cousineau, 2005; Morey, 2008) dar. Die gestrichelten Linien in a) zeigen die tatsächliche Dauer an. Zur besseren Interpretation wurde die Beschriftung der y-Achse in a) rücktransformiert.

zeigt die erwartete Richtung (siehe Tab. 9.1). Daher wird Hypothese IV.9 (i) ($\beta_{Usability} < 0$) zum Effekt der Usability angenommen.

Diese klare Beeinflussung konnte jedoch für die anderen Komponenten des Nutzererlebens nicht bestätigt werden. Für die wahrgenommene Attraktivität wurde weder ein Effekt der Dauer, noch der Usability (Beibehalten der $H_{0IV.9}$ (ii) mit $\beta_{Usability} < 0$) oder der Interaktion gefunden (siehe Tab. 9.1). Für die Valenz wurden sowohl ein nicht erwarteter signifikanter Effekte der Dauer als auch der erwartete Effekt der Usability (Hyp. IV.9 (iii) mit $\beta_{Usability} < 0$ angenommen) gefunden (siehe Tab. 9.1). Beide Effekte sind in Abbildung 9.5 c) bzw. d) dargestellt. Die Interaktion war nicht signifikant.

Zusammengefasst stützen die Ergebnisse zur Dauerschätzung und zur wahrgenommenen Usability die erwarteten und die nicht erwarteten Effekte. Die wahrgenommene Attraktivität hingegen wurde nicht wie vorhergesagt von den unabhängigen Variablen beeinflusst, während die Valenz von beiden Faktoren beeinflusst wurde. Bezogen auf die Mediatoranalyse gibt es also Evidenz für Pfad A zwischen der Dauer und der Dauerschätzung bzw. der Valenz sowie zwischen der Usability und der wahrgenommenen Usability bzw. der Valenz. Eine abschließende post-hoc Poweranalyse deutete darauf hin, dass bei vorliegenden Effektstärken die Analyse für die Betrachtung der Usability auf die UX-bezogenen Maße leicht unterpowert war (Anhang G.3).

9.2.2 Ergebnisse zur Teilfrage a): Abwägen der Erklärungsansätze durch Mediatoranalysen

Zum Abwägen der Erklärungsansätze für die Dauerbewertung und die Zeitflussbeurteilung wurden Mediatoranalysen durchgeführt. Hierzu wurden zunächst LMMs mit der gleichen Formalisierung wie für die Replikationshypothesen im Abschnitt 9.2.1 aufgesetzt, um Pfad C (vgl. Abb. 9.4) zu prüfen. Basierend auf diesen Ergebnissen und den Ergebnissen zu Pfad A (Replikationshypothesen) wurde dann für diejenigen Konstrukte eine Mediatoranalyse geprüft, für die sowohl Pfad A als auch Pfad C Signifikanz erreichte. Die Ergebnisse aller Analysen sind in Tabelle 9.2 einzusehen.

Die Analyse der Dauerbewertung zeigte einen Haupteffekt der Dauer, nicht

aber der Usabilitygruppe oder der Interaktion⁵. Der Effekt der Dauer ist in Abbildung 9.6 a) dargestellt. Auf Basis dieser Ergebnisse und der aus dem vorangegangenen Abschnitt kann nun die Mediatoranalyse durchgeführt werden. Es wird jedoch nur die Mediation des Dauereffekts auf die Dauerbewertung weiter betrachtet, da die Usability keinen signifikanten Einfluss auf die Dauerbewertung hatte (keine Evidenz für Pfad C). Daher kann dieser Effekt auch nicht mediert werden. Für die Dauer hingegen sind die beiden möglichen Mediatoren die Dauerschätzung und die Valenz, da es bei beiden einen Effekt der Dauer gab (Pfad A) und die Dauer die Dauerbewertung signifikant vorhersagte (Pfad C). Daher wurde für die Dauerschätzung sowie für die Valenz jeweils ein entsprechendes Mediatormodell berechnet (siehe Tab. 9.2).

Für beide möglichen Mediatoren zeigte sich, dass sie ein signifikanter Prädiktor der Dauerbewertung waren (Pfad B). So war die Dauerbewertung negativer (größere Werte), je länger die Dauer geschätzt wurde und je niedriger die Valenz war, siehe Abbildung 9.6 b) und c). Gleichzeitig reduzierte sich in den gemeinsamen Modellen (Mediatoranalyse Tab. 9.2) der Effekt der Dauer jeweils in geringem Maße (Pfad C). Daher können Dauerschätzung und Valenz als partielle Mediatoren für den Effekt der Dauer auf die Dauerbewertung gesehen werden. Vor allem für die Valenz ist die partielle Mediation jedoch sehr gering, da sich hier der Effekt der Dauer kaum reduzierte. Hier erhöhte sich allerdings die aufgeklärte Varianz, was darauf hindeutet, dass die Valenz einen anderen Teil der Varianz der Dauerbewertung aufklären kann als die Dauer.

Die Analyse des Zeitflusses wies keine signifikanten Effekte der unabhängigen Variablen auf⁶. Auf dieser Basis wurde keine weiterführende Mediatoranalyse durchgeführt.

⁵Für die Analyse der Dauerbewertung wurden zwei Werte mit NA ersetzt, da diese zu einer Abweichung der Residuen von der Normalverteilung führten. Dadurch veränderte sich jedoch nicht die Schlussfolgerung der Analyse.

⁶Für die Analyse der Zeitflussbeurteilung konnte auch mit einer Transformation keine Normalverteilung erreicht werden. Ein nicht-parametrisches Testverfahren (`bwtrim`) aus dem *WSR2*-Paket (Mair & Wilcox, 2018) für *R* zeigte jedoch die gleichen Effekte, wie die hier berichtete Analyse.

TABELLE 9.2: Effekte der unabhängigen Variablen auf die Dauerbewertung und das Zeitflussbeurteilung sowie Mediatoranalysen zur Dauerbewertung. Die obere und untere Analyse entspricht jeweils der Prüfung des Pfads C der Mediatoranalyse. Die mittleren Analysen entsprechen der Prüfung der Pfade B und C.

	β -Gewicht	95% Konf.	Std. Fehler	t-Wert	
Effekte der unabhängigen Variablen auf Dauerbewertung					
Intercept	39.07	[35.82, 42.32]	1.63	23.93	*
Dauer [Min.]	8.89	[7.29, 10.48]	0.80	11.10	*
Usab. hoch/niedrig	3.21	[-3.28, 9.72]	3.27	0.98	
Dauer:Usab.	-0.61	[-3.79, 2.59]	1.60	-0.38	
Random Effect Struktur: (1 <i>Versuchsperson</i>)					
$R^2_{GLMM(m)}$: 42.9 %, $R^2_{GLMM(c)}$: 60.2 %					
Dauerschätzung als Mediator zw. Dauer und Dauerbewertung					
Intercept	39.16	[35.78, 42.55]	1.70	23.08	*
Dauer [Min.]	6.89	[4.57, 9.27]	1.14	6.06	*
Dauersch. [Min.]	2.14	[0.2, 4.09]	0.92	2.33	*
Random Effect Struktur: (1 <i>Versuchsperson</i>)					
$R^2_{GLMM(m)}$: 43.7 %, $R^2_{GLMM(c)}$: 66.1 %					
Valenz als Mediator zwischen Dauer und Dauerbewertung					
Intercept	38.91	[35.84, 41.97]	1.54	25.28	*
Dauer [min]	8.52	[6.98, 10.06]	0.77	11.01	*
Valenz	-3.17	[-4.86, -1.49]	0.85	-3.72	*
Random Effect Struktur: (1 <i>Versuchsperson</i>)					
$R^2_{GLMM(m)}$: 48.8 %, $R^2_{GLMM(c)}$: 63.5 %					
Effekte der unabhängigen Variablen auf Zeitflussbeurteilung					
Intercept	50.65	[49.51, 51.79]	0.57	88.53	*
Dauer [Min.]	0.26	[-0.39, 0.92]	0.33	0.79	
Usab. hoch/niedrig	-0.75	[-3.03, 1.53]	1.14	-0.66	
Dauer:Usab.	-1.18	[-2.49, 0.13]	0.66	-1.79	
Random Effect Struktur: (1 <i>Versuchsperson</i>)					
$R^2_{GLMM(m)}$: 3.1 %, $R^2_{GLMM(c)}$: 17.2 %					

ANMERKUNG: Die Sterne geben an, ob der Effekt auf einem Signifikanzniveau von $\alpha = .05$ und zweiseitiger Testung signifikant ist.

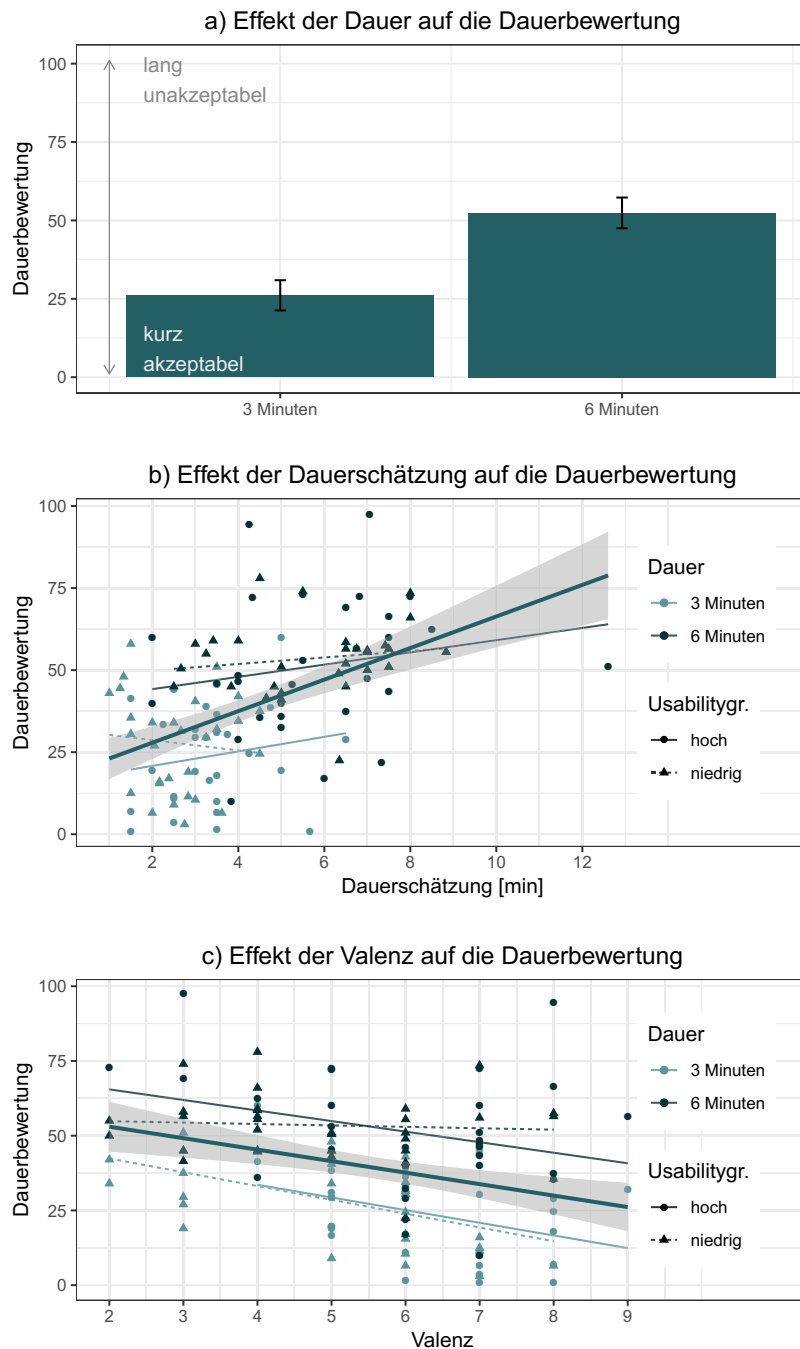


ABBILDUNG 9.6: Effekte der unabhängigen Variablen sowie der Dauerschätzung und Valenz auf die Dauerbewertung. a) Die Fehlerbalken stellen Konfidenzintervalle der Standardfehler (95 %) korrigiert für Innersubjektfaktoren (Cousineau, 2005; Morey, 2008) dar. b) und c) stellen ein Streudiagramm mit dem Zusammenhang (dicke Linie) samt Konfidenzintervall (grauer Bereich) der beiden Variablen dar. Als zusätzlichen Information sind die Punkte des Streudiagramms nach den vier Experimentalbedingungen aufgeteilt.

9.2.3 Ergebnisse zur Teilfrage b): Korrelationen

Die Korrelationen sind in Tabelle 9.3 dargestellt. Bezüglich der zeitbezogenen Maße zeigte sich nur eine signifikante Innersubjektkorrelation zwischen der geschätzten Aufgabendauer und der Dauerbewertung (Hyp. IV.4 (i): $r > 0$): Je länger die Probanden eine Dauer schätzten, umso negativer bewerteten sie diese. Es wurde jedoch weder Evidenz für einen Zusammenhang zwischen geschätzter Aufgabendauer und Zeitflussbeurteilung (Beibehalten der $H_{0IV.3}$ (iii): $r's \leq 0$) noch für den Zusammenhang zwischen Dauerbewertung und Zeitflussbeurteilung (Beibehalten der $H_{0IV.13}$: $r's \leq 0$) gefunden. Auch zwischen Probanden gab es keinen Hinweis auf einen Zusammenhang zwischen Dauerbewertung und Zeitflussbeurteilung.

Für die Zusammenhänge zwischen den zeit- und UX-bezogenen Variablen

TABELLE 9.3: Korrelationen zwischen den zeit-bezogenen und den UX-bezogenen Variablen.

Korrelationen innerhalb der Subjekte				
Hyp. IV.		geschätzte Aufg.dauer	Dauer- bewertung	Zeitfluss- beurteilung
4 (i)-(ii)	gesch. Aufg.dauer	–	.84 ***	.24, n.s.
13	Dauerbewertung	–	–	.22, n.s.
5 (ii)-(iv)	wahrg. Usability	.06, n.s.	-.14, n.s.	-.06, n.s.
6 (ii)-(iv)	wahrg. Attraktivität	-.11, n.s.	-.25 *	.01, n.s.
7 (ii)-(iv)	Valenz	-.24, n.s.	-.23, n.s.	.02, n.s.
8 (ii)-(iv)	Arousal	.07, n.s.	-.23, n.s.	.02, n.s.
Korrelationen zwischen Subjekten				
			Dauer- bewertung	Zeitfluss- beurteilung
13	Dauerbewertung		–	-.01, n.s.
5 (iii)-(iv)	wahrg. Usability		-.22, n.s.	-.18, n.s.
6 (iii)-(iv)	wahrg. Attraktivität		-.01, n.s.	.07, n.s.
7 (iii)-(iv)	Valenz		-.21, n.s.	.03, n.s.
8 (iii)-(iv)	Arousal		.02, n.s.	-.03, n.s.

ANMERKUNG: * bei $p < .05$, ** bei $p < .01$, *** bei $p < .001$, fett für $|r| \geq .3$

müssen fast alle Hypothesen abgelehnt werden (Beibehalten der $H_{0IV.5}$ (ii) bis (iv) zur wahrgenommenen Usability: $r's \geq 0$, Beibehalten der $H_{0IV.6}$ (ii) und (iv) zur wahrgenommenen Attraktivität: $r's \geq 0$, Beibehalten der $H_{0IV.7}$ (ii) bis (iv) zur Valenz: $r's \geq 0$ und Beibehalten der $H_{0IV.8}$ (ii) bis (iv) zum Arousal: $r's = 0$). Lediglich die Hypothese IV.6 (iii) ($r < 0$) zum Zusammenhang zwischen wahrgenommener Attraktivität und Dauerbewertung kann angenommen werden: Je attraktiver das System empfunden wurde, umso positiver war die Dauerbewertung. Dieser Zusammenhang zeigte sich jedoch nur innerhalb, nicht zwischen Probanden und war schwach ausgeprägt. Die Korrelationen zwischen den UX-bezogenen Variablen sind in Anhang G.4 zusammengefasst.

9.3 Diskussion

Ziel der vorliegenden Studie war die Dissoziation des Effekts der Dauerschätzung und des Effekts des Nutzererlebens auf die Dauerbewertung und die Zeitflussbeurteilung. Hintergrund hierfür war die Frage, ob Dauerbewertung und Zeitflussbeurteilung primär durch das Interaktionscharakteristikum Dauer oder durch andere Interaktionscharakteristika beeinflusst werden. Dazu wurden zwei unabhängige Variablen gewählt, die die Dauer von den anderen Charakteristika trennt und somit die Dauerschätzungen und das Nutzererleben als medierende Variablen unterschiedlich beeinflussen sollten: Dauer und Usability. Zusätzlich wurden die Korrelationen zwischen den zeit- sowie zwischen den zeit- und UX-bezogenen Variablen betrachtet.

9.3.1 Zusammenfassung und Beantwortung der Fragestellungen

Replikation der Effekte der UVn auf die Dauerschätzung und die Komponenten des Nutzererlebens

Die Ergebnisse zeigten eine erfolgreiche Manipulation der Dauerschätzung durch die Dauer und eine erfolgreiche Manipulation des Nutzererlebens durch die Usability bezüglich der wahrgenommenen Usability und der Valenz. Der Misserfolg der Manipulation der wahrgenommenen Attraktivität durch die

Usability kann durch das Versuchsdesign erklärt werden, da primär die Gebrauchstauglichkeit zwischen den Gruppen variiert wurde und keine der Versionen ein ansprechendes Design besaßen. Zusammengefasst konnten durch das experimentelle Design also die Dauer und das Nutzererleben bezüglich instrumenteller Qualitäten und Emotion getrennt voneinander beeinflusst werden. Dies ermöglicht die Beantwortung von Teilfrage a).

Teilfrage a): Wird die Dauerbewertung und die Zeitflussbeurteilung primär von der Dauerschätzung oder vom Nutzererleben beeinflusst?

Bezüglich des Abwägens der Einflusspfade aus Abbildung 9.1 (S. 138) wurden die Operationalisierungen für Zeitempfinden im Sinne einer Dauerbewertung und Zeitflussbeurteilung getrennt voneinander betrachtet. Für die Dauerbewertung stützte die Mediatoranalyse den Pfad vom Interaktionscharakteristikum Dauer über die Dauerschätzung zur Dauerbewertung. Emotionale Aspekte schienen in sehr schwachem Maße den Effekt der Dauer auf die Dauerbewertung ebenfalls partiell zu mediieren. Laut Baron und Kenny (1986) weisen partielle Mediationen darauf hin, dass es noch weitere Mediatoren gibt, die im vorliegenden Design nicht berücksichtigt wurden. Die Beeinflussung der Dauerbewertung durch die Dauern wird also wahrscheinlich nicht nur durch die Dauerschätzung und im geringen Maße durch die emotionale Bewertung der Interaktionsdauer vermittelt, sondern auch durch andere Prozesse, die hier nicht untersucht wurden.

Neben der Frage nach weiteren Mediatoren, bleibt auch die Frage offen, welche Interaktionscharakteristika neben der Dauer relevant für die Entstehung von Dauerbewertungen sind. Hier ist insbesondere der Effekt der Valenz interessant. Dieser scheint einen anderen Teil der Varianz aufzuklären, als die Dauer. Daher kann angenommen werden, dass Nutzer ihren emotionalen Zustand bei der Bewertung einer Interaktionsdauer berücksichtigen. Wodurch Unterschiede in der Emotion genau entstanden, ist in dieser Studie unklar. Hier ist weitere Forschung notwendig, um das Entstehen eines bestimmten emotionalen Zustandes und um die verschiedenen Einflüsse der Interaktionscharakteristika auf die Dauerbewertung besser zu verstehen.

Für die Zeitflussbeurteilung gab es keine Evidenz für eine Beeinflussung

durch die Dauer oder die Usability des Systems. Daher können keine Aussagen zur Teilfrage a) zur Entstehung der Zeitflussbeurteilung getroffen werden.

Zusammengefasst wird die Dauerbewertung primär durch die Dauer, mediert durch die Dauerschätzung, und durch den emotionalen Zustand des Nutzers beeinflusst. Der Effekt der Dauer scheint aber auch über weitere, noch unbekannte Variablen mediert zu werden. Bezogen auf die Zeitflussbeurteilung waren die Ergebnisse nicht eindeutig, da die Zeitflussbeurteilung weder durch die Dauer der Interaktion noch die Usability des Systems beeinflusst wurde. Daher kann Teilfrage a) zur Zeitflussbeurteilung nicht beantwortet werden.

Teilfrage b): Wie hängen die Operationalisierungen der Dauerwahrnehmung untereinander und mit den Komponenten des Nutzererlebens bei konstanten Dauern zusammen?

Die Betrachtung der Korrelationen zeigte überraschende Ergebnisse. So konnte zwar repliziert werden, dass Nutzer die erlebte Dauer umso negativer bewerten, je länger sie sie einschätzen. Andere Zusammenhänge, wie beispielsweise Korrelationen zwischen Dauerschätzung bzw. Dauerbewertung und Zeitflussbeurteilung sowie zwischen zeit- und UX-bezogenen Variablen wurden in Studie IV jedoch nicht gestützt.

Eine mögliche Erklärung könnte darin liegen, dass sich die betrachtete Interaktion in der vorliegenden Studie stark von denen der anderen Studien unterschied. Während in Studien I, II und III lebensnahe Aufgaben mit zumindest teilweise real existierenden Webseiten absolviert werden mussten, war die Aufgabe in Studie IV sehr künstlich und knüpfte weit weniger an Vorerfahrungen der Probanden an. Dadurch kann es den Probanden schwerer gefallen sein, Konstrukte wie Zeitflussbeurteilung oder Nutzererleben zu bewerten. Gegen solche Schwierigkeiten in der Bewertung sprechen jedoch die Korrelationen zwischen den UX-bezogenen Variablen. Diese waren vergleichbar zu den Korrelationen aus Studien I bis III. Darüber hinaus wurden die UX-bezogenen Variablen wie erwartet von der Manipulation des Systems beeinflusst. Dies deutet darauf hin, dass Probanden zumindest in Bezug auf ihre UX eine sinnvolle Bewertung abgeben konnten. Eine alternative Erklärung besteht darin, dass bei einer solchen Überwachungs- und Regulationsaufgabe die Dauer nicht die gleiche Bedeutung hat wie bei Interaktionen mit einer Webseite. Bei solchen

Aufgaben könnte Zeit grundsätzlich eine andere Rolle spielen, da hier zeitliche Abfolgen hauptsächlich durch das System und nicht durch den Nutzer gesteuert werden. Darüber hinaus ist hier die Gesamtdauer von der Aufgabe an sich losgelöst.

Abschließend muss auf Grund der sehr geringen Korrelationen zwischen zeit- und UX-basierten Maßen festgehalten werden, dass sich die Zusammenhänge, die in Studien I bis III gefunden wurden, nicht ohne weiteres auf andere Kontexte als Interaktionen mit Webseiten verallgemeinern lassen.

9.3.2 Limitationen von Studie IV

Studie IV weist verschiedene Limitationen auf, zum einen im Versuchsdesign und zum anderen in der statistischen Analyse. Bezogen auf das Versuchsdesign ist es möglich, dass der Zusammenhang zwischen Dauerschätzung und Dauerbewertung überschätzt wurde, da die Interaktion mit dem System sehr künstlich war und zwei extreme Dauerbedingungen getestet wurden. Einen Hinweis hierauf liefert die Tatsache, dass die Innersubjektkorrelation zwischen diesen beiden Konstrukten stärker war als in den vorangegangenen Studien. Möglicherweise hatten die Probanden in der hier untersuchten Simulationsumgebung eines Heizungssystems zu wenig Anhaltspunkte und zu wenig Erfahrung, um die Dauer bewerten zu können. Daher wurde das am einfachsten zugängliche Attribut der Interaktion genutzt, die geschätzte Dauer. Dies kann durch die unmittelbare zeitliche und räumliche Nähe der Erfassung beider Konstrukte noch begünstigt worden sein. Zusätzlich wurde die Dauer in zwei leicht unterscheidbaren Stufen variiert, wodurch sie besonders stark als Attribut der Interaktion hervorstach. Folgestudien sollten daher eine kleinere Differenz zwischen den Dauerstufen betrachten und eine alltagsnähere Aufgabe finden.

Die Dauerstufen waren in der vorliegenden Studie außerdem mit der Reihenfolge der Interaktionen konfundiert. Auf Grund einer weiteren Forschungsfrage, die hier nicht näher beschrieben wurde, musste die Dauer der Interaktionen für alle Probanden konstant gehalten werden, wodurch die erste Interaktion immer drei Minuten und die zweite Interaktion sechs Minuten betrug. Daher kann nicht ausgeschlossen werden, dass manche Effekte der Dauer durch diese Reihenfolge entstanden. Ein Beispiel hierfür könnte der Effekt der Dauer auf die Valenz sein: Die Abnahme der Valenz von drei zu sechs Minuten kann

an der längeren Dauer oder an der späteren Interaktion gelegen haben. Das vorliegende Versuchsdesign erlaubt hier keine Unterscheidung.

Darüber hinaus setzte die Manipulation der unabhängigen Variablen auf verschiedenen Stufen an. Während die Dauer ein Interaktionscharakteristikum darstellt, ist die Usability eine Systemeigenschaft, die die Interaktionscharakteristika zusammen mit Merkmalen der Person, des Kontexts und der Aufgabe beeinflusst. Daher ist in dem vorliegenden Versuchsdesign nicht klar, welche Interaktionscharakteristika durch die Usability-Manipulation beeinflusst wurden und wie diese Merkmale der Interaktion über die unterschiedlichen Probanden variierten. Der Vergleich zwischen dem Einfluss der Dauer und dem der Usability auf die Dauerbewertung ist somit nur unter Vorbehalt interpretierbar, da der Einflussweg der Usability wahrscheinlich über eine Vielzahl verschiedener Interaktionscharakteristika vermittelt wurde (vgl. Abb. 9.1). Hier wird jedoch eine Schwäche des CUE-Modells deutlich. In dem Modell werden Interaktionscharakteristika nicht näher spezifiziert, wodurch es schwierig ist, diese messbar zu machen und sie zu diskriminieren.

Eine weitere Limitation der Studie liegt in der Stichprobengröße. Die post-hoc Poweranalysen zeigten, dass ein Effekt der Usability auf die UX-bezogenen Variablen mit der vorliegenden Effekstärke so klein war, dass er mit der gegebenen Stichprobengröße nur mit einer mittleren Wahrscheinlichkeit (Power von $48\% \leq 1 - \beta \leq 79\%$) entdeckt werden kann. Dies deutet darauf hin, dass der Unterschied in der objektiven Usability zwischen den beiden Simulationen für die vorliegende Stichprobengröße eventuell nicht groß genug war. Dadurch können Beeinflussungen der Usability und damit auch des Nutzererlebens auf das Zeitempfinden unterschätzt worden sein. In Folgestudien sollte daher entweder der Unterschied zwischen den Simulationsumgebungen größer gemacht oder die Probandenanzahl erhöht werden.

Auf Seiten der statistischen Analysen müssen ebenfalls Limitationen beachtet werden. Ein mögliches Problem liegt in der Mediatoranalyse zur Dauerbewertung, da im Modell mit den Prädiktoren Dauer und Dauerschätzung Multikollinearität vorliegt. Diese tritt auf, wenn die Prädiktoren stark miteinander korrelieren, und kann zu Ungenauigkeiten in der Schätzung des β -Gewichts und zu Verzerrungen in den Teststatistiken führen (Bortz, 2005). Daher sind durch die genutzte Analysestrategie nach Baron und Kenny (1986) der tat-

sächliche Effekt der Dauer und der der Dauerschätzung nicht voneinander abgrenzbar. Um diese Probleme zukünftig zu umgehen, sollten weiterführende Studien größere Stichproben nutzen, um andere Verfahren, wie beispielsweise Strukturgleichungsmodelle, nutzen zu können.

9.3.3 Fazit Studie IV

In Studie IV konnte experimentell gezeigt werden, dass Nutzer eines Systems ihre Dauerbewertung primär auf der wahrgenommenen Dauer und nicht auf der wahrgenommenen Usability basieren. Es gibt jedoch auch weitere Einflussgrößen, zu denen beispielsweise die Valenz gehört. Für das Bewerten des Zeitflusses konnte dies nicht gezeigt werden. Zusätzlich konnten in Studie IV die gefundenen Zusammenhänge zwischen zeit- und UX-basierten Maßen aus den vorangegangenen Studien nicht repliziert werden. Sie scheinen also nicht auf jede Mensch-Computer-Interaktion verallgemeinerbar sind.

Insgesamt liefert Studie IV ein Beispiel, wie Interaktionscharakteristika gezielt untersucht werden können, um ihre Bedeutung für das Entstehen von UX zu klären und theoretische Ableitungen für beispielsweise eine Weiterentwicklung des CUE-Modells zu schaffen. Eine solche Ableitung ist hier für die Dauer der Interaktion als ein Charakteristikum gegeben: Die Dauer einer Interaktion beeinflusst das Nutzererleben in keiner relevanten Stärke, so lange die Dauer von Systemeigenschaften, wie der objektiven Usability, unabhängig ist. Da diese Ableitung jedoch nur auf Basis einer Studie erfolgte und hier die Aufgabe sehr künstlich war, sollte sie weiter geprüft werden. Trotzdem zeigt Studie IV auf, dass auch das Kästchen „Interaktionsmerkmale“ im CUE-Modell stärker beschrieben und experimentell untersucht werden kann, um das Modell weiter zu konkretisieren und es gegebenenfalls auch für eine mathematische Simulation der Entstehung von Nutzererleben nutzbar zu machen.

Teil III

Diskussion

Kapitel 10

Allgemeine Diskussion

Die Dauer von Ereignissen kann von Menschen wahrgenommen und bewertet werden. Sie ist ein wichtiger Bestandteil des Erlebens, der Einfluss auf Bewertungen, Entscheidungen und das Verhalten nehmen kann (Ariely & Zakay, 2001). Dabei ist die Wahrnehmung einer Dauer meist kein exaktes Abbild der objektiven Dauer, da sie verschiedenen kognitiven Prozessen unterliegt, die zu Interferenzen und damit Verzerrungen in der subjektiven Zeit führen kann. Im Kontext der Mensch-Computer-Interaktion werden Dauern häufig objektiv gemessen, während die subjektive Wahrnehmung und Bewertung von Dauern meist vernachlässigt wird (Liikkanen & Gómez Gómez, 2013).

Das Ziel der vorliegenden Arbeit ist es, den Einfluss von Interaktionsdauern und deren Wahrnehmung auf das Nutzererleben besser zu verstehen. Dies ist aus mehreren Perspektiven relevant. Zum einen ist unklar, welche zeitlichen Dimensionen es gibt und wie diese miteinander interagieren. Hierzu gibt es aus dem Grundlagenbereich viel Forschung zu Dauerschätzungen, jedoch weit aus weniger Arbeiten zum Zeitempfinden. Zwar hat sich in den letzten Jahren Forschung zu Zeitflussbeurteilungen etabliert, zu einer abschließenden Dauerbewertung gibt es jedoch bis jetzt kaum experimentelle Arbeiten. Hier leistet die vorliegende Arbeit einen Beitrag. Zum anderen können Dauern im Rahmen des CUE-Modell (Mahlke, 2008; Minge & Thüring, 2018; Thüring & Mahlke, 2007) als ein Interaktionscharakteristikum gesehen werden, welches während der Interaktion eines Nutzers mit einem System entsteht. Zwar stellt das CUE-Modell die Annahme auf, dass Interaktionscharakteristika als vermittelnde Komponente zwischen dem Nutzer, dem System und dem Kontext

auf der einen Seite und dem Nutzererleben auf der anderen Seite fungieren, es definiert jedoch nicht, was diese Interaktionscharakteristika genau sind und wie die zugehörigen Prozesse aussehen. In der vorliegenden Arbeit wird ein ausgewähltes Charakteristikum, die Dauer, in das Zentrum der Untersuchungen gestellt, um die hier stattfindenden Prozesse besser zu verstehen.

Im Mittelpunkt der Bearbeitung des Forschungsziels stehen Interaktionen zwischen Nutzern und Computern, in denen eine spezifische Aufgabe innerhalb von wenigen Minuten gelöst werden soll. Auch wenn kein Zeitdruck besteht, so hatte die Interaktionsdauer für die Nutzer eine hohe Relevanz, da die Interaktion an sich nur ein Hilfsmittel für eine möglichst schnelle und einfache Zielerreichung ist. Laut Zakay (2015) führt eine hohe temporale Relevanz zu einer verstärkten Zuwendung von Aufmerksamkeit auf das Verstreichen der Zeit, wodurch Dauern prospektiv wahrgenommen werden. Entsprechend steht das prospektive Paradigma der Dauerwahrnehmung im Fokus dieser Arbeit. Im Rahmen der ersten Forschungsfrage zur Wahrnehmung von Dauern während einer Mensch-Computer-Interaktion wurden in Studien I, II und IV verschiedene Interaktionscharakteristiken kreiert. Dies geschah über die Manipulation der Usability, die Gabe einer Zweitaufgabe oder die Kontrolle der Interaktionsdauer. Dabei wurde untersucht, wie sich solche Merkmale der Interaktion auf verschiedene zeitliche Dimensionen auswirken und ob diese zu Grundlagenbefunden und -modellen passen. Unter zeitlichen Dimensionen wurden sowohl prospektive Dauerschätzungen, als auch Bewertungen der erlebten Dauer und des Zeitflusses verstanden. Diese Operationalisierungen von Dauerwahrnehmung wurden in allen vier Studien auf ihren Zusammenhang untereinander untersucht. Im Rahmen der zweiten Forschungsfrage zum Zusammenhang zwischen Dimensionen der Dauerwahrnehmung und des Nutzererlebens wurden in Studien I, II, III und IV Korrelationen zwischen den verschiedenen Dimensionen der Dauerwahrnehmung und des Nutzererlebens untersucht.

Die Diskussion der Befunde der vier Studien wird im Folgenden anhand der Dimensionen der Dauerwahrnehmung gegliedert. Die Korrelationen aus allen vier Studien sind hier erneut dargestellt, damit sie studienübergreifend verglichen werden können (Tab. 10.1). Anschließend wird diskutiert, welche Konsequenzen auf Basis der vorliegenden Ergebnisse für die Konzeption des Nutzererlebens und für den Anwendungsbereich gezogen werden sollten.

10.1 Dauerwahrnehmung während Mensch-Computer-Interaktionen

10.1.1 Dauern schätzen - prospektive Dauerwahrnehmung

In allen vier Studien erlebten Probanden Interaktionen mit einem Computer und schätzten anschließend die Dauer der jeweiligen Interaktion prospektiv ein. Dabei variierten die Interaktionsdauern zwischen einem Minimalwert von 20 *Sek.* und einem Maximalwert von 487 *Sek.*, $M_{gesamt} = 179 \text{ Sek.}$, $SD_{gesamt} = 93 \text{ Sek.}$. In Abbildung 10.1 ist diese Bandbreite zusammen mit den Dauerschätzungen dargestellt. An ihr lassen sich verschiedene Eigenschaften von prospektiven Dauerschätzungen verdeutlichen, die größtenteils im Einklang mit Annahmen über kognitive Prozesse und mit Befunden im Bereich der prospektiven Dauerschätzung stehen. Diese umfassen die skalare Eigenschaften, Varianzen und Überschätzungen. Darüber hinaus werden in diesem Unterkapitel Effekte von Arousal und mentaler Beanspruchung diskutiert.

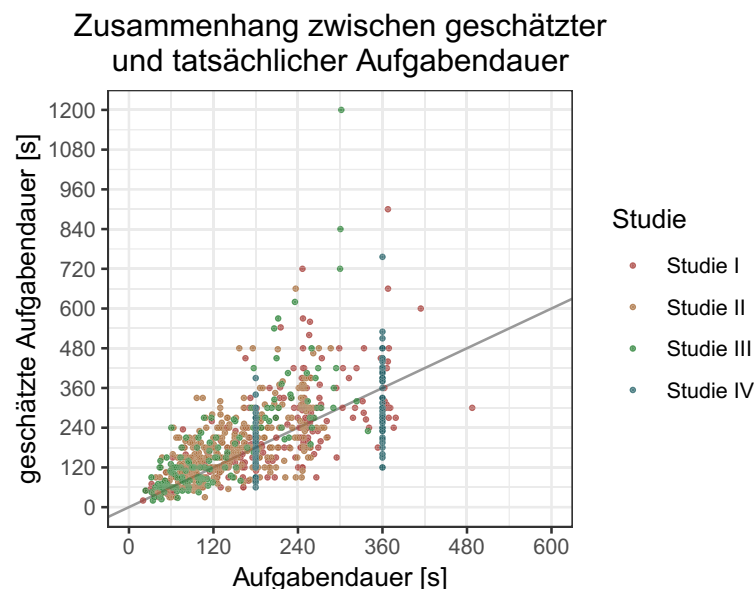


ABBILDUNG 10.1: Streudiagramm der Dauern und Dauerschätzungen aus Studien I bis IV. Die graue Linie zeigt die „perfekte Schätzung“ an.

Skalare Eigenschaften, Varianzen und Überschätzungen

Zunächst verdeutlicht die Abbildung 10.1, dass in allen vier Studien längere Dauern auch mit längeren prospektiven Schätzungen einhergingen. Dies wurde durch starke positive Korrelationen zwischen den beiden Maßen (Studien I bis III, siehe Tab. 10.1) bzw. durch eine deutliche Beeinflussung des einen Maßes durch das andere (Studie IV) gezeigt. Der Zusammenhang zwischen Aufgabendauer und Dauerschätzung stützt die theoretische Annahme, dass Dauerschätzungen skalare Eigenschaften haben, wie sie von Modellen mit einem internen Taktgeber, sogenannten „timing-with-a-timer“-Modellen (Block & Zakay, 1996), angenommen werden. Vertreter solcher Modelle sind das Informations-Prozess-Modell der Dauerwahrnehmung (Treisman, 1963, 2013), die STT (Gibbon et al., 1984) und das AGM (Block & Zakay, 1996; Zakay & Block, 1995, 1997). Sie nehmen an, dass ein interner Taktgeber in regelmäßigen Abständen Impulse generiert, die als Informationsquelle für die prospektive Schätzung dienen. Durch diesen Mechanismus skaliert die Dauerschätzung mit der Dauer. Die hier vorliegenden Befunde zeigen also, dass auch im Anwendungskontext von Mensch-Computer-Interaktionen die genannten Modelle und die kognitive Annahme eines Taktgebers mit der empirischen Evidenz vereinbar sind.

Darüber hinaus sagen „timing-with-a-timer“-Modelle voraus, dass längere Dauern mit einer größeren Varianz geschätzt werden als kürzere. Grundlage für diese Annahme sind veranschaulichte kognitive Prozesse, wie sie beispielsweise beim Zugriff auf Referenzdauern entstehen (vgl. Information-Process Model on Timing) oder beim Öffnen und Schließen des Schalters vorliegen, der den Beginn und das Ende eines zu schätzenden Intervalls kennzeichnet (vgl. STT). Auch diese Vorhersage ist im Einklang mit den vorliegenden Befunden: Betrachtet man Abbildung 10.1 wird eine trichterförmige Verteilung der Daten deutlich. Sie zeigt, dass die Streuung der Dauerschätzungen mit steigender Dauer größer wird. Auch dieser Befund entspricht bestehender experimenteller Evidenz (z.B. Allan, 1979; Brown, 1997; Droit-Volet et al., 2018; Wearden & Lejeune, 2008)

In Abbildung 10.1 ist zusätzlich die Diagonale eingezeichnet, die eine fehlerfreie Schätzung veranschaulicht. Die trichterförmige Verteilung zeigt eine leichte Verschiebung nach oben, was die Tendenz der Probanden widerspie-

TABELLE 10.1: Vergleich der Innersubjektkorrelationen zwischen den Studien.

Var. X	Var. Y	Hyp.	I	II	III	IV
Aufgaben- dauer	gesch. Auf.dauer	3(i)	+++	+++	+++	
	Dauerbewertung	3(ii)	++	++	+++	
	Zeitflussbeurteilung	3(iii)		+	++	
geschätzte	Dauerbewertung	4(i)	++	++	+++	+++
Aufgabendauer	Zeitflussbeurteilung	4(ii)		+	++	n.s.
Dauerbewertung	Zeitflussbeurteilung	13		++	+++	n.s.
wahrgen.	Aufgabendauer	5(i)	n.s.	-	- -	
Usability	gesch. Auf.dauer	5(ii)	-	-	- -	n.s.
	Dauerbewertung	5(iii)	- -	- -	- - -	n.s.
	Zeitflussbeurteilung	5(iv)		-	- -	n.s.
wahrgen.	Aufgabendauer	6(i)	n.s.	-	n.s.	
Attraktivität	gesch. Auf.dauer	6(ii)	n.s.	-	n.s.	n.s.
	Dauerbewertung	6(iii)	- -	- -	-	-
	Zeitflussbeurteilung	6(iv)		-	n.s.	n.s.
Valenz	Aufgabendauer	7(i)	-	- -	- - -	
	gesch. Auf.dauer	7(ii)	- -	-	- - -	n.s.
	Dauerbewertung	7(iii)	- -	- - -	- - -	n.s.
	Zeitflussbeurteilung	7(iv)		- -	- - -	n.s.
Arousal	Aufgabendauer	8(i)	n.s.	n.s.	-	
	gesch. Auf.dauer	8(ii)	n.s.	n.s.	-	n.s.
	Dauerbewertung	8(iii)	n.s.	n.s.	-	n.s.
	Zeitflussbeurteilung	8(iv)		-	- -	n.s.

ANMERKUNG: + und - weisen auf die Richtung des Zusammenhangs hin. +++ bzw. - - - für starke, sign. Koeffizienten mit $r \geq .5$, ++ bzw. - - für mittlere, sign. Koeffizienten mit $r \geq .3$ und + bzw. - für kleine, signifikante Koeffizienten mit $r \geq .1$. n.s. für nicht signifikant.

gelt, die Aufgabendauern zu überschätzen. Dies zeigte sich in den Studien I bis III durch mittlere PTR-Werte größer als eins. Der Befund, dass Dauern im Sekunden- bis Minutenbereich häufig überschätzt werden, steht mit anderen Arbeiten in Einklang. So zeigten Droit-Volet et al. (2018) vergleichbare Überschätzungen von Dauern zwischen zwei und 16 Minuten. Ähnlich dazu berichteten Hertzum und Holmegaard (2015) ebenfalls Überschätzungen bei ca. zweiminütigen Intervallen, in denen Mastermind gespielt wurde.

Droit-Volet et al. (2018) argumentieren, dass eine solche Überschätzung dafür spricht, dass der interne Taktgeber während eines mehrminütigen Intervall nicht kontinuierlich seine Pulse akkumuliert, sondern dass eine Unterteilung des Intervalls in kleinere Abschnitte erfolgt. Sie berufen sich bei diesem Argument darauf, dass „timing-with-a-timer“-Modelle die Annahme teilen, dass die Akkumulation der Impulse im Arbeitsgedächtnis erfolgt. Dieses unterliegt zeitlichen Beschränkungen (Baddeley, 2003), wodurch für eine mehrminütige Speicherung ein inneres Wiederholen (*rehearsal*) erforderlich wäre. Durch dieses Wiederholen würde eine Ablenkung von der Dauerschätzung erfolgen, wodurch Impulse verloren gingen und kürzere Dauern geschätzt würden. Daher argumentieren Droit-Volet et al. (2018), dass mehrminütige Intervalle nicht als ein Ereignis kodiert werden, sondern aus verschiedenen Ereignissen bestehen, für die jeweils eine Dauerschätzung vorliegt, die im Langzeitgedächtnis gespeichert und bei der Schätzung des gesamten Intervalls abgerufen und aufsummiert wird.

Die Ergebnisse aus Studie I bis III passen insofern zu dieser Argumentation, als dass sie auch eine Über- und keine Unterschätzung zeigten. Trotzdem stützen sie die Argumentation von Droit-Volet et al. (2018) nicht direkt, da die Überschätzung auch durch eine falsche Referenzdauer entstanden sein könnte und nicht zwangsläufig auf Basis einzeln kodierter Ereignisse entstanden sein muss. Um diese unterschiedlichen Erklärungen über kognitive Prozesse gegeneinander abwägen zu können, werden weitere Studien benötigt.

In Studie IV hingegen unterschätzten die Probanden die Dauer im Mittel. Dies kann verschiedene Ursachen haben, wie beispielsweise die bearbeitete Aufgabe. Die Aufgaben in Studien I bis III bestanden größtenteils aus der Suche nach Informationen auf verschiedenen Webseiten, wobei die Art der Interaktion bezüglich Aufgaben, Aufbau und Gestaltung der Webseiten sowie

Handlungsschritte für die meisten Nutzer sehr vertraut gewesen sein muss. Hier sollte die Aufgabe zwar möglichst schnell abgeschlossen werden, einzelne Teilhandlungen konnten jedoch im Tempo der Probanden ausgeführt werden. Im Gegensatz zu diesen recht alltäglichen Interaktionen bestand die Aufgabe in Studie IV darin, ein Heizungssystem zu überwachen und zu regulieren, womit kaum ein Proband Vorerfahrungen hatte. Darüber hinaus war hier jede Systemeingabe zeitlich anzupassen, da ein zu langsames oder zu schnelles Regulieren zu einem Fehler im System führen konnte, wodurch mehr Regulierungshandlungen erforderlich waren. Diese andere Art von Aufgabe könnte zum einen zu einer stärkeren mentalen Beanspruchung geführt haben, welche wiederum eine Verkürzung der Dauerschätzungen erzeugt hat. Dies wäre im Einklang mit dem AGM (Block & Zakay, 1996; Zakay & Block, 1995, 1997). Zum anderen kann mit der ungewohnten Aufgabe und der fehlenden Vorerfahrung auch eine unzuverlässliche Referenz für die Schätzung einhergegangen sein. Diese beiden Erklärungen ordnen die Entstehung der Unterschätzung verschiedenen Prozessschritten einer Dauerschätzung zu, und zwar Prozessen des Taktgebers und Prozessen des Gedächtnisses (*clock process* bzw. *memory process*, Gibbon et al., 1984). Beide Erklärungen schließen sich jedoch nicht aus und können gemeinsam zu einer Unterschätzung der Dauern in Studie IV geführt haben.

Effekte von Arousal

Eine wesentliche Annahme in den im Theorieteil vorgestellten „timing-with-a-timer“-Modellen liegt in der Funktionsweise des internen Taktgebers. Die Frequenz seiner Impulserzeugung wird vom Arousal-Level der Testperson beeinflusst, wodurch bei Erhöhung des Arousal-Levels eine Verlängerung von verbalen Dauerschätzungen entsteht. Dieser Effekt wurde meist im Rahmen von Veränderungen im emotionalen Zustand untersucht und für Dauern unter zwei Sekunden häufig experimentell bestätigt (Gil & Droit-Volet, 2012; Grommet et al., 2011; Trapp & Thüring, 2016). Bei längeren Dauern konnte eine Verlängerung jedoch nur bei sehr starken Veränderungen im emotionalen Zustand nachgewiesen werden (z.B. Castellà et al., 2017; Fayolle et al., 2015; Matell et al., 2004). Im Rahmen des Theorieteils wurde daher die Frage aufgeworfen, ob Emotionen, die während einer Mensch-Computer-Interaktion erlebt werden, stark genug sind, dass sie zu einem Effekt des Arousals auf prospektive

Dauerschätzungen bei mehrminütigen Intervallen führen.

In Studien I und II wurde der emotionale Zustand der Probanden indirekt durch die Usability des Systems manipuliert. Wie andere Autoren bereits gezeigt haben (z.B. Thüring & Mahlke, 2007; Brandenburg & Backhaus, 2016), wurde auch in Studien I und II der emotionale Zustand negativer, wenn Probanden mit einer Webseite mit niedriger Usability arbeiten mussten. Dies zeigte sich jedoch primär in Bezug auf die Valenz, nicht in Bezug auf das selbstberichtete Arousal. Das Arousal unterschied sich in Studie I vor allem zwischen den Usability-Gruppen, wenn die Probanden keine Zweitaufgabe während der Interaktion bearbeiten mussten. In Studie II wurde keine Evidenz für einen Unterschied zwischen den beiden Usability-Gruppen gefunden. Die post-hoc Betrachtung der erreichten Power für die Manipulation des Arousal durch die Usability zeigte darüber hinaus für beide Studien eine geringe bis mittlere Teststärke. Zusammengefasst war der vorliegende Effekt in beiden Studien zu klein und die Stichprobe zu gering, um eine verlässliche Aussage über die Beeinflussung des Arousal-Levels durch die Usability zu treffen. Die kleinen Effektstärken legen jedoch nahe, dass das Arousal-Level bei den hier untersuchten Mensch-Computer-Interaktionen keine große Rolle spielte und kaum variierte.

Basierend auf der fehlenden Evidenz für einen Effekt der Usability auf das Arousal-Level ist es nicht verwunderlich, dass keine Unterschiede in der Überschätzungstärke (PTR-Werte) zwischen den Usability-Gruppen gefunden wurden. Dies kann zum einen an den kleinen Varianzen im Arousal-Level gelegen haben. Zum anderen stimmt dies aber auch mit Befunden überein, die den Arousal-Effekt nur bei Intervallen unter 2 Sek. berichten (z.B. Angrilli et al., 1997; Nather et al., 2011). Die vorliegenden Ergebnisse stützen zwar nicht, widerlegen aber auch nicht die Annahme, dass Aufmerksamkeitsprozesse bei längeren Dauern eine zunehmend wichtigere Rolle spielen (Ivry & Schlerf, 2008; Noulhiane et al., 2007) und Arousal-Effekte überlagern können. Beide Interpretationsansätze (schwache Schwankungen in der Emotion und zunehmende Aufmerksamkeitsprozesse) stehen nicht im Widerspruch zueinander, sondern passen zu den vorgestellten Befunden.

Effekte der Aufmerksamkeitsallokation

Die Bedeutung der Aufmerksamkeitsallokation bei prospektiven Dauerschätzungen von mehrminütigen Intervallen konnte in den Studien I und II durch den Effekt von lautem Denken auf die Überschätzung von Dauern weiter gestützt werden. In beiden Studien zeigten Probanden, die während der Bearbeitung der Aufgaben laut dachten, eine geringere Überschätzung der Aufgabendauer als Probanden, die nicht laut dachten. Dieser Befund entspricht dem Interferenzeffekt bei prospektiven Dauerschätzungen durch Zweitaufgaben (Brown, 1997, 2008).

Die Befunde sind jedoch nicht nur mit den theoretischen Annahmen des AGM im Einklang, sondern sie zeigen auch auf, wie robust der Effekt von geteilter Aufmerksamkeit auf prospektive Dauerschätzungen ist. Selbst während einer mehrminütigen Mensch-Computer-Interaktion, in der es viele verschiedene Einflüsse auf den Nutzer und sein Erleben von Zeit gibt, führt eine Ablenkung von der Schätzaufgabe zu einer kleineren Dauerschätzung. Auf Grund dieser Robustheit kann der Zusammenhang zwischen geteilter Aufmerksamkeit und Dauerschätzungen im Sinne eines indirekten Maßes für mentale Beanspruchung im Anwendungskontext genutzt werden. Dieser Vorschlag basiert auf der Annahme, dass erhöhte mentale Beanspruchung Aufmerksamkeitsressourcen bindet, wodurch weniger Ressourcen für die Erstaufgabe, hier die Dauerschätzung, zur Verfügung stehen (Wickens, 2002, 2008). Lind und Sundvall (2007) berichteten bereits, dass die Stärke der Überschätzung hoch mit subjektiven Maßen der mentalen Beanspruchung während verschiedener Aufgaben am Computer korreliert. Baldauf et al. (2009) zeigten passend dazu in einem Fahrsimulator, dass mental beanspruchende Fahrsituationen zu einer Verkürzung von Dauerschätzungen führten. Die Autoren kontrollierten hier die mentale Beanspruchung der Fahrsituation sowohl mit subjektiven als auch mit physiologischen Maßen. In beiden Studien wurde die mentale Beanspruchung jedoch innerhalb von Subjekten variiert. Studien I und II erweitern hier die Befundlage, da sie aufzeigen, dass auch eine Variation der mentalen Beanspruchung zwischen Subjekten einen vergleichbaren Effekt auf Dauerschätzungen hat. Dauerschätzungen scheinen auch in einem Zwischensubjekt-Design mit frei variierenden Aufgabendauern als Indikator für die mentale Beanspruchung geeignet zu sein.

Fazit zu Dauerschätzungen

Zusammengefasst sind Nutzer dazu in der Lage, Interaktionsdauern im Minutenbereich einzuschätzen. Sie neigen jedoch zu Überschätzungen und zeigen entsprechend der Vorhersagen des Information-Process Model of Timing und der Scalar Expectancy Theory (SET) größere Varianzen in ihren Schätzungen, je länger die tatsächliche Dauer ist. Der Zustand der Person bezüglich Emotionen und Aufmerksamkeitsverteilung, der laut Information-Process Model of Timing, STT, Dual-Process Contingency Model und AGM die Verzerrung der Schätzung beeinflussen kann, ist im Anwendungsbereich nur teilweise von Relevanz. Emotionen, wie sie in den vorliegenden Studien auftraten, beeinflussen die Schätzverzerrung nicht signifikant. Eine erhöhte mentale Beanspruchung hingegen zeigt die erwarteten Verzerrungen in den Dauerschätzungen, wenn die mentale Beanspruchung durch eine aktive Zweitaufgabe erzeugt wird. Dieser Zusammenhang kann zur Messung von mentaler Beanspruchung genutzt werden.

10.1.2 Zeitempfinden

Im Zentrum der vorliegenden Arbeit stehen jedoch nicht nur Dauerschätzungen, sondern auch abstraktere Maße, wie das Zeitempfinden. Hierzu wurde in Studien I, II, III und IV die Dauerbewertung zusätzlich zur Dauerschätzung gemessen. Diese wurde ab Studie II durch das Konstrukt Zeitflussbeurteilung ergänzt.

Dauern bewerten

Dauerbewertungen beschreiben eine abschließende Evaluation einer Dauer, nachdem sie erlebt wurde. Laut Liikkanen und Gómez Gómez (2013) werden sie durch das Erleben der Zeit, Erwartungen, Vorerfahrungen und den Zustand des Nutzers beeinflusst. In den Studien I und II wurde gezeigt, dass die Dauerbewertung nicht durch die gleichen Faktoren beeinflusst wird wie die tatsächliche Dauer, wohl aber wie die geschätzte Dauer. Verzerrungen in der Dauerschätzung spiegeln sich also auch in der Dauerbewertung wider. Dies stützt die Aussage von Liikkanen und Gómez Gómez (2013) bezüglich der Bedeutung von erlebter Zeit für die Entstehung von Dauerbewertungen.

Die Befunde der Studien I, II, III und IV weisen darüber hinaus darauf hin, dass die Dauerbewertung umso negativer ausfällt, je länger die geschätzte Dauer ist. Dieser Schluss basiert zum einen auf mittleren bis starken Korrelationen (siehe Tab. 10.1). Zum anderen konnte in Studie IV gezeigt werden, dass der Effekt der tatsächlichen Dauer auf die Dauerbewertung zumindest teilweise durch die Dauerschätzung mediiert wird. Andere Autoren haben solche Zusammenhänge bereits zwischen Dauerschätzung und Zeitflussbeurteilung (PoTJ) berichtet (Droit-Volet et al., 2017, 2018; Sucala et al., 2011), welche eine andere Operationalisierung von Zeitempfinden darstellt. Das hier präsentierte Maß Dauerbewertungen passt insofern zur bestehenden Befundlage. Neben der Bedeutung von erlebter Zeit für die Dauerbewertung, deuten die Ergebnisse von Studie IV jedoch auf weitere Konstrukte hin, die zwischen der Dauer einer Interaktion und ihrer Bewertung durch den Nutzer mediierten, da hier nur eine partielle Mediation zwischen Dauer und Dauerbewertung über die Dauerschätzung vorlag (Baron & Kenny, 1986). Ein möglicher Mediator stellt die Valenz dar, die jedoch lediglich als sehr schwacher Mediator interpretiert werden konnte. Die Valenz scheint viel mehr einen eigenen Effekt auf die Dauerbewertung haben, der nicht auf die Interaktionsdauer zurückzuführen ist.

Zusammengefasst stützen die Befunde die Aussage von Liikkanen und Gómez Gómez (2013) bezüglich der Bedeutung des Erlebens der Zeit (Dauerschätzungen) und des Zustands des Nutzers (Valenz) bei der Bildung von Dauerbewertungen. Bezüglich der Bedeutung von Erwartungen und Vorerfahrungen kann in dieser Arbeit keine Aussage getroffen werden, da diese nicht explizit untersucht wurden. Diese Konstrukte stellen aber mögliche Mediatoren für den Effekt der tatsächlichen Dauer auf die Bewertung dieser dar. Hier werden noch weitere Studien benötigt, die neben den Erwartungen und Vorerfahrungen des Nutzers auch Aufgabencharakteristika und ihren Effekt auf Dauerbewertungen untersuchen sollten. Die vorliegenden Studien leisten jedoch einen ersten Beitrag zur systematischen Untersuchung der Entstehung von Dauerbewertungen und stellen durch die Operationalisierung durch zwei Items eine effiziente Art der Messung vor.

Zeitfluss beurteilen

Ein anderes Empfinden über Zeit beschreibt das Ziehen oder Verfliegen von Zeit (Passage of Time Judgement (PoTJ), Wearden et al., 2014). Diese Zeitflussbeurteilung zeigt Ähnlichkeiten zur Dauerbewertung, ist aber nicht deckungsgleich mit ihr. So lagen die Innersubjektkorrelationen zwischen den beiden Konstrukten in Studien II und III im mittleren Bereich: Je schneller die Zeit verflog, umso besser wurde die Dauer bewertet (siehe Tab. 10.1). Darüber hinaus wurde die Zeitflussbeurteilung in Studie II genau wie die Dauerbewertung stark von der Usability des Systems beeinflusst: Bei niedriger Usability fielen Dauerbewertung und Zeitflussbeurteilung negativer aus. Beide Befunde stützen die Argumentation, dass die Konstrukte Dauerbewertung und Zeitflussbeurteilung Ähnlichkeiten zueinander haben und unter dem gleichen Oberbegriff Zeitempfinden zusammengefasst werden können, sie zeigen jedoch auch, dass eine Unterscheidung sinnvoll ist, da sie nicht exakt das Gleiche abbilden.

Ähnlich zur Korrelation zwischen Dauerbewertung und Dauerschätzung liegt auch Evidenz für eine Korrelation zwischen Zeitflussbeurteilung und Dauerschätzung vor: Je kürzer die Dauer geschätzt wird, umso schneller scheint die Zeit zu verfliegen (siehe Tab. 10.1). Dieser Zusammenhang stimmt mit Befunden anderer Autoren zu mehrminütigen Intervallen in Richtung (Droit-Volet et al., 2017, 2018; Sucala et al., 2011) und Stärke (Sucala et al., 2011) überein. Der Zusammenhang war jedoch schwächer ausgeprägt als die Korrelationen zwischen Dauerschätzung und Dauerbewertung. Darüber hinaus zeigte sich in Studie II ein sehr ähnliches Befundmuster in den Effekten der unabhängigen Variablen auf die Dauerschätzung und auf die Dauerbewertung (Interaktion zwischen lautem Denken und Usability). Die Zeitflussbeurteilung wurde hingegen nur von der Usability signifikant beeinflusst. Daraus lässt sich schließen, dass die Zeitflussbeurteilung nicht in gleicher Weise auf Dauerschätzungen beruht wie die Dauerbewertung. Dies kann auch durch die Ergebnisse von Studie IV gestützt werden, da hier zwar ein starker Zusammenhang zwischen Dauerschätzung und Dauerbewertung gefunden wurde, es jedoch keine Evidenz für einen Zusammenhang zwischen Dauerschätzungen und Zeitflussbeurteilung gab.

Basierend auf diesen Befunden muss eine klare Abgrenzung zwischen der

Schätzung einer Dauer und der Zeitflussbeurteilung gezogen werden. In manchen Situationen kovariieren die beiden Maße zwar miteinander, dies kann aber nicht in jeder Situation erwartet werden. Tanaka und Yotsumoto (2017) argumentieren, dass die Zeitflussbeurteilung durch eine Diskrepanz zwischen erlebter und erwarteter Zeit entsteht. Legt man diese Argumentation den Ergebnissen von Studie IV zugrunde, müssen die Erwartungen der Probanden an die Dauer dafür gesorgt haben, dass es keinen Effekt der Dauer auf die Zeitflussbeurteilung gab. Da die beiden Durchgänge von drei bzw. sechs Minuten relativ adäquat eingeschätzt wurden, muss die erwartete Dauer an die erlebte Dauer angepasst worden sein, damit eine konstante Diskrepanz zwischen erlebter und erwarteter Dauer entstehen konnte. Tatsächlich hatten die Probanden in Studie IV zu wenige Informationen, um eine Erwartung über die Dauer auszubilden, da weder in den Instruktionen noch innerhalb der Aufgabenbearbeitung Hinweise auf die Dauer des jeweiligen Durchganges gegeben wurden. Daher ist es denkbar, dass die Probanden überhaupt keine präzise Erwartung an die Dauer hatten, sondern sie kontinuierlich an das Erlebte anpassten. Zusammengefasst gibt es zwar Situationen, in denen erwartete Dauern explizit vorhanden sind und einen starken Einfluss auf die Zeitflussbeurteilung haben (Sackett et al., 2010; Tanaka & Yotsumoto, 2017), in der untersuchten Nutzungssituation in Studie IV ist es jedoch wenig sinnvoll die Zeitflussbeurteilung als Diskrepanz zwischen erwarteter und erlebter Zeit zu definieren, da hier unklar ist, ob zeitliche Erwartungen überhaupt vorhanden oder bewusst waren. Daher ist der Ansatz von Tanaka und Yotsumoto (2017) zur Erklärung der Entstehung von Zeitflussbeurteilungen nicht in jedem Kontext sinnvoll einsetzbar.

Einen alternativen Ansatz lieferten Droit-Volet et al. (2018). Sie bieten zwar keine explizite Erklärung, wie die Zeitflussbeurteilung entsteht, beschreiben es aber als etwas, was zu jedem Zeitpunkt von Probanden auf Nachfrage hin berichtet werden kann, aber nur in manchen Situationen bewusst wird (Droit-Volet et al., 2018). Eine solche Situation entsteht, wenn mehrminütige Intervalle geschätzt werden sollen. Die Zeitflussbeurteilung kann während solcher Schätzaufgaben über die Zeit hinweg salienter werden, wodurch die Zeit langsamer zu vergehen scheint und Dauerschätzungen länger werden. Bezogen auf Studie IV kann jedoch nicht abgeschätzt werden, wie präsent der aktuelle Zeit-

fluss im Bewusstsein der Probanden war. Geht man jedoch davon aus, dass es in beiden Versuchsdurchgängen vergleichbar war, so kann dies erklären, warum die Zeitflussbeurteilung über die Versuchsdurchgänge hinweg keine signifikanten Veränderungen erfuhr. Tatsächlich unterschieden sich die beiden Durchgänge bis auf die Dauer kaum voneinander, da die Aufgabe exakt die gleiche war. Dieser Ansatz kann die ähnlichen Angaben zur Zeitflussbeurteilung erklären, ohne Annahmen über die Dauererwartungen zu machen.

Zusammengefasst zeigen die Ergebnisse von Studie IV auf, dass Zeitflussbeurteilungen nicht zwangsläufig von der Dauer beeinflusst werden müssen, wie es der Ansatz von Tanaka und Yotsumoto (2017) nahelegt. Wovon eine Zeitflussbeurteilung jedoch abhängt bzw. beeinflusst wird und wie sie entsteht, kann auf Basis der hier vorliegenden Studien nicht abschließend beantwortet werden. Darüber hinaus sind die Ergebnisse von Studie IV bezüglich der Zeitflussbeurteilung ohne Replikation mit einer größeren Stichprobe nicht ausreichend, um verlässliche Schlüsse ziehen zu können. Weiterführende Studien könnten sich neben einer Replikation auch auf das Testen der eben diskutierten Ansätze konzentrieren. Dies kann zum einen durch die Manipulation der erwarteten Zeit passieren (Sackett et al., 2010; Tanaka & Yotsumoto, 2017). Zum anderen könnte auch das Argument von Droit-Volet et al. (2018) bezüglich einem verlangsamten Zeitflussempfinden bei erhöhtem Bewusstsein des Zeitflusses getestet werden. Hier wären beispielsweise Studien denkbar, die die Menge an zugewiesener Aufmerksamkeit auf den Zeitfluss experimentell variieren. Methodische Ansätze hierfür lassen sich aus der Forschung zu Dual-Tasks entleihen, in denen Probanden instruiert werden eine bestimmte Menge ihrer Aufmerksamkeit (z.B. 70%) auf eine Aufgabe zu richten (z.B. Wickens & Gopher, 1977).

Fazit zum Zeitempfinden

Das Zeitempfinden eines Nutzers während einer Mensch-Computer-Interaktion kann in Bezug auf eine abschließende Bewertung der Gesamtdauer und des Zeitflusses bewertet werden. Die vorliegenden Ergebnisse deuten darauf hin, dass die beiden Konzeptionalisierungen von Zeitempfinden, Dauerbewertungen und Zeitflussbeurteilung, während typischer Interaktionen mit Webseiten miteinander kovariieren, aber nicht deckungsgleich sind. Darüber hinaus zeigte

sich in den ersten drei Studien der erwartete Zusammenhang zur tatsächlichen bzw. geschätzten Dauer. Dies stützt die Annahme, dass Zeitempfinden auf der Wahrnehmung von Dauern beruht, wie sie beispielsweise von „timing-with-a-timer“-Modelle auf kognitiver Ebene beschrieben werden (Droit-Volet et al., 2018). Die Unterschiede der Befunde zu Studie IV bezüglich des Zeitflusses weisen zudem darauf hin, dass die Bewertung von sich ziehender oder verfliegender Zeit auf weiteren kognitiven Prozessen beruht, die im Rahmen dieser Arbeit nicht untersucht wurden.

10.2 Dauerwahrnehmung und Nutzererleben

Das zweite Ziel der vorliegenden Arbeit war die Beantwortung der Frage, wie sich die Wahrnehmung von Interaktionsdauern in das Nutzererleben während einer Mensch-Computer-Interaktion integriert. Hintergrund für diese Frage bildet das CUE-Modell (Mahlke, 2008; Minge & Thüring, 2018; Thüring & Mahlke, 2007). Im CUE-Modell wird postuliert, dass auf Basis des Systems, des Nutzers und des Kontextes spezifische Interaktionscharakteristika entstehen, die als Grundlage für die Entstehung der Komponenten des Nutzererlebens fungieren. Diese Interaktionscharakteristika werden im Rahmen des CUE-Modells nicht näher spezifiziert. Die Interaktionsdauer kann jedoch als ein solches Charakteristikum angesehen werden. Das zweite Ziel der Arbeit war zum einen den Einfluss von Interaktionsdauern auf das Nutzererleben besser zu verstehen, um verlässliche und fundierte Vorhersagen über die Bedeutung von Dauern und Dauerwahrnehmung während einer Mensch-Computer-Interaktion machen zu können. Zum anderen sollte anhand der Interaktionsdauer beispielhaft für ein Interaktionscharakteristikum gezeigt werden, wie solche Charakteristika experimentell untersucht werden können. Dies kann für eine Spezifizierung des CUE-Modells genutzt werden. Dazu wurden in den Studien I bis III Innersubjektkorrelationen zwischen zeitbezogenen Variablen der Dauerwahrnehmung und UX-bezogenen Variablen im Rahmen typischer Interaktionen mit Webseiten untersucht. In Studie IV wurde zusätzlich versucht abzuschätzen, wie groß der Effekt des Nutzererlebens auf das Zeitempfinden ist, wenn für die Dauer kontrolliert wird.

10.2.1 Zusammenhänge mit den Komponenten des Nutzererlebens

Im CUE-Modell werden verschiedene Komponenten des Erlebens einer Mensch-Computer-Interaktion beschrieben, und zwar die Wahrnehmung instrumenteller Qualitäten, die Wahrnehmung nicht-instrumenteller Qualitäten und Emotionen. In den vorliegenden Studien wurden diese Komponenten jeweils über ein Konstrukt abgebildet: die wahrgenommene Usability, die wahrgenommene Attraktivität und die erlebte Valenz der Versuchsperson.

In den ersten drei Studien zeigten sich substantielle Zusammenhänge zwischen den Variablen der Dauerwahrnehmung und diesen UX-bezogenen Variablen (Tab. 10.1). Im Großen und Ganzen können die Zusammenhänge so zusammengefasst werden, dass ein positiveres Nutzererleben einhergeht mit kürzeren Aufgabendauern, als kürzer geschätzten Aufgabendauern, einer besseren Bewertung der Dauer und einer schneller verfließenden Zeit. Diese Zusammenhänge zeigte sich in unterschiedlicher Stärke für die verschiedenen Komponenten des Nutzererlebens: Für die wahrgenommene Attraktivität waren die Korrelationen am schwächsten und für die Valenz am stärksten. In Studie III trat dies am deutlichsten zu Tage (siehe Tab 10.1). Neben den starken Zusammenhängen zur Valenz zeigte Studie III zudem, dass die wahrgenommene Usability vor allem mit der Dauerbewertung sehr stark korrelierte. Daher kann geschlossen werden, dass die Dauerbewertung einen Teil der wahrgenommenen instrumentellen Qualitäten einer Interaktion bildet. Ähnliches gilt für Zeitflussbeurteilungen, hier war der Zusammenhang zur wahrgenommenen Usability jedoch schwächer.

Die ersten drei Studien erlauben nur eine korrelative Betrachtung der Zusammenhänge. In Studie IV wurde daher geprüft, in wie weit das Nutzererleben die Dauerbewertung und die Zeitflussbeurteilung vorhersagen können, wenn die Dauer der Interaktion experimentell kontrolliert wird. In dieser Studie konnten die Zusammenhänge zwischen Dauerbewertung bzw. Zeitflussbeurteilung zu den UX-bezogenen Variablen nicht repliziert werden. Wie vorher bereits diskutiert, wurde in Studie IV jedoch eine Aufgabe genutzt, in der Zeit kaum eine Rolle spielte und in der das Ende der Aufgabe nicht durch die Probanden kontrolliert wurde. Hierdurch entstand ein großer Unterschied zu Studien I

bis III. Die Ergebnisse zeigen daher auf, dass der Zusammenhang zwischen Zeitempfinden und UX nicht auf alle Kontexte übertragbar ist.

Zusammengefasst zeigen die Ergebnisse, dass die Dauer einer Mensch-Computer-Interaktion von Nutzern nicht nur wahrgenommen, sondern auch in Form eines Zeitempfindens in das Gesamterleben der Interaktion eingebettet wird. Dies zeigt sich primär durch korrelativen Zusammenhänge. Bezüglich des CUE-Modells wird auf Basis der Ergebnisse die Hypothese aufgestellt, dass Probanden ihr Zeitempfinden als eine instrumentelle Qualität der Interaktion wahrnehmen. Wie im Theorieteil jedoch bereits diskutiert, korrelieren die Komponenten des Nutzererlebens mittel bis stark miteinander (z.B. Aranyi & van Schaik, 2015; Lavie & Tractinsky, 2004). Daher ist es nicht verwunderlich, dass das Zeitempfinden zwar der instrumentellen Komponente zugewiesen werden kann, trotzdem aber mittlere bis starke Korrelationen zur Emotion und teilweise auch zu Konstrukten der wahrgenommenen nicht-instrumentellen Qualitäten zeigen.

10.2.2 Spezifizierung des Begriffs Interaktionscharakteristika

Doch was heißt dies nun für Interaktionscharakteristika im Allgemeinen? Zunächst zeigen Studien I und II, dass eine systematische Variation des Systems und des Kontextes, hier über die Gebrauchstauglichkeit des Systems und über eine Zweitaufgabe, und eine unsystematische Variation des Nutzers, hier über das zufällige Ziehen von Probanden, zu einer individuellen Ausprägung des Interaktionscharakteristikums Dauer führt. Daher stützen die Ergebnisse zumindest für dieses Charakteristikum die Annahme des CUE-Modells, dass System, Nutzer und Kontext die Interaktionscharakteristika beeinflussen. Darüber hinaus zeigen Studien I bis III, dass die Wahrnehmung des Charakteristikums Dauer mit anderen Komponenten des Nutzererlebens kovariieren, was den im CUE-Modell angenommenen Einfluss der Interaktionscharakteristika auf die Komponenten stützt.

Im Gegensatz dazu lieferte Studie IV für diesen zweiten Teil der Argumentation keine Evidenz. Hier wurde die Dauer vom Einfluss des Systems getrennt, woraufhin es keine substantiellen Zusammenhänge zwischen Dauer-

wahrnehmung und Nutzererleben gab. Einzig die Valenz war ein signifikanter Prädiktor der Dauerbewertung. Wie oben aber bereits diskutiert, war die Aufgabendauer in Studie IV nicht aufgabenrelevant, da die Probanden keinen Einfluss auf sie nehmen konnten und die Aufgabe unabhängig von der Dauer bearbeitet wurde. Daher sprechen die Ergebnisse dafür, dass die Dauer einer Interaktion nur dann eine Bedeutung für das Entstehen von Nutzererleben hat, wenn sie essentieller Bestandteil der Aufgabe ist.

Zusammengefasst zeigt die vorliegende Arbeit, dass einzelne Interaktionscharakteristiken untersucht werden können. Diese Untersuchungen bilden eine wichtige Grundlage für das Verständnis, wie ein spezifisches Nutzererleben entsteht. Solch ein Verständnis kann im Idealfall helfen, einfache Kästchenmodelle zu übersteigen und zu einem computationalen Modell führen, welches eine mathematische Modellierung und Vorhersage von einzelnen Komponenten des Nutzererlebens auf Basis der Interaktionscharakteristika erlaubt.

Hiervon ist die vorliegende Arbeit jedoch noch weit entfernt. Da nur ein Charakteristikum beispielhaft betrachtet wurde, kann kein Schluss auf Basis der präsentierten Daten gezogen werden, was weitere Charakteristiken sind. Dies scheint jedoch auch eine konzeptionelle Frage zu sein. Um dieser gerecht zu werden, werden Interaktionscharakteristika als objektiv messbare Größen definiert. Beispiele hierfür sind die Gesamtdauer, wie sie hier untersucht wurde, oder die Anzahl und Dauer systeminhärenter Latenzen (Thüring, 2013). Andere Beispiele sind die Häufigkeit von Fehlbedienungen oder die Anzahl und Dauer von Suchprozessen nach einem gewünschten Interaktionselement. Solche Interaktionscharakteristika können durch Beobachtung der Nutzungssituation oder durch die Messung anderer behavioraler Maße (z.B. Blickbewegungen) während der Interaktion messbar gemacht werden. Auch diese Interaktionscharakteristika beeinflussen vermutlich eher die Wahrnehmung instrumenteller Qualitäten.

Es sind jedoch auch Interaktionscharakteristika denkbar, die einen starken Einfluss auf nicht-instrumentelle Qualitäten haben. Hierzu könnte die Anzahl der als typischerweise ästhetisch empfundenen Elemente gehören, wie ähnliche Seitenverhältnisse von Objekten, Regelmäßigkeit der Abstände oder Symmetrie (Ngo, 2001). Auch die Neuheit des Produkts könnte ein Interaktionscharakteristikum sein, welches durch den Nutzer wahrgenommen und gewichtet wird,

wodurch wiederum die Wahrnehmung des eigenen Statuszuwachses durch die Nutzung entsteht. Eine systematische Variation dieser verschiedenen Kandidaten von Interaktionscharakteristiken könnte als Grundlage für eine Weiterentwicklung des CUE-Modells dienen. Hierzu liefert die vorliegende Arbeit einen Baustein.

10.3 Anwendungsaspekte der Ergebnisse

Auf Basis der vorgestellten Ergebnisse werden im Folgenden verschiedene Konsequenzen für den Anwendungskontext Mensch-Computer-Interaktionen gezogen. Diese basieren zum einen auf bereits diskutierten Ergebnissen, wie dem Zusammenhang zwischen Dauerbewertung und wahrgenommener Usability oder dem Effekt von lautem Denken auf Verzerrungen der Schätzung. Zum anderen gibt es Konsequenzen, die auf Basis noch nicht diskutierter Ergebnisse gezogen werden können, wie dem Zusammenhang zwischen Dauerwahrnehmung und mentaler Beanspruchung.

10.3.1 Dauerbewertung als indirektes Maß für die wahrgenommene Usability

In drei von vier Studien wurden mittlere bis starke Korrelationen zwischen der Dauerbewertung und der wahrgenommenen Usability gefunden. Die beiden Konstrukte scheinen also nicht deckungsgleich zu sein. Sie weisen in typischen Interaktionen mit einer Webseite aber einen konstanten Zusammenhang auf. Daher wird die Annahme aufgestellt, dass bei Interaktionen mit Webseiten die Dauerbewertung als indirektes Maß für Usability genutzt werden kann. Diese Annahme wird zudem durch Fragebögen unterstützt, in denen Items der Dauerbewertung zur Erfassung der Usability naher Konstrukte genutzt werden, wie der Effizienz oder Erlernbarkeit des Systems (z.B. PSSUQ: Lewis, 1992; QUIS: Chin et al., 1988; UEQ: Laugwitz et al., 2006, USE: Lund, 2001).

Ein indirektes Maß für wahrgenommene Usability scheint auf den ersten Blick nicht notwendig, da Usability relativ einfach und kostengünstig mit einem bis wenigen Items reliabel abgefragt werden kann (Single Item siehe z.B. Minge & Thüring, 2018). Dies geschah auch in der vorliegenden Arbeit in Stu-

dien I, II und IV. Es gibt jedoch Untersuchungssettings, in denen das Untersuchungsziel nicht offenkundig ersichtlich sein soll, um zum Beispiel Antworttendenzen zu reduzieren (Moosbrugger & Kelava, 2012). In solchen Situationen kann es von Vorteil sein, wenn die wahrgenommene Usability nicht mit Items zur Gebrauchstauglichkeit oder der Einfachheit der Nutzung erfragt wird, sondern mit allgemeineren Attributen. Hier können Dauerbewertungen hilfreich sein, da sie lediglich eine Bewertung der Dauer bezüglich den Dimensionen kurz und lang bzw. der Akzeptabilität erfragen.

Ein anderer Anwendungskontext für ein solches indirektes Maß könnte in der Forschung mit Probanden liegen, die über ein eingeschränktes Vokabular verfügen. Beispiele hierfür können die Forschung mit Kindern oder mit Menschen mit Behinderung sein, aber auch Befragungen auf Sprachen, die von den Probanden gerade erst erlernt werden. In solchen Kontexten ist es wichtig, eine einfache Sprache zu nutzen (Moosbrugger & Kelava, 2012). Hier kann es leichter fallen, anzugeben, ob eine Bearbeitungsdauer eher kurz oder lang war, als anzugeben, ob das System während der Interaktion als gebrauchstauglich erlebt wurde oder nicht.

Für beide Anwendungskontexte muss jedoch zunächst untersucht werden, in wie weit die Annahme stimmt, dass Dauerbewertungen als indirektes Maß für Usability nutzbar sind. Da die Annahme lediglich auf korrelativen Zusammenhängen basiert, werden experimentelle Arbeiten benötigt, die dies weiter prüfen. Darüber hinaus kann auf Basis der vorliegenden Studien nur davon ausgegangen werden, dass ein solches indirektes Maß für Interaktionen mit Webseiten sinnvoll ist. In wie weit dies jedoch weiter verallgemeinerbar ist, kann nur mit weiterer Forschung beantwortet werden.

10.3.2 Lautes Denken in Usability- und UX-Testszenarien

Auf Basis der Befunde der Studien I und II kann auch die Frage weiter diskutiert werden, in wie weit lautes Denken die Interaktion und das Erleben der Probanden verändert. Hertzum und Holmegaard (2015) zeigten vergleichbare Ergebnisse zu den hier vorliegenden Befunden: Probanden, die während einer Knobelaufgabe laut dachten, überschätzen die Aufgabendauer weniger stark als solche, die nicht laut dachten. Die Autoren interpretierten dies so, dass zeitbasierte Dimensionen nicht zusammen mit lautem Denken gemessen wer-

den sollten, da lautes Denken zu einer Veränderung im Dauererleben führen kann. Sie folgern, dass das Erleben einer Dauer andere Erlebenskomponenten beeinflussen kann, was zu einer Verzerrung führt. Die Ergebnisse der Studien I und II stützen zwar den ersten Teil der Argumentation, nicht aber den zweiten, da sich lautes Denken in keiner der beiden Studien auf die verschiedenen UX-Komponenten auswirkte.

Darüber hinaus untersuchten Hertzum und Holmegaard (2015) lediglich Dauerschätzungen und kein Zeitempfinden, welches enger mit Erlebenskomponenten verbunden ist. Bezüglich Zeitempfinden zeigte sich weder in Studie I noch in Studie II ein Effekt des lauten Denkens, obwohl die tatsächliche Dauer mit lautem Denken höher war. Wie vorher bereits diskutiert, kann dies auf eine Verzerrung in der Dauerschätzung zurückgeführt werden, was die Warnung von Hertzum und Holmegaard (2015) stützen würde. Die Ergebnisse können aber auch dafür sprechen, dass Nutzer lautes Denken als eine Zweitaufgabe verstehen, die sie in der Bewertung von Dauern berücksichtigen. Auch hier sind also Folgestudien notwendig, um die beiden Erklärungsansätze gegeneinander abwägen zu können, in dem zum Beispiel erwartete Dauern erhoben werden. Basierend auf dieser Abwägung könnte die Aussage von Hertzum und Holmegaard (2015) weiter gestützt oder abgeschwächt werden, dass lautes Denken nicht mit zeitbasierten Maßen gemessen werden sollte.

In Einklang mit den Befunden von Hertzum und Holmegaard (2015), die Level 1 bis 2 der Verbalisierung (Levels siehe Hertzum et al., 2009) nutzten, konnte jedoch in Studie I und II gezeigt werden, dass lautes Denken mit Begründungen der Handlungen (vgl. Level 3 der Verbalisierung) die mentale Beanspruchung der Nutzer erhöht. Forscher müssen sich bei der Nutzung von lautem Denken dieser zusätzlichen Belastung der Nutzer bewusst sein, da sie unter Umständen die Interaktion beeinflussen oder behindern kann.

10.3.3 Kognitive Prozesse während einer Mensch-Computer-Interaktion durch Interferenzen besser verstehen

In Studien I und II wurde zudem die Frage aufgeworfen, ob eine verminderte Usability nicht auch immer zu einer Erhöhung der mentalen Beanspruchung

führt. Diese Annahme basiert auf theoretischen Überlegungen: So definiert beispielsweise die DIN, EN ISO 9241-11 (2018) Usability so, dass sie eine Zielerreichung mit einer angemessenen mentalen Beanspruchung erlaubt. Auch in der Norm zur Dialoggestaltung (DIN, EN ISO 9241-110, 2008) werden unnötige Informationen in der Dialogführung, die ein Anzeichen für niedrige Usability darstellen, als eine Quelle für unnötige mentale Beanspruchungen genannt.

Die Befunde der Studien I, II und III liefern jedoch Indizien, dass eine verminderte Usability nicht zwangsläufig mit einer erhöhten mentalen Belastung einhergeht. So zeigte sich keine Evidenz für einen Effekt der objektiven Usability auf die Verzerrungen der Schätzungen in Studien I und II und kein übereinstimmendes Muster der Beeinflussung der Aufgaben auf die wahrgenommene Usability und die Verzerrungen in Studie III. Nimmt man an, dass eine Veränderung der mentalen Beanspruchung immer zu einer Verzerrung der Dauerschätzung führt, dann deuten diese Ergebnisse an, dass die verminderte Usability nicht die mentale Beanspruchung erhöht hat.

Eine alternative Erklärung könnte aber auch sein, dass mentale Beanspruchung, die durch niedrige Usability hervorgerufen wird, nicht gleichzusetzen ist mit mentaler Beanspruchung, die durch lautes Denken Level 3 entsteht. So können andere Arten von kognitiven Ressourcen (Wickens, 2002, 2008) und eine aktivere Verarbeitung (vgl. Meta-Analyse Block et al., 2010) beim lauten Denken benötigt werden, wodurch es durch lautes Denken, nicht aber im Falle von niedriger Usability, zu einer Interferenz mit der Schätzaufgabe kommt. Grundlagenarbeiten zu Verzerrungen bei Dauerschätzungen zeigen zudem, dass vor allem solche Zweitaufgaben zu Interferenzen führen, die exekutive Funktionen beanspruchen (Brown, 2008). Dies scheint bei niedriger Usability nicht oder nur im geringen Maße der Fall zu sein.

Zusammengefasst erlaubt die Interferenz zwischen Aufgaben auch im Kontext von Mensch-Computer-Interaktionen einen Rückschluss auf kognitive Prozesse zu ziehen. In der hier vorliegenden Arbeit wurden Interferenzen zwischen Dauerschätzung und Zweitaufgabe verglichen (Bearbeitung einer Aufgabe auf einer Webseite mit vs. ohne zusätzliches Verbalisieren). Dauerschätzungen wurden dabei als Indikator für die mentale Beanspruchung gesehen. Da der Zusammenhang zwischen Dauerschätzungen und mentaler Beanspruchung bereits sehr gut untersucht ist (Reviews siehe z.B. Brown, 2008; Grondin, 2010;

Matthews & Meck, 2016) und auch die beteiligten Prozesse bei einer Dauerschätzung ausführlich erforscht wurden (z.B. zu Exekutivfunktionen, Brown, 2008), bieten sich Dauerschätzungen zur Untersuchung der Interferenzen mit anderen Aufgaben an. Dies erfolgte in der vorliegenden Arbeit jedoch nur für lautes Denken, da ein anderes Forschungsziel im Fokus stand. Für das integrative Verstehen kognitiver Prozesse bei einer Interaktion mit einem Computer bieten solche Interferenzstudien jedoch eine vielversprechende Methodik.

10.4 Limitationen

Die vorliegende Arbeit ist nicht ohne Limitationen. Diese sind teilweise inhaltlicher Natur und schränken die Anwendungskontexte ein, für die die diskutierten Ergebnisse und Schlüsse gelten. Es gibt jedoch auch Limitationen für die Verallgemeinbarkeit der Ergebnisse, die auf Basis methodischer Entscheidungen im Design beachtet werden müssen.

10.4.1 Inhaltliche Limitationen

Die wohl größte inhaltliche Limitation liegt in den untersuchten Szenarien und abhängigen Maßen. So wurden in den ersten drei Studien nur Interaktionen mit Webseiten betrachtet. Andere Mensch-Computer-Interaktionen, wie beispielsweise die Bedienung eines Textverarbeitungsprogramms oder eines Betriebssystems waren nicht teil der gewählten Szenarien. Auch ganz andere Interaktionsarten, wie beispielsweise Spiele oder Interaktionen mit Mobilgeräten wurden nicht untersucht. Die Aufgabe in Studie IV unterschied sich zwar stark von den anderen Aufgaben, sie stellte gleichzeitig aber eine so abstrakte Aufgabe dar, dass sie kaum auf einen Anwendungskontext übertragen werden kann. Daher ist der untersuchte Anwendungskontext relativ klein im Vergleich zu dem möglichen Spektrum an Anwendungsbereichen und schränkt die Verallgemeinbarkeit der Ergebnisse ein.

Auf Seiten der abhängigen Maße müssen Beschränkungen vor allem für die Erfassung des Nutzererlebens erwähnt werden. Nutzererleben ist wie eingangs beschrieben ein sehr breit aufgestelltes Konstrukt (vgl. Definition nach DIN, EN ISO 9241-11, 2018, siehe Kap. 2, S. 7). In der vorliegenden Arbeit wurde jedoch ein reduktionistisches Modell von Nutzererleben als Rahmenmodell

genutzt (CUE-Modell). Darüber hinaus wurden die UX-Komponenten durch jeweils nur ein Maß abgebildet. Wie weiter oben bereits diskutiert, stellt dies eine starke Vereinfachung der Komponenten dar. Auch hier wäre eine breitere Aufstellung für eine weitere Generalisierbarkeit und Prüfung der gezogenen Schlüsse wünschenswert. Auf der anderen Seite gab es in der Studie eine Vielzahl an abhängigen Maßen. Eine breitere Abbildung von UX hätte auch eine noch größere Zahl an abhängigen Maßen bedeutet, was die Komplexität der Versuche erhöht und die Interpretation der Ergebnisse erschwert hätte.

Auch auf Seiten der zeitbezogenen Maße erfolgte eine starke theoretische Einschränkung: Hier wurde das prospektive Paradigma der Dauerwahrnehmung in den Fokus gestellt. Im Rahmen des Dual-Process Contingency Models (Zakay, 1993a) wurde bereits angesprochen, dass es im Kontext der Mensch-Computer-Interaktion zu Mischformen zwischen dem prospektiven und retrospektiven Paradigma kommen kann. Ist die Interaktion zeitweise so beanspruchend, dass keine Aufmerksamkeit mehr auf das Verstreichen von Zeit gerichtet werden kann, müssen diese Episoden durch eine retrospektive Schätzung ergänzt werden. Solche gemischten Schätzungen sind bis jetzt jedoch wenig untersucht worden.

In kognitiven Modellen der retrospektiven Dauerschätzung werden teilweise gegensätzliche Vorhersagen für die Auswirkung bestimmter Faktoren auf die Dauerschätzung im Vergleich zu Modellen im prospektiven Paradigma gemacht. Ein Einflussfaktor, für den die Vorhersagen gegenläufig sind, ist die Anzahl der Aufgabenwechsel (*task switching*) während des zu schätzenden Intervalls (Zakay & Block, 2004). Im retrospektiven Paradigma sollte nach dem *Contextual-Change-Modell* ein häufiger Wechsel zu einer höheren Anzahl an enkodierten kontextuellen Veränderungen führen. Vergleichbar zu den kritischen Ereignissen (HPEs) aus dem Dual-Process Contingency Model (Zakay, 1993a) umfassen solche kontextuellen Veränderungen externe und interne Aspekte, wie das Laden einer neuen Webseite oder eine Veränderung des emotionalen Zustands. Bei der Schätzung werden diese Veränderungen abgerufen (dekodiert) und für die Rekonstruktion der Dauer genutzt. Dadurch entstehen längere retrospektive Schätzungen, wenn mehr kontextuelle Veränderungen abgerufen werden können (Zakay & Block, 2004). Im prospektiven Paradigma hingegen sollte laut dem AGM ein häufiges Wechseln zwischen Aufgaben mehr Aufmerk-

samkeit von der Zeitschätzaufgabe abziehen, wodurch die prospektive Schätzung sich verkürzt. Zakay und Block (2004) berichten empirische Ergebnisse, die diese Vorhersagen stützen: Wenn Probanden eine Dauer von 12 Sek. retrospektiv schätzten, sind diese Schätzungen länger bei mehr erlebten Aufgabenwechseln im Vergleich zu wenigen Wechseln. Bei prospektiven Schätzungen ist es umgekehrt.

Eine ähnliche Logik kann auch auf die Ergebnisse der Studien I und II angewandt werden. Auch hier mussten die Probanden zwischen Zeitschätzaufgabe und Durchführung der Interaktion immer wieder wechseln, um beide Aufgaben parallel zu erfüllen. In der Gruppe mit lautem Denken kam hier noch eine weitere Aufgabe und damit mehr Wechsel zwischen Aufgaben hinzu. Die Ergebnisse der Studien zeigen, dass sich die Dauerschätzungen mit lautem Denken (mehr Wechsel) gegenüber den Dauern ohne lautes Denken (weniger Wechsel) verkürzen. Dies spricht dafür, dass Probanden tatsächlich prospektiv die Dauer schätzten. Es kann jedoch nicht ausgeschlossen werden, dass einzelne Episoden der geschätzten Dauer retrospektiv wahrgenommen und ergänzt wurden. Auf Grund der Gegenläufigkeit des Effekts des Aufgabenwechsels könnte dies zu einer kleineren Differenz zwischen den Gruppen mit und ohne lautem Denken geführt haben als bei einer rein prospektiven Schätzung.

Hier stellt sich jedoch ein grundsätzliches Problem. In keiner der hier vorgestellten Studien wurde kontrolliert, ob während der gesamten Interaktionsdauer Aufmerksamkeit auf das Verstreichen der Zeit gerichtet wurde. Eine solche Kontrolle ist schwer möglich, da sie ein kontinuierliches Maß für die Aufmerksamkeitszuwendung voraussetzt. Gerade hierzu ist die empirische Befundlage jedoch unzureichend, da sich die meisten Arbeiten klar zu einem der beiden Paradigmen zuordnen. Im Rahmen der hier untersuchten Interaktionen ist es jedoch wahrscheinlich, dass Mischformen zwischen den Paradigmen entstanden sind. Diese Mischformen empirisch erfassen zu können, stellt ein wichtiges Forschungsziel für zukünftige Arbeiten dar.

Neben der klaren Positionierung der Arbeit im prospektiven Paradigma mit Abgrenzung zum retrospektiven Paradigma, ist aber auch die Betrachtung und Operationalisierung von Zeit eine Limitation. Wie in der Theorie bereits beschrieben, kann die Wahrnehmung von Dauern nicht nur aus der Perspektive von Dauerschätzungen betrachtet werden (Block, 1990). Auch Ab- oder

Zeitfolgen von Ereignissen oder die zeitliche Perspektive mit Bezug zur Vergangenheit, Gegenwart und Zukunft könnten im Bereich Dauerwahrnehmung diskutiert und untersucht werden. Hierzu gibt es bereits Arbeiten, wie Studien zum Effekt von Regelmäßigkeit von erlebten Dauern auf die Effizienz und das Nutzerleben während einer Mensch-Computer-Interaktion (Thomaschke & Haering, 2014; Weber, Haering & Thomaschke, 2013). Andere Arbeiten propagieren die Berücksichtigung der zeitlichen Perspektive des Nutzers während der Entwicklung eines technischen Systems, um ein positives Nutzererleben zu ermöglichen (Hermosa Perrino & Burmester, 2017). Im Rahmen der vorliegenden Arbeit wurden diese Ansätze allerdings nicht berücksichtigt.

10.4.2 Methodische Limitationen auf Grund des Versuchdesigns

Die vermutlich stärkste methodische Limitation ist die Nutzung nicht validierter Items für die Erfassung des Zeitempfindens. Hier konnte anders als bei den UX-bezogenen Single-Items auch nicht die externe Validität der Skalen geprüft werden, da es für beide Operationalisierungen von Zeitempfinden keine validierten Erfassungsmethoden gibt. Daher kann für beide Maße nur von einer Augenscheinvalidität gesprochen werden (Moosbrugger & Kelava, 2012). Bezüglich der Reliabilität zeigten die Maße zwar eine hohe Konsistenz über die Mehrfachmessung innerhalb jeder Studie und Ähnlichkeiten in den Befunden zwischen den Studien, dies ersetzt jedoch auch keine ausführliche Testung der Reliabilität. Daher kann nicht mit Sicherheit gesagt werden, dass die beiden Konstrukte verlässlich das gemessen haben, was gemessen werden sollte und wofür die Schlussfolgerungen gezogen wurden.

Eine weitere Schwäche in der vorliegenden Arbeit besteht in der Veränderung der selbstentwickelten Maße für Zeitempfinden und UX-bezogene Maße. So wurden Items teilweise ergänzt oder auf Grund von Redundanz weggelassen und die Skalierung der Items in Studie IV verändert. Auch wenn dies keine großen Veränderungen waren, könnten sie teilweise für die Unterschiede in den Befunden zwischen den Studien verantwortlich sein.

Eine weitere Limitation liegt in der erreichten Teststärke ($Power\ 1 - \beta$). In allen Studien wurde eine Vielzahl an verschiedenen Effekten untersucht, zu de-

nen jeweils eine Poweranalyse zur Bestimmung der benötigten Stichprobengröße a priori hätte durchgeführt werden müssen. Hier bestand jedoch ein Problem darin, dass teilweise keine Effektstärke vor der Untersuchung bestimmt werden konnte. Dies lag daran, dass zum Teil selbstentwickeltes Versuchsmaterial in zwei Versionen verwendet wurde, für welches der zu erwartenden Unterschied zwischen den Versionen in seiner Stärke nicht vorher bekannt war. Zum anderen gab es Maße und Konstrukte, die in dieser Form noch nicht untersucht worden waren, wodurch ebenfalls keine empirisch bestimmte Effektstärke vorlag. Ein weiteres Problem lag darin, dass die untersuchten Effekten sehr unterschiedliche Effektstärken hatten, wodurch keine einheitliche Stichprobengröße bestimmt werden konnte: Die gleiche Stichprobengröße kann bei einem großen Effekt zu einem überpowererten und bei einem kleinen Effekt zu einem unterpowererten Test führen (Bortz, 2005).

Daher wurde für alle Studien eine Stichprobengröße gewählt, die zur Entdeckung von mittleren bis großen Effekten eine ausreichende Power von $1 - \beta \leq 80\%$ gewähren sollte. Zusätzlich wurde für jeden Effekt die Power post-hoc anhand des beobachteten Effekts berechnet. Hier zeigte sich, dass die gewählte Stichprobengröße für viele der untersuchten Effekte groß genug war und zu einem überpowererten Test führte. Dies galt vor allem für signifikante Effekte, was nicht verwunderlich ist, da bei post-hoc Poweranalysen die Teststärke äquivalent zum p -Wert ist (Hoenig & Heisey, 2001). Für nicht signifikante Effekte hingegen kann auf Basis der vorliegenden Studien kaum eine Aussage getroffen werden. Durch das Berichten der Kennwerte der Unterschiede und der erreichten Power können die vorliegenden Studien jedoch in Folgeuntersuchungen als Basis für eine a priori Poweranalyse genutzt werden.

Neben der Anzahl der Probanden muss auch die Anzahl der untersuchten Aufgaben als eine Limitation der vorliegenden Studien gesehen werden. In den ersten drei Studien wurden die zu bearbeitenden Aufgaben zwar innerhalb der Probanden variiert, jedoch wurden auf Grund von zeitlichen Begrenzungen nur wenige Aufgaben gewählt. Dies führte zu einer Einschränkung der Generalisierbarkeit. Seitens der Analysen konnte zudem kein Random Intercept über Aufgaben geschätzt werden. Solch ein Random Intercept über die Aufgabe kann jedoch theoretisch interessant sein, da er die Varianz zwischen Aufgaben in der Population aller Aufgaben schätzt. Bei der Betrachtung von mehreren

Random Effects über die Aufgaben könnten darüber hinaus auch die Korrelationsparameter zwischen den Random Effects von Interesse sein. So könnte beispielsweise nicht nur geschätzt werden, wie groß die Varianz zwischen Aufgaben in der Dauer ist (Random Intercept), sondern auch ob Aufgaben, die besonders lange brauchen, auch besonders stark durch die Usability des Systems beeinflusst werden (Korrelation zwischen Random Intercept und Random Slope für Effekt der Usability). Für solche Betrachtungen ist in den vorliegenden Studien jedoch die Aufgabenanzahl zu gering. Hier hätten pro Versuchsperson wesentlich mehr Aufgaben erhoben werden (Bolker, 2019).

10.5 Fazit

Im Kontext von Mensch-Computer-Interaktionen werden Interaktionsdauern häufig als objektiver Indikator für die Effizienz, die Verständlichkeit oder die intuitive Bedienbarkeit eines Systems genutzt. Die Dauer kann jedoch auch als subjektives Konstrukt betrachtet werden, da sie ein Interaktionscharakteristikum darstellt, das vom Nutzer wahrgenommen, bewertet und in das Erleben integriert werden kann. Vor diesem Hintergrund war es das Ziel der vorliegenden Arbeit zu untersuchen, wie Nutzer Interaktionsdauern wahrnehmen und wie diese Wahrnehmung mit anderen Erlebenskomponenten zusammenhängt. Dazu wurde verschiedene Dimensionen von Dauerwahrnehmung in den Fokus gerückt: Dauerschätzungen, Dauerbewertungen und Zeitflussbeurteilungen. Zunächst zeigte sich, dass kognitionspsychologische Modelle und Theorien auch im Bereich Mensch-Computer-Interaktionen gut dazu in der Lage sind, vorherzusagen, wie Nutzer eine Interaktionsdauer einschätzen werden. Darüber hinaus konnte gezeigt werden, dass die Bewertung einer Dauer zu großen Teilen von der geschätzten Dauer beeinflusst wird. Neben der geschätzten Dauer beeinflusst jedoch auch die emotionale Valenz die Bewertung von Dauern. Der Zeitfluss hingegen kovariiert zwar mit den Dauerschätzungen und -bewertungen, zeigte aber auch deutliche Unterschiede zu den beiden Konstrukten. Zusammengefasst kann daher geschlussfolgert werden, dass Nutzer die Dauer von Interaktionen im Großen und Ganzen adäquat wahrnehmen können und Aspekte der Dauer, wie ihre Länge und das Verstreichen der Zeit während der Dauer, unabhängig voneinander bewerten können. Darüber

hinaus konnte gezeigt werden, dass diese Aspekte von Dauerwahrnehmung mit verschiedenen Komponenten des Nutzererlebens kovariieren. Vor allem die Dauerbewertung scheint in typischen Interaktionen mit Webseiten als eine instrumentelle Qualität der Interaktion interpretiert werden zu können.

Die vorliegende Arbeit ist an der Schnittstelle zwischen zwei verschiedenen Forschungsfeldern anzusiedeln, Dauerwahrnehmung und Nutzererleben. Für beide Forschungsfelder konnten Erkenntnisse generiert werden. Für den eher grundlagenorientierten Bereich der Dauerwahrnehmung konnte die Gültigkeit von angenommen kognitiven Prozessen auch im Anwendungskontext gestützt werden. Für den Bereich des Nutzererlebens konnte gezeigt werden, wie durch die Verknüpfung der Konstrukte dieses Bereiches mit grundlagenorientierten Arbeiten eine Präzisierung der betrachteten Prozesse erfolgen kann. Genauer gesagt wurde hier für das Interaktionscharakteristikum Dauer gezeigt, wie es zu einer Wahrnehmung führt, die wiederum in abstraktere Bewertungsformen integriert wird. Solche Präzisierungen der Prozesse können auf lange Sicht eine Verfeinerung der Modelle zum Nutzererleben erlauben, sodass diese in computationale Modelle überführt werden können.

Teil IV

Anhang

Literaturverzeichnis

- Agarwal, A. & Meyer, A. (2009). Beyond usability: evaluating emotional response as an integral part of the user experience. In *Chi'09 extended abstracts on human factors in computing systems* (S. 2919–2930). ACM.
- Albert, W., Tullis, T. & Tedesco, D. (2009). *Beyond the usability lab: Conducting large-scale online user experience studies*. Burlington, MA: Morgan Kaufmann.
- Allan, L. G. (1979). The perception of time. *Perception & Psychophysics*, 26 (5), 340–354. doi: 10.3758/BF03204158
- Angrilli, A., Cherubini, P., Pavese, A. & Manfredini, S. (1997). The influence of affective factors on time perception. *Perception & Psychophysics*, 59 (6), 972–982. doi: 10.3758/BF03205512
- Antonides, G., Verhoef, P. & van Aalst, M. (2002). Consumer perception and evaluation of waiting time : A field experiment. *Journal of Consumer Psychology*, 12 (3), 193–202. doi: 10.1207/S15327663JCP1203_02
- Aranyi, G. & van Schaik, P. (2015). Testing a model of user-experience with news websites. *Journal of the Association for Information Science and Technology*, 67 (7), 1555–1575. doi: 10.1002/asi.23462
- Ariely, D. & Zakay, D. (2001). A timely account of the role of duration in decision making. *Acta Psychologica*, 108 (2), 187–207. doi: 10.1016/S0001-6918(01)00034-8
- Baayen, H., Bates, D. M., Kliegl, R. & Vasishth, S. (2015). Repsychling: Data sets from psychology and linguistics experiments [Software-Handbuch]. Zugriff auf <https://github.com/dmbates/RePsychLing> (R package version 0.0.4)
- Backhaus, N. (2017). *Nutzervertrauen und -erleben im Kontext technischer Systeme : empirische Untersuchungen am Beispiel von Webseiten und*

- Cloudspeicherdiensten* (Dissertation, Technische Universität Berlin, Berlin). doi: <http://dx.doi.org/10.14279/depositonce-5706>
- Backhaus, N. & Brandenburg, S. (2014). Zeitliche Verläufe emotionaler Aktivierung in der Mensch-Technik-Interaktion. *i-com*, 13 (1), 63–69. doi: 10.1515/icom-2014-0009
- Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature reviews neuroscience*, 4 (10), 829–839. doi: 10.1038/nrn1201
- Bakdash, J. Z. & Marusich, L. R. (2018). rmcorr: Repeated measures correlation [Software-Handbuch]. Zugriff auf <https://CRAN.R-project.org/package=rmcorr> (R package version 0.3.0)
- Balashova, T. (2018). *Auswirkungen der Zeitwahrnehmung auf das Nutzererleben in der Mensch-Computer-Interaktion* (Unveröffentlichte Masterarbeit). Technische Universität Berlin.
- Baldauf, D., Burgard, E. & Wittmann, M. (2009). Time perception as a workload measure in simulated car driving. *Applied Ergonomics*, 40 (5), 929–935. doi: 10.1016/j.apergo.2009.01.004
- 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* (S. 2689–2698).
- Baron, R. M. & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51 (6), 1173–1182. doi: DOI:10.1037//0022-3514.51.6.1173
- Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68 (3), 255–278. doi: 10.1016/j.jml.2012.11.001
- Bartoń, K. (2018). Mumin: Multi-model inference [Software-Handbuch]. Zugriff auf <https://CRAN.R-project.org/package=MumIn> (R package version 1.42.1)
- Bates, D. M. (2010). *lme4: Mixed-effects modeling with R*. New York: Springer.
- Bates, D. M. (2011). *Mixed models in R using the lme4 package part 3: Inference based on profiled deviance*. Online verfügbares Kapi-

- tel. Zugriff am 26.11.2018 auf <http://lme4.r-forge.r-project.org/slides/2011-03-16-Amsterdam/3ProfilingH.pdf>
- Bates, D. M., Kliegl, R., Vasishth, S. & Baayen, H. (2015). *Parsimonious mixed models*. Online verfügbarer Artikel. Zugriff am 30.01.2016 auf <http://arxiv.org/abs/1506.04967>
- Bates, D. M., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67 (1), 1–48. doi: 10.18637/jss.v067.i01
- Ben-Bassat, T., Meyer, J. & Tractinsky, N. (2006). Economic and subjective measures of the perceived value of aesthetics and usability. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 13 (2), 210–234. doi: 10.1145/1165734.1165737
- Bevan, N. (2009). Extending quality in use to provide a framework for usability measurement. In M. Kurosu (Hrsg.), *Human centered design. hcd 2009. lecture notes in computer science* (Bd. 5619, S. 13–22). Berlin, Heidelberg. doi: 10.1007/978-3-642-02806-9_2
- Blackler, A., Popovic, V. & Mahar, D. (2010). Investigating users’ intuitive interaction with complex artefacts. *Applied ergonomics*, 41 (1), 72–92. doi: 10.1016/j.apergo.2009.04.010
- Bland, J. M. & Altman, D. G. (1994). Correlation, regression, and repeated data. *BMJ: British Medical Journal*, 308 (6933), 896. doi: 10.1136/bmj.308.6933.896
- Bland, J. M. & Altman, D. G. (1995a). Statistics notes: Calculating correlation coefficients with repeated observations: part 1—correlation within subjects. *Bmj*, 310 (6977), 446. doi: 10.1136/bmj.310.6977.446
- Bland, J. M. & Altman, D. G. (1995b). Statistics notes: Calculating correlation coefficients with repeated observations: part 2—correlation between subjects. *BMJ*, 310 (6980), 633. doi: 10.1136/bmj.310.6980.633
- Block, R. A. (1990). Models of psychological time. In R. A. Block (Hrsg.), *Cognitive models of psychological time* (S. 1–35). Hillsdale , NJ: Lawrence Erlbaum Associates.
- Block, R. A. (1992). Prospective and retrospective duration judgment: The role of information processing and memory. In F. Macar, V. Pouthas & W. J. Friedman (Hrsg.), *Time, action and cognition: Towards bridging*

- the gap* (S. 141–152). Dordrecht: Kluwer.
- Block, R. A., George, E. J. & Reed, M. A. (1980). A watched pot sometimes boils: a study of duration experience. *Acta Psychologica*, 46 (2), 81–94. doi: 10.1016/0001-6918(80)90001-3
- Block, R. A., Hancock, P. A. & Zakay, D. (2010). How cognitive load affects duration judgments: A meta-analytic review. *Acta Psychologica*, 134 (3), 330–343. doi: 10.1016/j.actpsy.2010.03.006
- Block, R. A. & Zakay, D. (1996). Models of psychological time revisited. In H. Helfrich (Hrsg.), *Time and mind* (S. 171–195). Kirkland, WA: Hogrefe & Huber.
- Block, R. A. & Zakay, D. (1997). Prospective and retrospective duration judgments: A meta-analytic review. *Psychonomic Bulletin & Review*, 4 (2), 184–197. doi: 10.3758/BF03209393
- Bolker, B. (2019). *GLMM FAQ*. Webseite. Zugriff am 28. März 2019 auf <http://bbolker.github.io/mixedmodels-misc/glmmFAQ.html>
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler* (6. Aufl.). Heidelberg: Springer.
- Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Aufl.). Heidelberg: Springer.
- Brandenburg, S. & Backhaus, N. (2016). The dynamics of film-induced affect and its effect on the interaction with tablet pcs. *Behaviour & Information Technology*, 35 (5), 410–421. doi: 10.1080/0144929X.2016.1151076
- Brave, S. & Nass, C. (2003). Emotion in human-computer interaction. In J. A. Jacko & A. Sears (Hrsg.), *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications* (1. Aufl., S. 53–68). London: Lawrence Erlbaum Associates.
- Brouwer, A.-M., Van Wouwe, N., Muehl, C., Van Erp, J. & Toet, A. (2013). Perceiving blocks of emotional pictures and sounds: effects on physiological variables. *Frontiers in Human Neuroscience*, 7, 295. doi: 10.3389/fnhum.2013.00295
- Brown, S. W. (1985). Time perception and attention: The effects of prospective versus retrospective paradigms and task demands on perceived duration. *Perception & Psychophysics*, 38 (2), 115–124. doi: 10.3758/BF03198848
- Brown, S. W. (1997). Attentional resources in timing: Interference effects in

- concurrent temporal and nontemporal working memory tasks. *Perception & psychophysics*, 59 (7), 1118–1140. doi: 10.3758/BF03205526
- Brown, S. W. (2006). Timing and executive function: bidirectional interference between concurrent temporal production and randomization tasks. *Memory & Cognition*, 34 (7), 1464–1471. doi: 10.3758/BF03195911
- Brown, S. W. (2008). Time and attention: Review of the literature. In S. Grondin (Hrsg.), *Psychology of time* (S. 111–138). Bingley: Emerald.
- Brown, S. W. & Boltz, M. G. (2002). Attentional processes in time perception: Effects of mental workload and event structure. *Journal of Experimental Psychology: Human Perception and Performance*, 28 (3), 600–615. doi: 10.1037/0096-1523.28.3.600
- Bühner, M. (2011). *Einführung in die Test-und Fragebogenkonstruktion*. München: Pearson.
- Burmester, M., Jäger, K., Mast, M., Peissner, M. & Sproll, S. (2010). Design verstehen – Formative Evaluation der User Experience. In H. Brau, S. Diefenbach, K. Göring, M. Peissner & K. Petrovic (Hrsg.), *Tagungsband Usability Professionals 2010* (S. 206–211). Stuttgart: Fraunhofer Verlag.
- Castellà, J., Cuello, C. & Sanz, A. (2017). Does time fly 20 m above the ground? Exploring the role of affective response on time perception in a high-risk sport. *Applied Cognitive Psychology*, 31 (6), 644–652. doi: 10.1002/acp.3367
- Chin, J. P., Diehl, V. A. & Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of the SIGCHI conference on human factors in computing systems* (S. 213–218). ACM.
- Church, R. M., Meck, W. H. & Gibbon, J. (1994). Application of scalar timing theory to individual trials. *Journal of Experimental Psychology: Animal Behavior Processes*, 20 (2), 135–155. doi: 10.1037//0097-7403.20.2.135
- Cohen, J. (1992). A power primer. *Psychological bulletin*, 112 (1), 155–159. doi: 10.1037/0033-2909.112.1.155
- Courtney, M. G. R. & Gordon, M. (2013). Determining the number of factors to retain in EFA: Using the SPSS R-Menu v2. 0 to make more judicious estimations. *Practical assessment, research & evaluation*, 18 (8), 1–14.

- Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in quantitative methods for psychology*, 1 (1), 42–45. doi: 10.20982/tqmp.01.1.p042
- Csikszentmihalyi, M. (2014). Toward a psychology of optimal experience. In *Flow and the foundations of positive psychology* (S. 209–226). Dordrecht: Springer. doi: 10.1007/978-94-017-9088-8_14
- Dellaert, B. G. & Kahn, B. E. (1998). How tolerable is delay?: Consumers' evaluations of internet web sites after waiting. *Journal of interactive marketing*, 13 (1), 41–54. doi: 10.1002/(SICI)1520-6653(199924)13:1<41::AID-DIR4>3.0.CO;2-S
- DIN, EN ISO 9241-11. (2018). *Ergonomie der Mensch-System-Interaktion – Teil 11: Gebrauchstauglichkeit: Begriffe und Konzepte*. Berlin: Beuth Verlag.
- DIN, EN ISO 9241-110. (2008). *Ergonomie der Mensch-System-Interaktion – Teil 110: Grundsätze der Dialoggestaltung*. Berlin: Beuth Verlag.
- Droit-Volet, S., Bigand, E., Ramos, D. & Bueno, J. L. O. (2010). Time flies with music whatever its emotional valence. *Acta Psychologica*, 135 (2), 226–232. doi: 10.1016/j.actpsy.2010.07.003
- Droit-Volet, S., Brunot, S. & Niedenthal, P. M. (2004). Perception of the duration of emotional events. *Cognition & Emotion*, 18 (6), 849–858. doi: 10.1080/02699930341000194
- Droit-Volet, S., Fayolle, S. L. & Gil, S. (2011). Emotion and time perception: Effects of film-induced mood. *Frontiers in Integrative Neuroscience*, 5. Zugriff am 2014-03-03 auf <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3152725/> doi: 10.3389/fnint.2011.00033
- Droit-Volet, S. & Gil, S. (2015). The emotional body and time perception. *Cognition & Emotion*, 1–13. doi: 10.1080/02699931.2015.1023180
- Droit-Volet, S. & Meck, W. H. (2007). How emotions colour our perception of time. *Trends in Cognitive Sciences*, 11 (12), 504–513. doi: 10.1016/j.tics.2007.09.008
- Droit-Volet, S., Meck, W. H. & Penney, T. B. (2007). Sensory modality and time perception in children and adults. *Behavioural Processes*, 74 (2), 244–250. doi: 10.1016/j.beproc.2006.09.012
- Droit-Volet, S., Mermillod, M., Cocenas-Silva, R. & Gil, S. (2010). The effect

- of expectancy of a threatening event on time perception in human adults. *Emotion*, 10 (6), 908–914. doi: 10.1037/a0020258
- Droit-Volet, S., Monceau, S., Berthon, M., Trahanias, P. & Maniadakis, M. (2018). The explicit judgment of long durations of several minutes in everyday life: Conscious retrospective memory judgment and the role of affects? *PloS one*, 13 (4), e0195397. doi: 10.1371/journal.pone.0195397
- Droit-Volet, S., Ramos, D., Bueno, L. J. & Bigand, E. (2013). Music, emotion, and time perception: the influence of subjective emotional valence and arousal? *Emotion Science*, 4, 417. doi: 10.3389/fpsyg.2013.00417
- Droit-Volet, S., Trahanias, P. & Maniadakis, M. (2017). Passage of time judgments in everyday life are not related to duration judgments except for long durations of several minutes. *Acta psychologica*, 173, 116–121. doi: 10.1016/j.actpsy.2016.12.010
- Droit-Volet, S. & Wearden, J. H. (2003). Les modèles d’horloge interne en psychologie du temps. *L’année psychologique*, 104, 617–654. doi: 10.3406/psy.2003.29656
- Droit-Volet, S. & Wearden, J. H. (2015). Experience sampling methodology reveals similarities in the experience of passage of time in young and elderly adults. *Acta Psychologica*, 156, 77–82. doi: 10.1016/j.actpsy.2015.01.006
- Droit-Volet, S. & Wearden, J. H. (2016). Passage of time judgments are not duration judgments: Evidence from a study using experience sampling methodology. *Frontiers in Psychology*, 7, 176. doi: 10.3389/fpsyg.2016.00176
- Elias, N. (1988). *Über die Zeit*. Baden-Baden: suhrkamp taschenbuch wissenschaft.
- Fayolle, S., Gil, S. & Droit-Volet, S. (2015). Fear and time: fear speeds up the internal clock. *Behavioural processes*, 120, 135–140. doi: 10.1016/j.beproc.2015.09.014
- Field, A., Miles, J. & Field, Z. (2012). *Discovering statistics using R*. London: SAGE.
- Fox, J. & Weisberg, S. (2011). *An R companion to applied regression* (2. Aufl.). Thousand Oaks CA: Sage. Zugriff am 29.03.2019 auf <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>

- Frøkjær, E., Hertzum, M. & Hornbæk, K. (2000). Measuring usability: are effectiveness, efficiency, and satisfaction really correlated? In *Proceedings of the SIGCHI conference on human factors in computing systems* (S. 345–352). ACM.
- Gibbon, J. (1977). Scalar expectancy theory and weber's law in animal timing. *Psychological Review*, 84 (3), 279–325. doi: 10.1037/0033-295X.84.3.279
- Gibbon, J. & Church, R. M. (1984). Sources of variance in an information processing theory of timing. In H. Roitblat, H. Terrace & T. Bever (Hrsg.), *Animal cognition* (S. 465–488). Taylor & Francis. Zugriff am 29.03.2019 auf <https://books.google.de/books?id=60JHAWAAQBAJ>
- Gibbon, J., Church, R. M. & Meck, W. H. (1984). Scalar timing in memory. *Annals of the New York Academy of Sciences*, 423 (1), 52–77. doi: 10.1111/j.1749-6632.1984.tb23417.x
- Gil, S. & Droit-Volet, S. (2009). Time perception, depression and sadness. *Behavioural Processes*, 80 (2), 169–176. doi: 10.1016/j.beproc.2008.11.012
- Gil, S. & Droit-Volet, S. (2011). "Time flies in the presence of angry faces"... depending on the temporal task used! *Acta psychologica*, 136 (3), 354–362. doi: 10.1016/j.actpsy.2010.12.010
- Gil, S. & Droit-Volet, S. (2012). Emotional time distortions: The fundamental role of arousal. *Cognition & Emotion*, 26 (5), 847–862. doi: 10.1080/02699931.2011.625401
- Gorn, G. J., Chattopadhyay, A., Sengupta, J. & Tripathi, S. (2004). Waiting for the web: How screen color affects time perception. *Journal of Marketing Research*, 41 (2), 215–225. doi: 10.1509/jmkr.41.2.215.28668
- Green, P. & MacLeod, C. J. (2016). Simr: an r package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7 (4), 493–498. doi: 10.1111/2041-210X.12504
- Grieb, F. T. (2015). *Zeitwahrnehmung bei Usability-Tests: Auswirkung von Usability und Lautem Denken auf Zeitschätzungen* (Unveröffentlichte Bachelorarbeit). Technische Universität Berlin.
- Grommet, E. K., Droit-Volet, S., Gil, S., Hemmes, N. S., Baker, A. H. & Brown, B. L. (2011). Time estimation of fear cues in human observers. *Behavioural Processes*, 86 (1), 88–93. doi: 10.1016/j.beproc.2010.10.003

- Grondin, S. (2010). Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception, & Psychophysics*, 72 (3), 561–582. doi: 10.3758/APP.72.3.561
- Grondin, S., Laflamme, V. & Gontier, É. (2014). Effect on perceived duration and sensitivity to time when observing disgusted faces and disgusting mutilation pictures. *Attention, Perception, & Psychophysics*, 1–13. doi: 10.3758/s13414-014-0682-7
- Hagendorf, H., Krummenacher, J., Müller, H.-J. & Schubert, T. (2011). Selektive aufmerksamkeit. In H. Hagendorf, J. Krummenacher, H.-J. Müller & T. Schubert (Hrsg.), *Wahrnehmung und aufmerksamkeit: Allgemeine psychologie für bachelor* (S. 179–201). Berlin, Heidelberg: Springer Berlin Heidelberg. Zugriff auf https://doi.org/10.1007/978-3-642-12710-6_15 doi: 10.1007/978-3-642-12710-6_15
- Hamborg, K.-C., Hülsmann, J. & Kaspar, K. (2014). The interplay between usability and aesthetics: More evidence for the “what is usable is beautiful” notion. *Advances in Human-Computer Interaction, 2014*, 15. doi: 10.1155/2014/946239
- Hassenzahl, M. (2003). The thing and I: Understanding the relationship between user and product. In M. A. Blythe, K. Overbeeke, A. F. Monk & P. C. Wright (Hrsg.), *Funology: From usability to enjoyment* (S. 31–42). Netherlands: Springer. doi: 10.1007/1-4020-2967-5
- Hassenzahl, M. (2004). The interplay of beauty, goodness, and usability in interactive products. *Human-computer interaction*, 19 (4), 319–349. doi: 10.1207/s15327051hci1904_2
- Hassenzahl, M. (2008). User experience (UX): towards an experiential perspective on product quality. In *Proceedings of the 20th Conference on l'Interaction Homme-Machine* (S. 11–15). ACM.
- Hassenzahl, M., Burmester, M. & Koller, F. (2008). Der User Experience (UX) auf der Spur: Zum Einsatz von www.attrakdiff.de. In H. Brau, S. Diefenbach, M. Hassenzahl, F. Koller, M. Peissner & K. Röse (Hrsg.), *Tagungsband Usability Professionals 2008* (S. 78–82). Fraunhofer Verlag.
- Hassenzahl, M., Eckoldt, K. & Thielsch, M. T. (2009). User Experience und Experience Design – Konzepte und Herausforderungen. In *Tagungsband up09* (S. 233–237). Fraunhofer Verlag.

- Hermosa Perrino, C. & Burmester, M. (2017). Designing for Time Perspectives – Gestaltung der zeitlichen Dimension der UX. In *Usability Professionals 2017, Regensburg, Germany, September 10-13, 2017*. Gesellschaft für Informatik eV. doi: 10.18420/muc2017-up-0176
- Hertzum, M., Hansen, K. D. & Andersen, H. H. (2009). Scrutinising usability evaluation: does thinking aloud affect behaviour and mental workload? *Behaviour & Information Technology*, 28 (2), 165–181. doi: 10.1080/01449290701773842
- Hertzum, M. & Holmegaard, K. D. (2013). Thinking aloud in the presence of interruptions and time constraints. *International Journal of Human-Computer Interaction*, 29 (5), 351–364.
- Hertzum, M. & Holmegaard, K. D. (2015). Thinking aloud influences perceived time. *Human Factors*, 57 (1), 101–109. doi: 10.1177/0018720814540208
- Hoenig, J. M. & Heisey, D. M. (2001). The abuse of power: the pervasive fallacy of power calculations for data analysis. *The American Statistician*, 55 (1), 19–24. doi: 10.1198/000313001300339897
- Horr, N. K. & Di Luca, M. (2015). Filling the blanks in temporal intervals: the type of filling influences perceived duration and discrimination performance. *Perception Science*, 6, 114. doi: 10.3389/fpsyg.2015.00114
- Hoxmeier, J. A. & DiCesare, C. (2000). System response time and user satisfaction: An experimental study of browser-based applications. *AMCIS 2000 Proceedings*, 140–145.
- Hudlicka, E. (2003). To feel or not to feel: The role of affect in human–computer interaction. *International journal of human-computer studies*, 59 (1-2), 1–32. doi: 10.1016/S1071-5819(03)00047-8
- Ivry, R. B. & Schlerf, J. E. (2008). Dedicated and intrinsic models of time perception. *Trends in cognitive sciences*, 12, 273–280. doi: 10.1016/j.tics.2008.04.002
- Johnson, P. C. (2014). Extension of Nakagawa & Schielzeth’s R2GLMM to random slopes models. *Methods in Ecology and Evolution*, 5 (9), 944–946. doi: 10.1111/2041-210X.12225
- Jokinen, J. P. (2015). Emotional user experience: traits, events, and states. *International Journal of Human-Computer Studies*, 76, 67–77. doi: 10.1016/j.ijhcs.2014.12.006

- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.
- Karapanos, E., Zimmerman, J., Forlizzi, J. & Martens, J.-B. (2010). Measuring the dynamics of remembered experience over time. *Interacting with Computers*, 22 (5), 328–335. doi: 10.1016/j.intcom.2010.04.003
- Kim, E. (2017). *Der Einfluss manipulierter Zeitempfindens auf das Nutzererleben* (Unveröffentlichte Masterarbeit). Technische Universität Berlin.
- Kliegl, R., Masson, M. E. & Richter, E. M. (2010). A linear mixed model analysis of masked repetition priming. *Visual Cognition*, 18 (5), 655–681. doi: 10.1080/13506280902986058
- Kliegl, R., Wei, P., Dambacher, M., Yan, M. & Zhou, X. (2011). Experimental effects and individual differences in linear mixed models: estimating the relationship between spatial, object, and attraction effects in visual attention. *Quantitative Psychology and Measurement*, 1:238, 1–12. doi: 10.3389/fpsyg.2010.00238
- Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E. & Sinnelä, A. (2011). UX Curve: A method for evaluating long-term user experience. *Interacting with Computers*, 23 (5), 473–483. doi: 10.1016/j.intcom.2011.06.005
- Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K. & Sinnelä, A. (2011). Identifying hedonic factors in long-term user experience. In *Proceedings of the 2011 conference on designing pleasurable products and interfaces* (S. 17).
- Kurusathianpong, P. & Tangmanee, C. (2018). Comparison of perceived waiting time between two lengths of progress indicator and two styles of graphics animation with perceived uncertainty as a covariate. In *Seventh ICT International Student Project Conference (ICT-ISPC)* (S. 1–6). IEEE.
- Lallemant, C., Gronier, G. & Koenig, V. (2015). User experience: A concept without consensus? Exploring practitioners’ perspectives through an international survey. *Computers in Human Behavior*, 43, 35–48. doi: 10.1016/j.chb.2014.10.048
- Laugwitz, B., Schrepp, M. & Held, T. (2006). Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten. In *Mensch & Computer* (S. 125–134). doi: 10.1524/9783486841749.125

- Lavie, T. & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. *International journal of human-computer studies*, 60 (3), 269–298. doi: 10.1016/j.ijhcs.2003.09.002
- Law, E. L.-C., Roto, V., Hassenzahl, M., Vermeeren, A. P. & Kort, J. (2009). Understanding, scoping and defining user experience: a survey approach. In *Proceedings of the SIGCHI conference on human factors in computing systems* (S. 719–728). ACM.
- Ledesma, R. D. & Valero-Mora, P. (2007). Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis. *Practical assessment, research & evaluation*, 12 (2), 1–11.
- Lee, Y., Chen, A. N. & Ilie, V. (2012). Can online wait be managed? the effect of filler interfaces and presentation modes on perceived waiting time online. *MIS Quarterly*, 365–394.
- Lesemann, E., Woletz, N. & Koerber, S. (2007). Combining methods to evaluate mobile usability. In *Proceedings of the 9th international conference on human computer interaction with mobile devices and services* (S. 444–447). ACM.
- Lewis, J. R. (1992). Psychometric evaluation of the post-study system usability questionnaire: The PSSUQ. In *Proceedings of the human factors and ergonomics society annual meeting* (Bd. 36, S. 1259–1260). Los Angeles, CA: SAGE Publications.
- Lewis, J. R. & Mayes, D. K. (2014). Development and psychometric evaluation of the emotional metric outcomes (emo) questionnaire. *International Journal of Human-Computer Interaction*, 30 (9), 685–702. doi: 10.1080/10447318.2014.930312
- Liikkanen, L. A. & Gómez Gómez, P. (2013). Designing interactive systems for the experience of time. In *Proceedings of the 6th international conference on designing pleasurable products and interfaces* (S. 146–155). ACM. doi: 10.1145/2513506.2513522
- Lind, M. & Sundvall, H. (2007). Time estimation as a measure of mental workload. In *International Conference on Engineering Psychology and Cognitive Ergonomics* (S. 359–365). Heidelberg: Springer. doi: 10.1007/978-3-540-73331-7_39

- Lund, A. M. (2001). Measuring usability with the USE questionnaire. *Usability interface*, 8 (2), 3–6.
- Mahlke, S. (2008). *User experience of interaction with technical systems* (Dissertation, Technische Universität Berlin, Berlin). doi: <http://dx.doi.org/10.14279/depositonce-1793>
- Mair, P. & Wilcox, R. (2018). WRS2: Wilcox robust estimation and testing [Software-Handbuch]. (R package version 0.10-0)
- Matell, M. S., King, G. R. & Meck, W. H. (2004). Differential modulation of clock speed by the administration of intermittent versus continuous cocaine. *Behavioral neuroscience*, 118 (1), 150–156. doi: 10.1037/0735-7044.118.1.150
- Mattes, S. & Ulrich, R. (1998). Directed attention prolongs the perceived duration of a brief stimulus. *Perception & Psychophysics*, 60 (8), 1305–1317. doi: <https://doi.org/10.3758/BF03207993>
- Matthews, W. J. & Meck, W. H. (2016). Temporal cognition: Connecting subjective time to perception, attention, and memory. *Psychological bulletin*, 142 (8), 865–907. doi: 10.1037/bul0000045
- McDonald, S., Zhao, T. & Edwards, H. M. (2013). Dual verbal elicitation: the complementary use of concurrent and retrospective reporting within a usability test. *International Journal of Human-Computer Interaction*, 29 (10), 647–660. doi: 10.1080/10447318.2012.758529
- Meck, W. H. (1984). Attentional bias between modalities: Effect on the internal clock, memory, and decision stages used in animal time discrimination a. *Annals of the New York Academy of sciences*, 423 (1), 528–541. doi: 10.1111/j.1749-6632.1984.tb23457.x
- Mella, N., Conty, L. & Pouthas, V. (2011). The role of physiological arousal in time perception: Psychophysiological evidence from an emotion regulation paradigm. *Brain and Cognition*, 75 (2), 182–187. doi: 10.1016/j.bandc.2010.11.012
- Minge, M. (2008). Dynamics of user experience. In *Proceedings of the Workshop on Research Goals and Strategies for Studying User Experience and Emotion, NordiCHI 2008* (Bd. 8, S. 1–5).
- Minge, M. (2011). *Dynamische aspekte des nutzungserlebens der interaktion mit technischen systemen* (Dissertation, Technische Universität Berlin,

- Berlin). doi: <http://dx.doi.org/10.14279/depositonce-2992>
- Minge, M. & Riedel, L. (2013). meCUE-Ein modularer Fragebogen zur Erfassung des Nutzungserlebens. In *Mensch & Computer* (S. 89–98).
- Minge, M. & Thüring, M. (2018). Hedonic and pragmatic halo effects at early stages of user experience. *International Journal of Human-Computer Studies*, 109, 13–25. doi: 10.1016/j.ijhcs.2017.07.007
- Minge, M., Thüring, M. & Wagner, I. (2016). Developing and validating an english version of the mecue questionnaire for measuring user experience. In *Proceedings of the human factors and ergonomics society 2016 annual meeting* (Bd. 60, S. 2063–2067). Los Angeles, CA: SAGE Publications.
- Mioni, G., Stablum, F., McClintock, S. M. & Grondin, S. (2014). Different methods for reproducing time, different results. *Attention, perception & psychophysics*, 76 (3), 675–681. doi: 10.3758/s13414-014-0625-3
- Moosbrugger, H. & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion*. Berlin: Springer.
- Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, 4 (2), 61–64. doi: 10.20982/tqmp.04.2.p061
- Morey, R. D. & Rouder, J. N. (2018). Bayesfactor: Computation of bayes factors for common designs [Software-Handbuch]. Zugriff auf <https://CRAN.R-project.org/package=BayesFactor> (R package version 0.9.12-4.2)
- Moshagen, M. & Thielsch, M. (2013). A short version of the visual aesthetics of websites inventory. *Behaviour & Information Technology*, 32 (12), 1305–1311. doi: 10.1080/0144929X.2012.694910
- Mrowetz, L. (2016). *Einfluss von Usability und Workload auf die Zeitwahrnehmung während Usability-Tests* (Unveröffentlichte Bachelorarbeit). Technische Universität Berlin.
- Müsseler, J. (2008). *Allgemeine Psychologie* (2. Aufl.). Springer.
- Nakagawa, S. & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4, 133–142. doi: 10.1111/J.2041-210x.2012.00261.X
- Nather, F. C., Bueno, J. L. O., Bigand, E. & Droit-Volet, S. (2011). Time changes with the embodiment of another’s body posture. *PLoS ONE*, 6

- (5), e19818. doi: 10.1371/journal.pone.0019818
- Ngo, D. C. L. (2001). Measuring the aesthetic elements of screen designs. *Displays*, 22 (3), 73–78. doi: 10.1016/S0141-9382(01)00053-1
- Nielsen, J. (1994). *Usability engineering*. Cambridge, MA: AP Professional.
- Nielsen, J. (2012). *Usability 101: Introduction to usability*. Online verfügbares Dokument. Zugriff am 24.08.2018 auf http://dockerby.com/web/Unit%206%20Validating/Usability%20101_%20Introduction%20to%20Usability.pdf
- Noulhiane, M., Mella, N., Samson, S., Ragot, R. & Pouthas, V. (2007). How emotional auditory stimuli modulate time perception. *Emotion*, 7 (4), 697–704. doi: 10.1037/1528-3542.7.4.697
- Nürck, K. (2015). *Emotionales Nutzererleben und Stimme: Analysen von Stimmparametern beim Lauten Denken in der Mensch-Computer-Interaktion* (Unveröffentlichte Bachelorarbeit). Technische Universität Berlin.
- O'Connor, B. P. (2017). paramap: paramap [Software-Handbuch]. Zugriff auf <https://people.ok.ubc.ca/briocconn/nfactors/paramap.pdf> (R package version 1.4)
- Ono, F. & Kawahara, J.-I. (2007). The subjective size of visual stimuli affects the perceived duration of their presentation. *Perception & psychophysics*, 69 (6), 952–957. doi: 10.3758/BF03193932
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y. & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of applied psychology*, 88 (5), 879–903. doi: 10.1037/0021-9101.88.5.879
- Podsakoff, P. M., MacKenzie, S. B. & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual review of psychology*, 63, 539–569. doi: 10.1146/annurev-psych-120710-100452
- Pohlmeier, A. E., Hecht, M. & Blessing, L. (2009). User experience lifecycle model ContinUE [Continuous User Experience]. In *Der mensch im mittelpunkt technischer systeme. fortschritt-berichte vdi reihe* (Bd. 22, S. 314–317). Düsseldorf: VDI-Verlag.
- R Core Team. (2018). R: A language and environment for statistical com-

- puting [Software-Handbuch]. Vienna, Austria. Zugriff auf <https://www.R-project.org/> (R version 3.5.1)
- Rammsayer, T. H. & Verner, M. (2014). The effect of nontemporal stimulus size on perceived duration as assessed by the method of reproduction. *Journal of Vision*, *14* (5), 1-10. doi: 10.1167/14.5.17
- Revelle, W. (2018). psych: Procedures for psychological, psychometric, and personality research [Software-Handbuch]. Zugriff auf <https://CRAN.R-project.org/package=psych> (R package version 1.8.4)
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review*, *110* (1), 145—172. doi: 10.1037//0033-295X.110.1.145
- Russell, J. A. & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of personality and social psychology*, *76* (5), 805–819. doi: 10.1037//0022-3514.76.5.805
- Russell, J. A., Weiss, A. & Mendelsohn, G. A. (1989). Affect grid: a single-item scale of pleasure and arousal. *Journal of personality and social psychology*, *57* (3), 493–502. doi: 10.1037/0022-3514.57.3.493
- Sackett, A. M., Meyvis, T., Nelson, L. D., Converse, B. A. & Sackett, A. L. (2010). You're having fun when time flies the hedonic consequences of subjective time progression. *Psychological Science*, *21* (1), 111–117. doi: 10.1177/0956797609354832
- Sarkar, D. (2008). *Lattice: Multivariate data visualization with R*. New York: Springer. Zugriff auf <http://lmdvr.r-forge.r-project.org> (ISBN 978-0-387-75968-5)
- Schad, D. J., Hohenstein, S., Vasisht, S. & Kliegl, R. (2018). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *arXiv preprint arXiv:1807.10451*.
- Schafer, J., Opgen-Rhein, R., Zuber, V., Ahdesmaki, M., Silva, A. P. D. & Strimmer, K. (2017). corpcor: Efficient estimation of covariance and (partial) correlation [Software-Handbuch]. Zugriff auf <https://CRAN.R-project.org/package=corpcor> (R package version 1.6.9)
- Scherer, K. R. (2005). What are emotions? and how can they be measured? *Social Science Information*, *44* (4), 695–729. doi: 10.1177/

0539018405058216

- Schielzeth, H. & Forstmeier, W. (2009). Conclusions beyond support: overconfident estimates in mixed models. *Behavioral Ecology*, 20 (2), 416–420. doi: 10.1093/beheco/arn145
- Schorling, L. (2018). *On the influence of manipulated time perception on the evaluation of an online tool* (Unveröffentlichte Masterarbeit). Technische Universität Berlin.
- Sears, A., Jacko, J. A. & Borella, M. S. (1997). Internet delay effects: how users perceive quality, organization, and ease of use of information. In *CHI'97 Extended Abstracts on Human Factors in Computing Systems* (S. 353–354). ACM.
- Seow, S. C. (2008). *Designing and engineering time - the psychology of time perception in software*. Pearson Education.
- Shi, Z., Jia, L. & Müller, H. J. (2012). Modulation of tactile duration judgments by emotional pictures. *Frontiers in Integrative Neuroscience*, 6. doi: 10.3389/fnint.2012.00024
- Sonderegger, A. & Sauer, J. (2010). The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. *Applied ergonomics*, 41 (3), 403–410. doi: 10.1016/j.apergo.2009.09.002
- Steinbach, A. B. (2017). *Untersuchung zum Zusammenhang von Usability und Zeitwahrnehmung* (Unveröffentlichte Masterarbeit). Technische Universität Berlin.
- Sucala, M., Scheckner, B. & David, D. (2011). Psychological time: interval length judgments and subjective passage of time judgments. *Current psychology letters. Behaviour, brain & cognition*, 26 (2, 2010). Zugriff auf <http://journals.openedition.org/cpl/4998>
- Szameitat, A. J., Rummel, J., Szameitat, D. P. & Sterr, A. (2009). Behavioral and emotional consequences of brief delays in human–computer interaction. *International Journal of Human-Computer Studies*, 67 (7), 561–570. doi: 10.1016/j.ijhcs.2009.02.004
- Tanaka, R. & Yotsumoto, Y. (2017). Passage of time judgments is relative to temporal expectation. *Frontiers in Psychology*, 8, 187. doi: 10.3389/fpsyg.2017.00187
- Thielsch, M. T. & Moshagen, M. (2011). Erfassung visueller Ästhetik mit

- dem VisAWI. In H. Brau, A. Lehmann, K. Petrovic & M. C. Schroeder (Hrsg.), (S. 260–265). Stuttgart: German UPA eV.
- Thomaschke, R. & Haering, C. (2014). Predictivity of system delays shortens human response time. *International Journal of Human-Computer Studies*, 72, 358–365. doi: 10.1016/j.ijhcs.2013.12.004
- Thönes, S., Castell, C., Iflinger, J. & Oberfeld, D. (2018). Color and time perception: Evidence for temporal overestimation of blue stimuli. *Scientific reports*, 8 (1), 1688. doi: 10.1038/s41598-018-19892-z
- Thüring, M. (2013). Nutzererleben Komponenten, Phasen, Phänomene. In S. Boll, S. Maaß & R. Malaka (Hrsg.), *Workshopband Mensch & Computer 2013* (S. 113–120). München: Oldenbourg Verlag.
- Thüring, M. & Mahlke, S. (2007). Usability, aesthetics and emotions in human–technology interaction. *International Journal of Psychology*, 42 (4), 253–264. doi: 10.1080/00207590701396674
- Tipples, J. (2008). Negative emotionality influences the effects of emotion on time perception. *Emotion*, 8 (1), 127–131. doi: 10.1037/1528-3542.8.1.127
- Tractinsky, N., Katz, A. S. & Ikar, D. (2000). What is beautiful is usable. *Interacting with computers*, 13 (2), 127–145. doi: 10.1016/S0953-5438(00)00031-X
- Trapp, A. K. & Thüring, M. (2016). Emotional influences on time perception. In A. Papafragou, D. Grodner, D. Mirman & J. Trueswell (Hrsg.), *Proceedings of the 38th annual conference of the cognitive science society*. (S. 2657–2657). Austin, TX: Cognitive Science Society.
- Treisman, M. (1963). Temporal discrimination and the indifference interval. implications for a model of the “internal clock”. *Psychological Monographs: General and Applied*, 77, 1–31. doi: 10.1037/h0093864
- Treisman, M. (2013). The information-processing model of timing (treisman, 1963): Its sources and further development. *Timing & Time Perception*, 1, 131–158. doi: 10.1163/22134468-00002017
- Tullis, T. & Albert, W. (2010). *Measuring the user experience: Collecting, analyzing, and presenting usability metrics*. Morgan Kaufmann.
- Ulbrich, P., Churan, J., Fink, M. & Wittmann, M. (2007). Temporal reproduction: Further evidence for two processes. *Acta Psychologica*, 125 (1),

- 51–65. doi: 10.1016/j.actpsy.2006.06.004
- Ullrich, F. K. (2018). *Der Zusammenhang zwischen Usability und Zeitwahrnehmung bei konstanten Bearbeitungsdauern in der Mensch-Computer-Interaktion* (Unveröffentlichte Masterarbeit). Technische Universität Berlin.
- van der Geest, T. & van Dongelen, R. (2009). What is beautiful is useful-visual appeal and expected information quality. In *Professional Communication Conference, IPCC 2009* (S. 1–5). IEEE International.
- van Schaik, P., Hassenzahl, M. & Ling, J. (2012). User-experience from an inference perspective. *ACM Transactions on Computer-Human Interaction*, 19 (2), 11. doi: 10.1145/2240156.2240159
- Venables, W. N. & Ripley, B. D. (2002). *Modern applied statistics with s* (Fourth Aufl.). New York: Springer. Zugriff auf <http://www.stats.ox.ac.uk/pub/MASS4>
- von Wilamowitz-Moellendorff, M., Hassenzahl, M. & Platz, A. (2006). Dynamics of user experience: How the perceived quality of mobile phones changes over time. In E. L.-C. Law, E. T. Hvannberg & M. Hassenzahl (Hrsg.), *UX WS NordiCHI'06: COST294-MAUSE* (S. 74–78).
- Walsh, T., Varsaluoma, J., Kujala, S., Nurkka, P., Petrie, H. & Power, C. (2014). Axe UX: Exploring long-term user experience with iScale and AttrakDiff. In *Proceedings of the 18th international academic mindtrek conference: Media business, management, content & services* (S. 32–39). ACM.
- Wearden, J. H. (2015). Passage of time judgements. *Consciousness and Cognition*, 38, 165 – 171. doi: 10.1016/j.concog.2015.06.005
- Wearden, J. H. (2016). *The psychology of time perception*. London: palgrave macmillan.
- Wearden, J. H. & Lejeune, H. (2008). Scalar properties in human timing: Conformity and violations. *The Quarterly Journal of Experimental Psychology*, 61 (4), 569–587. doi: 10.1080/17470210701282576
- Wearden, J. H., O'Donoghue, A., Ogden, R. & Montgomery, C. (2014). Subjective duration in the laboratory and the world outside. In V. Arstila & D. Lloyd (Hrsg.), *Subjective time: The philosophy, psychology, and neuroscience of temporality* (S. 287–306). MA, USA: MIT Press.

- Weber, F., Haering, C. & Thomaschke, R. (2013). Improving the human–computer dialogue with increased temporal predictability. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 55 (5), 881–892. doi: 10.1177/0018720813475812
- Wei, T. & Simko, V. (2017). R package "corrplot": Visualization of a correlation matrix [Software-Handbuch]. Zugriff auf <https://github.com/taiyun/corrplot> (R package version 0.84)
- Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, 3 (2), 159–177. doi: 10.1080/14639220210123806
- Wickens, C. D. (2008). Multiple resources and mental workload. *Human factors*, 50 (3), 449–455. doi: 10.1518/001872008X288394
- Wickens, C. D. & Gopher, D. (1977). Control theory measures of tracking as indices of attention allocation strategies. *Human Factors*, 19 (4), 349–365. doi: 10.1177/001872087701900404
- Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, 21 (12). Zugriff auf <http://www.jstatsoft.org/v21/i12/paper>
- Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, 40 (1), 1–29. Zugriff auf <http://www.jstatsoft.org/v40/i01/>
- Wickham, H. (2017). tidyverse: Easily install and load the 'tidyverse' [Software-Handbuch]. Zugriff auf <https://CRAN.R-project.org/package=tidyverse> (R package version 1.2.1)
- Wickham, H., Hester, J. & Chang, W. (2018). devtools: Tools to make developing r packages easier [Software-Handbuch]. Zugriff auf <https://CRAN.R-project.org/package=devtools> (R package version 1.13.6)
- Winter, B. (2013). A very basic tutorial for performing linear mixed effects analyses. *arXiv preprint arXiv:1308.5499*.
- Wittmann, M. (2013). *Gefühlte Zeit - Kleine Psychologie des Zeitempfindens*. München: becksche reihe.
- Xuan, B., Zhang, D., He, S. & Chen, X. (2007). Larger stimuli are judged to last longer. *Journal of vision*, 7 (10), 1–5. doi: 10.1167/7.10.2
- Zakay, D. (1993a). Relative and absolute duration judgments under prospective

- and retrospective paradigms. *Perception & Psychophysics*, 54 (5), 656–664. doi: 10.3758/BF03211789
- Zakay, D. (1993b). Time estimation methods—do they influence prospective duration estimates? *Perception*, 22 (1), 91–101. doi: 10.1068/p220091
- Zakay, D. (2005). Attention et jugement temporel. *Psychologie française*, 50 (1), 65–79. doi: 10.1016/j.psfr.2004.10.004
- Zakay, D. (2015). The temporal-relevance temporal-uncertainty model of prospective duration judgment. *Consciousness and cognition*, 38, 182–190. doi: 10.1016/j.concog.2015.10.006
- Zakay, D. & Block, R. A. (1995). An attentional-gate model of prospective time estimation. In V. de Keyser, G. d’Ydewakke & A. Vandierendonck (Hrsg.), *Time and the dynamic control of behavior* (S. 167–178). Liège, Belgium: Universite de Liege.
- Zakay, D. & Block, R. A. (1997). Temporal cognition. *Current Directions in Psychological Science*, 6 (1), 12–16. doi: 10.1111/1467-8721.ep11512604
- Zakay, D. & Block, R. A. (2004). Prospective and retrospective duration judgments: an executive-control perspective. *Acta Neurobiologiae Experimentalis*, 64 (3), 319–328.
- Zhang, M., Lu, A. & Hodges, B. H. (2019). Lifting, tasting, and carrying: The interaction of magnitude and valence effects in time perception. *Acta psychologica*, 193, 1–10. doi: 10.1016/j.actpsy.2018.11.010
- Ziefle, M. & Bay, S. (2005). How older adults meet complexity: aging effects on the usability of different mobile phones. *Behaviour & information technology*, 24 (5), 375–389. doi: 10.1080/0144929042000320009

Anhang A

Abkürzungen

AGM Attentional Gate Model

ALM Allgemeines Lineares Modell

ANCOVA Kovarianzanalyse

ANOVA Varianzanalyse

AV abhängige Variable

bzw. beziehungsweise

ContinUE-Modell User Experience Lifecycle Model

CUE-Modell Components of User Experience Model

d.h. das heißt

ggf. gegebenenfalls

HPE high-priority event

Hyp. Hypothese

Kap. Kapitel

KMO Kaiser-Meyer-Olkin-Koeffizienten

Konf. Konfidenzintervall

LMM Linear Mixed-Effects Model

Min. Minuten

ML Maximum Likelihood Schätzung

ms Millisekunden

NA not available

PCA Principle Component Analysis

PoTJ Passage of Time Judgement

PSE Point of Subjective Equality

PTR Perceived Time Ratio

Sek. Sekunden

SET Scalar Expectancy Theory

Std. Fehler Standardfehler

STT Scalar Timing Theory

UV unabhängige Variable

UX User Experience

vgl. vergleiche

z.B. zum Beispiel

Anhang B

Statistische Methoden

Das folgende Kapitel ist ein Exkurs, welches den Hintergrund der statistischen Methoden beleuchtet und als Nachschlagewerk für die bessere Verständlichkeit der genutzten statistischen Analysen dient.

Zur Analyse der Studien stehen eine Reihe statistischer Analysemethoden zur Auswahl, die entsprechend des Studiendesigns und der Datenqualität gewählt werden können. Hierbei müssen vor allem die jeweiligen Annahmen der Analysemethode berücksichtigt werden, um eine sinnvolle Interpretation des statistischen Modells zu erlauben (Field et al., 2012). Da in den vorliegenden empirischen Arbeiten meist nicht nur Zwischensubjekt-Faktoren, sondern auch Innersubjektfaktoren manipuliert werden, schränkt dies die Auswahl an möglichen statistischen Verfahren ein. In solchen Designs liefert jede Versuchsperson Daten zu mehreren Messpunkten, die dadurch voneinander abhängig sind, dass sie von der gleichen Person stammen. Parametrische Testverfahren setzen jedoch in der Regel die Unabhängigkeit der Datenpunkte voneinander voraus (Field et al., 2012), die in einem Messwiederholungsdesign nicht gegeben ist. Daher muss bei diesen Designs eine Adjustierung des statistischen Modells als Korrektur für die Abhängigkeit der Messwerte einer Person erfolgen. Klassischerweise wird hierzu in der Psychologie eine *Varianzanalyse* (ANOVA) mit Messwiederholung genutzt. Dieses Verfahren stellt einen Spezialfall des Allgemeinen Linearen Modells dar, welches bei diskreten Stufen der unabhängigen Variable angemessen ist (z.B. vor und nach einem Training). Eine Varianzanalyse (ANOVA) birgt allerdings auch einige Nachteile, wie beispiels-

weise einen unflexiblen Umgang mit fehlenden Werten oder eine fehlende Gewichtung von Probanden mit mehr oder weniger Messwerten.

Eine geeignete Alternative zur ANOVA ist in diesem Kontext ein gemischtes lineares Modell, *Linear Mixed-Effects Model* (LMM). Diese Analysemethode wird in der vorliegenden Arbeit genutzt. Da LMMs in der Psychologie erst in den letzten Jahren an Gewicht gewonnen haben, sollen sie im Folgenden näher beschrieben und im Vergleich zur ANOVA mit Messwiederholung diskutiert werden. Zusätzlich soll die Analysestrategie für die vorliegenden empirischen Arbeiten vorgestellt werden. Darüber hinaus sind in dieser Arbeit auch Zusammenhangsmaße von Bedeutung. Auch für Korrelationen gilt, dass eine Unabhängigkeit der Datenpaare angenommen wird. Diese Annahme wird bei Mehrfachmessungen innerhalb eines Probanden verletzt, weswegen in der vorliegenden Arbeit eine alternative Berechnung für Korrelationen gewählt wird, die Berechnung von Innersubjektkorrelationen. Da Innersubjektkorrelationen in der Psychologie jedoch nicht häufig verwendet werden, werden auch diese im Folgenden beschrieben und diskutiert.

B.1 Gemischtes lineares Modell - LMM

Das *Allgemeines Lineares Modell* (ALM), mit Sonderfällen wie der linearen Regression, dem LMM oder der ANOVA, gehört zu den varianzanalytischen Verfahren, in denen die Varianz von Messwerten (z.B. Unterschiede zwischen Personen) auf die Ausprägungen einer oder mehrerer Prädiktoren oder unabhängigen Variablen zurückgeführt werden soll (Bortz, 2005). Die Kernidee des Allgemeinen Linearen Modells (ALMs) besteht darin, dass Messwerte einer mindestens intervallskalierten, abhängigen Variable durch die Variation einer oder mehrerer Prädiktoren mit Hilfe einer linearen Zerlegung vorhergesagt werden können. Am Beispiel eines einzelnen Prädiktors, kann dieser Ansatz wie folgt formalisiert werden ¹,

$$\begin{aligned} Y_i &= \alpha + \beta \cdot X_i + \varepsilon_i \\ \varepsilon_i &\sim \text{Gaussian}(0, \sigma_\varepsilon^2) \end{aligned} \tag{B.1}$$

wobei Y_i den i 'ten Messwert der abhängigen Variable darstellt, der sich aus

¹Notation in Anlehnung an Nakagawa und Schielzeth (2013)

einem vorhersagbaren Wert $Y'_i = \alpha + \beta \cdot X_i$ und dem Fehlerterm ε_i zusammensetzt. Der vorhersagbare Wert Y'_i basiert auf einer linearen Gleichung, die von der Ausprägung des Prädiktors X_i , dem Intercept (Y-Achsenabschnitt) α und der Steigung β abhängt. Wenn X_i also den Wert 0 annimmt, dann wird Y'_i den Wert α annehmen. Steigt X_i um eins an, dann wird sich Y'_i um den Wert β verändern. Der Fehlerterm ε_i stellt die Abweichung des Messwertes Y_i von der Vorhersage Y'_i dar. Er ist sozusagen der Teil des Messwerts, der nicht durch das lineare Modell erklärt werden kann. Für den Fehlerterm wird angenommen, dass er aus einer Normalverteilung mit dem Mittelwert 0 und einer Standardabweichung von σ_ε stammt. Diese Annahme spiegelt das zufällige Ziehen der Messwerte aus der Grundgesamtheit wieder und basiert auf der Annahme, dass die Grundgesamtheit um den wahren Wert mit der Standardabweichung σ_ε streut.

Neben der Annahme der Normalverteilung der Fehlerterme besteht eine weitere Annahme des ALMs darin, dass die Messwerte unabhängig voneinander sind (Field et al., 2012). Diese Annahme ist verletzt, sobald eine Versuchsperson mehrere Messwerte liefert. So kann Person A beispielsweise grundsätzlich schneller antworten als Person B. Diese Systematik in den Messwerten, die durch die wiederholte Messung einer Person entsteht, sollte auch im statistischen Modell widerspiegelt werden, da dies die Güte der Vorhersage verbessern kann. LMMs als Spezialfall des ALM erlauben solche Systematiken statistisch durch einen oder mehrere Schätzwerte zu berücksichtigen. Diese Schätzwerte werden als *Random Effects* bezeichnet. Sie sind abzugrenzen von den sogenannten *Fixed Effects*, die die Effekte der Prädiktoren beschreiben, die über alle Versuchspersonen hinweg gleich sind.

Am oben genannten Beispiel mit einem Fixed Effect (durch einen Prädiktor) und mit einem Random Effect für die Messwiederholung innerhalb der gleichen Versuchsperson, kann der Ansatz eines LMMs wie folgt formalisiert werden:

$$\begin{aligned} Y_{ij} &= \alpha + \beta \cdot X_{ij} + \gamma_{0j} + \varepsilon_{ij} \\ \gamma_{0j} &\sim \text{Gaussian}(0, \sigma_\gamma^2) \\ \varepsilon_{ij} &\sim \text{Gaussian}(0, \sigma_\varepsilon^2) \end{aligned} \tag{B.2}$$

wobei Y_{ij} den i 'ten Messwert der j 'ten Person der abhängigen Variable darstellt, der sich aus dem vorhersagbaren Wert $Y'_{ij} = \alpha + \beta \cdot X_{ij}$, dem Random

Effect γ_{0j} der Person j und dem Fehlerterm ε_{ij} zusammensetzt. Ähnlich wie für den Fehlerterm ε_{ij} wird für den Random Effect γ_{0j} angenommen, dass er aus einer Normalverteilung mit Mittelwert 0 und einer Standardabweichung von σ_γ stammt. Auch diese Annahme spiegelt ein zufälliges Ziehen wider, hier das zufällige Ziehen der Versuchspersonen aus der Grundgesamtheit.

Durch die Berücksichtigung von Random und Fixed Effects können systematische Varianzen in Messwerten aufgeklärt werden, wobei Fixed Effects klassischerweise die Effekte sind, die von Interesse sind, und Random Effects der weiteren Varianzaufklärung dienen. Random Effects können weiter in zwei Kategorien unterteilt werden: *Random Intercepts* und *Random Slopes* (Bates, 2010; Winter, 2013). Im oben genannten Beispiel (Gleichung B.2) wurde ein Random Intercept eingeführt, der für jede Versuchsperson den Intercept α adjustiert. Es ist allerdings auch vorstellbar, dass sich der Wert der Steigung β zwischen Versuchspersonen unterscheidet. So könnte beispielsweise für alle Versuchspersonen ein positiver Zusammenhang zwischen X_{ij} und Y_{ij} bestehen, wobei dieser Zusammengang für Person A stärker ist als für Person B. Dies kann ebenfalls zu einer systematischen Varianz in den Daten führen, die mit Hilfe eines Random Slopes abgebildet wird. Dieser adjustiert die Steigung β für jede Versuchsperson. Neben der Schätzung von Random Intercepts und Random Slopes erlauben LMMs auch das Beachten der Korrelationen zwischen diesen Schätzungen im Rahmen des statistischen Modells (Bates, Kliegl et al., 2015).

Die bisher genannten Beispiele beziehen sich auf eine Messwiederholung innerhalb der gleichen Versuchsperson. In vielen Versuchsdesigns gibt es aber auch noch weitere Quellen für Random Effects. Ein Beispiel hierfür sind die Aufgaben oder Items die jede Versuchsperson bearbeiten muss (Winter, 2013). Auch hier liegt ein Messwiederholungsdesign vor, da jede Aufgabe pro Versuchsperson einmal und damit mehrfach zur Messung genutzt wird. Dabei ist es naheliegend anzunehmen, dass bestimmte Aufgaben einen spezifischen Effekt auf die abhängige Variable Y_{ij} haben, der über alle Versuchspersonen hinweg gleich ist. LMMs erlauben es, die Schätzung von Random Effects basierend auf verschiedenen Quellen von systematischer Varianz zu definieren, wie hier im Beispiel basierend auf der Messwiederholung innerhalb einer Versuchsperson und auf der Messwiederholung innerhalb einer Aufgabe. Auch für

Random Effects, die durch die Aufgaben entstehen, gilt dabei die Annahme, dass die Auswahl der Aufgaben eine zufällige Ziehung aus allen möglichen Aufgaben darstellt. Damit gilt auch hier, dass Random Effects über die Aufgabe aus einer Normalverteilung mit dem Mittelwert 0 und einer spezifischen Streuung stammen.

Für ein einfaktorielles Design mit dem Faktor X_{ijk} , in dem jede Versuchsperson j jede Aufgabe k i -mal löst, könnte ein LMM also vier Random Effects enthalten: einen Random Intercept für die Versuchspersonen γ_{0j} , einen Random Slope für die Versuchspersonen γ_{1j} , einen Random Intercept für die Aufgabe ω_{0k} und einen Random Slope für die Aufgaben ω_{1k} . Die entsprechende Formalisierung für ein solches Design lautet:

$$\begin{aligned}
 Y_{ijk} &= \alpha + \gamma_{0j} + \omega_{0k} + (\beta + \gamma_{1j} + \omega_{1k}) \cdot X_{ijk} + \varepsilon_{ijk} \\
 \gamma_{0j} &\sim \text{Gaussian}(0, \sigma_{\gamma_0}^2) \\
 \gamma_{1j} &\sim \text{Gaussian}(0, \sigma_{\gamma_1}^2) \\
 \omega_{0k} &\sim \text{Gaussian}(0, \sigma_{\omega_0}^2) \\
 \omega_{1k} &\sim \text{Gaussian}(0, \sigma_{\omega_1}^2) \\
 \varepsilon_i &\sim \text{Gaussian}(0, \sigma_{\varepsilon}^2)
 \end{aligned} \tag{B.3}$$

Hier wird deutlich, dass die Random Intercepts γ_{0j} und ω_{0k} eine Adjustierung des Intercepts α darstellen, wobei für jede Versuchsperson und für jede Aufgabe eine spezifische Adjustierung erfolgt. Ähnliches gilt für die Anpassung der Steigung β durch die Random Slopes γ_{1j} und ω_{1k} . Für alle Random Effects gilt auch hier die Annahme, dass sie aus einer Normalverteilung mit dem Mittelwert 0 stammen.

Eine andere Formalisierungsart des gleichen Sachverhalts, die in der vorliegenden Arbeit genutzt wird, basiert auf der Schreibweise des *lme4*-Pakets (Bates, Mächler et al., 2015) für *R*:

$$AV \sim 1 + UV + (1 + UV|ID) + (1 + UV|Aufgabe) \tag{B.4}$$

wobei die abhängige Variable (AV) durch einen Intercept von 1 sowie die unabhängige Variable (UV) vorhergesagt wird. Die Vorhersage wird ergänzt durch den Random Intercept und Random Slope für jede Versuchsperson ($1 + UV|ID$) und die Random Effects für jede Aufgabe ($1 + UV|Aufgabe$). Der

Fehlerterm ε_{ijk} für die Residuen muss in dieser Formalisierung nicht explizit genannt werden, wird aber in der Berechnung, wie in Formel B.3 angegeben, berücksichtigt.

B.1.1 Abgrenzung zur ANOVA mit Messwiederholung

Das ALM als Grundlage für sowohl das LMM und die ANOVA wirft die Frage auf, worin sich ein LMM und eine ANOVA unterscheiden. Eine wesentlicher Unterschied besteht in der Anzahl der Random Effects. Während in der ANOVA mit Messwiederholung nur für Random Effects basierend auf einer Zufallsvariable kontrolliert werden kann (typischerweise Versuchspersonen), können mit LMMs mehrere Quellen von Random Effects berücksichtigt werden (zum Beispiel Versuchspersonen und Aufgaben).

Ein weiterer wichtiger Unterschied zwischen einer ANOVA mit Messwiederholung und einem LMM liegt in der Anzahl der Datenpunkte, die pro Zelle des Designs vorliegen müssen. Eine ANOVA mit Messwiederholung setzt voraus, dass jede Versuchsperson in jeder Zelle des Designs genau einen Messwert geliefert hat. Dies hat zwei Konsequenzen: Zum einen stellen Datensätze mit einem oder mehreren fehlenden Datenpunkten ein Problem für die Analyse mit einer ANOVA mit Messwiederholung dar. Daher müssen entweder die betroffenen Versuchsperson aus der Analyse ausgeschlossen oder die fehlenden Werte ersetzt werden. Zum anderen muss über mehrere Datenpunkte gemittelt werden, wenn jede Zelle mehrfach pro Versuchsperson gemessen wurde. Auch dies hat Auswirkungen, wenn fehlende Werte auftreten. So werden alle Zellen mit dem gleichen Gewicht in die Analyse eingehen, auch wenn sich die Sicherheit, mit der der Mittelwert berechnet wurde, unterscheiden kann (z.B.: eine der Zellen basiert auf einer Mittelung über 6 Datenpunkte, wohingegen eine andere Mittelung auf nur 2 Datenpunkten beruht). Bei der Analyse mit LMMs ist keine Mittelung notwendig, da hier eine Trial-by-Trial-Analyse erfolgt. Hierdurch können auch Datensätze von Versuchspersonen, deren Datensatz nicht vollständig ist, in ihrer unvollständigen Form in die Analyse eingehen. LMMs sind damit im Gegensatz zur ANOVA mit Messwiederholung robuster gegenüber fehlenden Datenpunkten und erlauben mehrere Messwerte pro Zelle des Designs und pro Versuchsperson. Gehen Versuchspersonen mit einer unterschiedlichen Anzahl an Messpunkten in die Analyse mit einem LMM ein, so

erfolgt eine unterschiedliche Gewichtung.

Zusammengefasst lässt sich sagen, dass die Berücksichtigung von mehreren Quellen für Random Effects in einem LMM dazu führt, dass mehr Fehlervarianz als in einer klassischen ANOVA mit Messwiederholung aufgeklärt werden kann (Kliegl, Wei, Dambacher, Yan & Zhou, 2011), wodurch die statistische Power steigt (Bates, Kliegl et al., 2015). Dies hilft, die Gefahr eines Fehlers zweiter Art zu verringern. Darüber hinaus reduziert sich durch die Aufnahme von Random Intercepts und Random Slopes auch die Gefahr, einen experimentellen Effekt zu überschätzen (Fehler erster Art; Schielzeth & Forstmeier, 2009). Ein weiterer Vorteil liegt in der Trial-by-Trial-Analyse, da sie wesentlich flexibler gegenüber fehlenden Daten ist und besser die Häufigkeiten der Messung innerhalb der Daten widerspiegelt. Durch eine Trial-by-Trial-Analyse können darüber hinaus Kovariaten berücksichtigt werden, die sich innerhalb einer Versuchsperson zwischen den Trials unterscheiden. Dies ist in der klassischen Kovarianzanalyse (ANCOVA) mit Messwiederholung nicht möglich.

B.1.2 Selektion, Kodierung und Interpretation der Fixed Effects

Die Auswahl der Fixed Effects im Rahmen von LMMs sollte die Hypothesen in Bezug auf unabhängige Variablen bzw. Prädiktoren widerspiegeln. Der Wertebereich eines möglichen Prädiktors X_i (Gleichungen B.1, B.2 und B.3) hängt dabei von seinem Skalenniveau ab. Während bei einem intervall- oder verhältnisskalierten Prädiktor der Wert der Variable X_i direkt in die Gleichung B.1 eingesetzt werden kann, ist dies bei einem niedrigeren Skalenniveau nicht sinnvoll.

Um die Stufen einer nominalskalierten, unabhängigen Variable X als linearen Prädiktor nutzbar zu machen, muss eine Umcodierung der Variable erfolgen, wie es beispielsweise Teil der ANOVA ist. Beispiele für eine solche Umcodierung sind die Dummycodierung, die Effektcodierung und die Kontrastcodierung (Bortz & Schuster, 2010). Sowohl bei der Dummycodierung als auch bei der Effektcodierung werden bei k verschiedenen Gruppen, die die unabhängige Variable X umfasst, $k - 1$ Dummy-Variablen erstellt, die als neue Prädiktoren zur Vorhersage von Y'_i genutzt werden. Diese Dummy-Variablen

erlauben je nach Kodierung den Einfluss einer Stufe oder eines spezifischen Effekts zu testen (Schad et al., 2018). So können beispielsweise über einen Summenkontrast (*sum contrast*) für einen Faktor mit $k = 3$ Stufen die ersten beiden Stufen durch jeweils eine Dummyvariable gegen den Gesamtmittelwert getestet werden (Schad et al., 2018). Der dritte Vergleich wäre redundant, da der Unterschied der dritten Stufe zum Gesamtmittelwert aus den ersten beiden Vergleichen erschlossen werden kann. Wird statt eines Summenkontrasts ein wiederholter Kontrast (*repeated contrast* oder *difference contrast*) genutzt, testet die erste Dummyvariable den Unterschied zwischen Stufe eins und zwei, während die zweite Dummyvariable den Unterschied zwischen Stufe zwei und drei testet (Schad et al., 2018).

Die Art der Kodierung der Dummyvariablen und die Wahl ihrer Werte beeinflusst die mögliche Interpretation der statistischen Kennwerte des linearen Modells. Abbildung B.1 stellt drei Möglichkeiten der Kodierung einer zweistufigen unabhängigen Variable dar. Für die Betrachtung der Kontraste ist hier also nur eine Dummyvariable notwendig. Alle drei dargestellten Möglichkeiten nutzen einen Abstand von eins zwischen den Stufen. Durch diesen Abstand kann das β -Gewicht b im linearen Modell als Unterschied zwischen den Mittelwerten

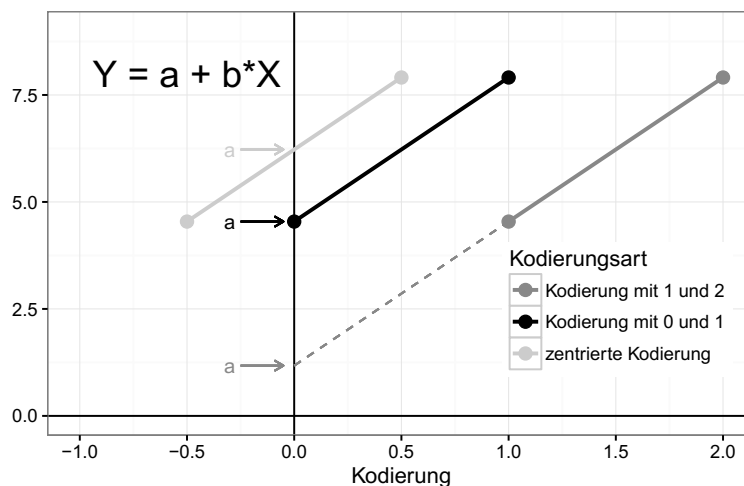


ABBILDUNG B.1: Verschiedene Kodierungsarten für zweistufige, nominalskalierte Prädiktoren in einer linearen Regression. Die Abbildung verdeutlicht die inhaltliche Entsprechung der Kennwerte der linearen Gleichung in Abhängigkeit von der Kodierung.

der Stufen interpretiert werden. Der Intercept a ist jedoch je nach Kodierungsart unterschiedlich zu interpretieren: Er entspricht dem Gesamtmittelwert bei einer zentrierte Kodierung, dem Mittelwert einer der Stufen bei einer Kodierung mit 0 und 1 oder einem nur nach Kontext interpretierbaren Wert bei einer Kodierung, in der keine der Kodierungsstufen mit 0 gleichgesetzt wird. Um in den vorliegenden Analysen eine leichte und direkte Interpretierbarkeit der statistischen Kennwerte zu ermöglichen, wird für alle nominalskalierten, unabhängigen Variablen bei Analysen mit LMMs eine zentrierte Kodierung durchgeführt, bei der die Unterschiede zwischen den Stufen mit einem Abstand von eins kodiert wird. Dies gilt auch für Variablen mit mehr als zwei Stufen. Das heißt der Intercept a kann in allen Analysen als Gesamtmittelwert der unabhängigen Variable interpretiert werden. Verhältnisskalierte Prädiktoren werden zu diesem Zweck ebenfalls zentriert. Darüber hinaus kann der Steigungsparameter b immer als Unterschied zwischen den jeweils angegebenen Stufen interpretiert werden.

B.1.3 Selektion der Random Effects

Bei der Entscheidung, welche Random Effects im statistischen Modell berücksichtigt werden sollen, stellt sich zunächst die Frage, welche Random Effects geschätzt werden können und welche Struktur der Random Effects sinnvoll ist. Dabei gilt es abzuwägen, wie viele Schätzparameter notwendig sind, um möglichst viel Varianz aufzuklären und gleichzeitig eine Überparametrisierung des statistischen Modells zu vermeiden. Gleichzeitig muss entschieden werden, welche Variablen für den Analysierenden von Interesse sind, da der Einfluss einzelner Variablen häufig sowohl als Fixed als auch als Random Effect in der statistischen Modellspezifikation berücksichtigt werden könnte.

In einem Messwiederholungsdesign besteht die minimale Random Effect Struktur in einem Random Intercept, um die Abhängigkeit zwischen den Messwiederholungen aufzufangen. Die maximale Struktur besteht aus einem Random Intercept, allen möglichen Random Slopes und den Korrelationen zwischen den Random Effects. Random Slopes können sowohl für Haupt- als auch für Interaktionseffekte berücksichtigt werden. Jede Variable mit Messwiederholung auf Seiten der Fixed Effects, kann also auch Teil der Random Effects Struktur in Rahmen eines Random Slopes sein. Bei größeren Designs kann es

daher zu sehr komplexen Random Effect Strukturen kommen.

Bates, Kliegl et al. (2015) argumentieren, dass sehr komplexe Random Effect Strukturen zu einer Überparametrisierung des Modells oder im schlimmsten Fall zu einer fehlenden Konvergenz der Schätzung führen können. Daher widersprechen sie der Empfehlung von Barr, Levy, Scheepers und Tily (2013) immer die maximale Random Effect Struktur zu nutzen. Stattdessen schlagen sie eine daten-basierte Selektion der Random Slopes vor, bei der nur diejenigen Random Effects im Model belassen werden, die einen Beitrag zur Varianzaufklärung leisten. Bates, Kliegl et al. (2015) empfehlen zunächst mit Hilfe einer Hauptkomponentenanalyse zu ermitteln, wie viele Random Effects keine weitere Varianz aufklären, um die grundlegende Dimensionalität der Random Effect Struktur zu überblicken und um unnötige Random Effects aus dem Modell zu nehmen. Dabei sollten höherwertige Terme vor niederwertigen Termen entfernt werden (z.B. zweifach Interaktion vor Haupteffekten). Das resultierende Modell kann erneut mit einer Hauptkomponentenanalyse betrachtet werden. Danach empfehlen Bates, Kliegl et al. (2015) ein iteratives Verfahren, in dem reduzierte Modelle mit dem jeweils vorangegangenen Modell verglichen werden. Die Reduzierung erfolgt dabei zunächst in Bezug auf die Korrelationen zwischen den Random Effects und anschließend in Bezug auf alle anderen Random Effects. Abschließend empfehlen Bates, Kliegl et al. (2015) das reduzierte Modell erneut um die Korrelationen zwischen den Random Effects zu erweitern. Dadurch kann geprüft werden, ob die Berücksichtigung der Korrelationen einen substantiellen Beitrag zur Random Effect Struktur leistet. Durch das Vorgehen wird eine Random Effect Struktur ausgewählt, die durch die vorliegenden Daten gestützt wird. Es ist damit ein datengetriebenes Verfahren.

Die Aufnahme von Random Effects in das statistische Modell dient neben der Korrektur bei Verletzung der Unabhängigkeit der Daten vor allem dazu, möglichst viel Varianz aufzuklären und dadurch einen möglichst starken Test für die Fixed Effects zu gewährleisten (Bates, Kliegl et al., 2015). Daher wird in der vorliegenden Arbeit nicht weiter auf ihre Interpretation eingegangen, obwohl sie in manchen Kontexten inhaltlich relevant sein können.

B.1.4 Normalverteilung der Residuen und Transformationen

Neben der Annahme von Unabhängigkeit der Datenpunkte voneinander setzt das ALM auch eine Normalverteilung der Residuen voraus (Bortz, 2005). Dies kann beispielsweise mit Hilfe eines Shapiro-Wilk-Tests geprüft werden (Field et al., 2012). Liegt keine Normalverteilung der Residuen vor, so kann dies die Interpretierbarkeit der Ergebnisse einschränken. Eine Normalverteilung kann aber in bestimmten Fällen durch eine Transformation der abhängigen Variable Y erzeugt werden (Bortz, 2005; Field et al., 2012). In der vorliegenden Arbeit wird für alle berichteten Modelle geprüft, ob die Residuen bei jeweils gegebener Modellspezifizierung normalverteilt sind oder durch eine Transformation der abhängigen Variable näher an eine Normalverteilung gebracht werden können. Diese Prüfung wird mit Hilfe der Box-Cox-Power-Transformation aus dem *MASS*-Paket (Venables & Ripley, 2002) für R überprüft. Dabei wird der optimale Parameter λ geschätzt, für den eine Transformation der Form

$$Y' = \begin{cases} \frac{(Y^\lambda - 1)}{\lambda} & , \text{ für } \lambda \neq 0 \\ \ln(Y) & , \text{ für } \lambda = 0 \end{cases} \quad (\text{B.5})$$

die Residuen näher an eine Normalverteilung bringt. Alle Transformationen behalten hier die Rangreihenfolge der nicht-transformierten Daten bei. Auf Basis des geschätzten λ s wird eine Standardtransformation gewählt: Für $\lambda > 0.5$ keine Transformation, für $-0.5 \leq \lambda \leq 0.5$ eine $\ln$ -Transformation mit $\lambda = 0$ und für $\lambda < -0.5$ eine reziproke Transformation mit $\lambda = -1$. Diese Standardtransformationen (vgl. Field et al., 2012) werden anstelle einer Transformation auf Basis des ermittelten Werts von λ genutzt, um eine bessere Vergleichbarkeit zwischen Studien und Interpretierbarkeit der transformierten Werte zu gewährleisten.

B.1.5 t-Werte, Signifikanzniveaus, Profiling, Varianzaufklärung und Power

Für die Analyse der LMMs wird in der vorliegenden Arbeit das *lme4*-Paket (Bates, Mächler et al., 2015) genutzt, welches für jeden Fixed Effect einen t -Wert ausgibt. Zu diesem t -Wert wird kein p -Wert angegeben, da die Frei-

heitsgrade in einem LMM nur geschätzt werden können (Kliegl, Masson & Richter, 2010). Auf Grund der Annäherung an eine Normalverteilung bei einer großen Menge an Datenpunkten können aber t -Werte, die in den äußeren 5% der t -Verteilung liegen ($|t| > 2$) auf einem Signifikanzniveau von $\alpha = .05$ als signifikant angenommen werden (Kliegl et al., 2010, S. 4).

Zusätzlich werden für jeden Fixed Effekt 95%-Konfidenzintervalle der Parameterschätzung des β -Gewichts angegeben. Diese Konfidenzintervalle werden auf Basis des `profile` Befehls (R Core Team, 2018) ermittelt, in dem systematisch die einzelnen Schätzparameter des Modells variiert werden, um den Fit des Modell bei einem bestimmten Parameterwert mit dem Fit des optimalen Modells zu vergleichen (Bates, 2010, S. 16). Sie sind ein Maß der Sensitivität des Modellfits auf Veränderungen in den Schätzparametern (Bates, 2010) und geben damit einen Einblick darin, wie genau die Schätzung der β -Gewichte in den berichteten LMMs sind (Bates, 2011).

Um die Größe der Varianzaufklärung durch das jeweilige Modell aufzuzeigen, werden für jedes LMM zwei verschiedene R^2 -Werte nach Nakagawa und Schielzeth (2013) und Johnson (2014) berichtet: marginales R^2 ($R^2_{GLMM(m)}$) und konditionales R^2 ($R^2_{GLMM(c)}$). Das marginale R^2 beschreibt die Varianz, die alleine durch die Fixed Effekte aufgeklärt wird (Bartoń, 2018). Das konditionale R^2 beschreibt die Varianzaufklärung des gesamten Modell mit Fixed und Random Effects (Nakagawa & Schielzeth, 2013).

Zusätzlich wird für jeden Effekt der unabhängigen Variablen, der mit einem LMM untersucht wird, eine post-hoc Poweranalyse durchgeführt. Die Power $1-\beta$ gibt dabei an, wie wahrscheinlich es für einen tatsächlich existierenden Effekt ist, in einem Testverfahren als signifikant klassifiziert zu werden (Green & MacLeod, 2016). β steht hier für die Auftretenswahrscheinlichkeit eines Fehlers 2. Art und nicht für das β -Gewicht, welches als Steigungsparameter in einem LMM den Effekt einer Variable beschreibt. Die Berechnung der Power erfolgt in der vorliegenden Arbeit post-hoc auf Basis des gefundenen Unterschieds in den Daten, der untersuchten Stichprobengröße und einem kritischen Signifikanzniveau von $\alpha = 5\%$. Hierzu wird das *simr*-Paket (Green & MacLeod, 2016) genutzt, welches die Power auf Basis von Monte Carlo Simulationen ermittelt.

Die Power wird aus zwei Gründen berichtet. Zum einen kann eine kleine Power als Indiz dafür gesehen werden, dass ein Effekt, falls es ihn gibt, nicht

sehr wahrscheinlich mit der vorliegenden Stichprobengröße gefunden werden konnte (Bortz, 2005). Vor allem bei kleinen Effekten führt die begrenzte Versuchspersonenanzahl häufig zu einem unterpowertem Testverfahren, was bei der Interpretation zu berücksichtigen ist. Jedoch darf auch bei ausreichender Power und nicht-signifikanten Ergebnissen keine Interpretation im Sinne der Nullhypothese erfolgen (Hoenig & Heisey, 2001). Zum anderen wird die Power für mögliche Replikationen berichtet, damit zusammen mit den Kennwerten für das β -Gewicht und für dessen Standardfehler die benötigte Stichprobengröße für folgende Arbeiten abgeleitet werden kann.

B.2 Korrelationen bei Messwiederholung

Korrelationen sind in der Psychologie ein weit verbreitetes Maß für den linearen Zusammenhang zwischen zwei Variablen. Sie werden in Form des Korrelationskoeffizienten r ausgedrückt, der im Beispiel des *Pearsons Produkt-Moment Koeffizienten* (Bortz, 2005) das Ausmaß des linearen Zusammenhangs zwischen den Variablen X und Y als das Verhältnis zwischen der Kovarianz der beiden Variablen $cov(X, Y)$ zu dem Produkt der Streuungen der Variablen σ_X und σ_Y beschreibt (Bortz, 2005). Der Korrelationskoeffizient r kann dabei einen Wert im Intervall $[-1, 1]$ annehmen, wobei der Zusammenhang zwischen den Variablen umso größer ist, je größer $|r|$ ist.

Für die Berechnung von r nach Pearson müssen mindestens intervallskalierte Wertepaare vorliegen, bestehend aus einem X- und einem Y-Wert pro Mess-einheit (z.B. pro Versuchsperson). Um das so gewonnene r einem Signifikanztest unterziehen zu können, müssen die Wertepaare darüber hinaus unabhängig voneinander aus einer bivariat normalverteilten Grundgesamtheit gezogen worden sein.

Im Falle eines Messwiederholungsdesigns ist diese Unabhängigkeit jedoch nicht gegeben, da das Werte-Paar eines Probanden zu einem Messzeitpunkt das Werte-Paar des gleichen Probanden zu einem anderen Messzeitpunkt beeinflussen kann, wohingegen es keinen Einfluss auf die Werte-Paare eines anderen Probanden hat. Dadurch ist die Variabilität der Wertepaare zwischen Probanden größer als innerhalb eines Probanden (Bland & Altman, 1994). Ähnliches gilt für verschiedene Messzeitpunkte oder Aufgaben, zu denen die

Werte-Paare erhoben werden: Werte-Paare, die zum gleichen Messzeitpunkt oder nach der gleichen Aufgabe erhoben wurden, sollten eine geringere Variabilität aufweisen als Werte-Paare, die zu unterschiedlichen Messzeitpunkten oder nach unterschiedlichen Aufgaben gemessen wurden. Mit anderen Worten finden sich bei der Betrachtung von Korrelationen im Messwiederholungsdesign die gleichen Probleme durch Abhängigkeiten in den Daten wieder, wie sie in Abschnitt B.1 bereits im Bezug auf LMMs diskutiert wurden.

Bland und Altman (1994, 1995a, 1995b) stellen zwei verschiedene Lösungsvorschläge vor, die in Abhängigkeit der Fragestellung mit diesen Problemen umgehen. Die erste mögliche Fragestellung lautet: Zeigen Probanden, die auf Variable X hohe Werte zeigen, auch hohe Werte auf Variable Y? Zur Beantwortung dieser Frage kann der Zusammenhang zwischen den Variablen X und Y unabhängig von Messwiederholungsfaktoren wie Person oder Messzeitpunkt betrachtet werden. Dazu wird klassischerweise über die Messwiederholung pro Versuchsperson gemittelt, so dass nur ein Werte-Paar pro Person vorliegt (Bland & Altman, 1994). Das Verfahren kann auch angewendet werden, wenn die Versuchspersonen unterschiedlich viele Wertepaare generiert haben und dadurch die gemittelten Wertepaare auf einer unterschiedlichen Anzahl an Datenpunkten basieren. Bland und Altman (1995b) empfehlen in diesem Fall eine gewichtete Korrelation zu berechnen.

Die zweite mögliche Fragestellung lautet: Führen Veränderungen der Variable X innerhalb eines Probandens über mehrere Messzeitpunkte hinweg zu einer parallelen Veränderung in der Variable Y? Um diese Frage beantworten zu können, muss der Zusammenhang innerhalb des Messwiederholungsfaktors Person betrachtet werden. Die Fragestellung ist vor allem dann relevant, wenn jeder Proband unter sehr verschiedenen Bedingungen Messdaten geliefert hat und sich der Forscher für den Zusammenhang zwischen X und Y über diese Situation hinweg interessiert. Gerade im Kontext von UX, in dem der Zusammenhang zwei verschiedener Bewertungsmaße von Interesse sein kann, ist diese Fragestellung also hoch relevant.

Bland und Altman (1995a) schlagen für die Berechnung des entsprechenden Korrelationskoeffizienten ein statistisches Verfahren vor, in dem ein lineares Modell mit Y als Kriterium sowie mit X und einer faktorisierten Variable ID für die Probanden als Prädiktoren berechnet wird. Dieses Modell ähnelt teil-

weise einem Random-Intercept-Modell, bei dem der Effekt der Variable X (die Steigung) für alle Probanden gleich ist, der Intercept für jeden Probanden aber variieren kann. Die von Bland und Altman (1995a) vorgeschlagene Berechnung unterscheidet sich aber von der Berechnung eines LMMs, da hier der Effekt der einzelnen Versuchspersonen nicht als Random Effect, sondern als kontrastkodierter Fixed Effect in das Modell miteingeht. Der Korrelationskoeffizient r wird abschließend basierend auf den Quadratsummen des Effekts der Variable X und der Residuen berechnet (siehe Gleichung B.6). Im Folgenden wird diese Art der Korrelationsberechnung Innersubjektkorrelation genannt.

$$r = \frac{\text{Quadratsumme von } X}{\text{Quadratsumme von } X + \text{Quadratsumme der Residuen}} \quad (\text{B.6})$$

In Abbildung B.2 sind die beiden verschiedenen Arten der Berechnungen von r mit Hilfe simulierter Daten beispielhaft dargestellt. In beiden Grafiken der Abbildung sind die gleichen Datenpunkte zu sehen, wobei die Farben die unterschiedlichen Probanden kodieren. In der linken Grafik sind zusätzlich die Mittelwerte der einzelnen Probanden durch Kreuze markiert, die zur Berechnung der Korrelation zwischen Probanden genutzt wurden. Hier zeigt sich eine positive Korrelation mit $r = .33$, d.h. Probanden, die im Mittel hohe Werte auf der Variable X zeigen, zeigen im Mittel auch hohe Werte auf der Variable Y. In der rechten Grafik ist die Korrelation innerhalb der Probanden dargestellt: Je höher die Werte der Probanden in der Variable X sind, desto niedriger sind ihre Werte in der Variable Y. Dies zeigt sich in einem negativen Korrelationskoeffizienten von $r = -.85$. Die beispielhaften Daten zeigen, dass die verschiedenen Arten der Berechnung von Korrelationskoeffizienten in einem Messwiederholungsdesign je nach Datenlage zu stark unterschiedlichen Schlussfolgerungen führen können. Welche Analyse jedoch für einen spezifischen Datensatz angemessen ist, hängt jeweils von der Fragestellung ab.

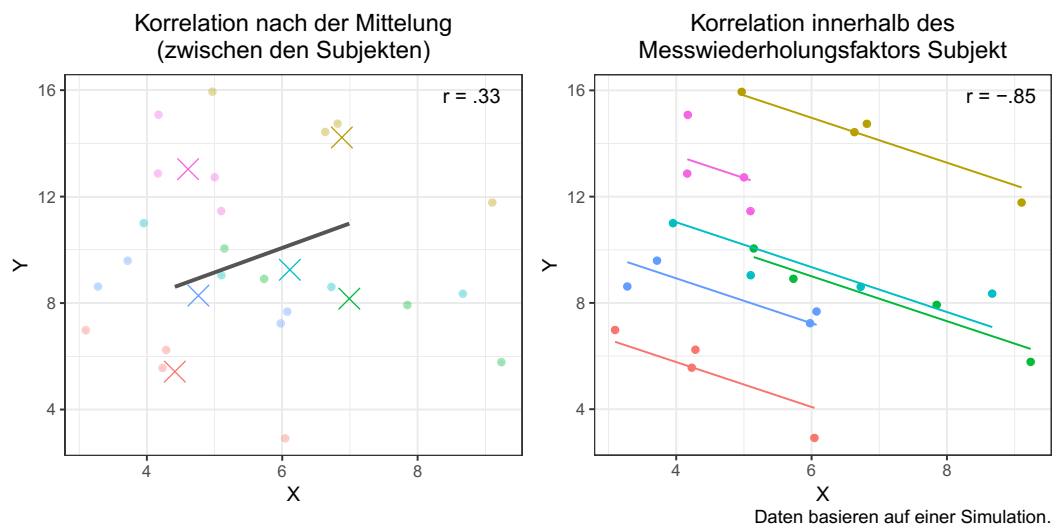


ABBILDUNG B.2: Unterschied zwischen Korrelationen basierend auf gemittelten Werten und Innersubjektkorrelationen.

Anhang C

Zusammenfassung aller Hypothesen

C.0.1 Unterschiedshypothesen

*Hypothese X.1: Die **Aufgabendauer** ist signifikant länger in der Gruppe mit lautem Denken als in der Gruppe ohne lautes Denken. Dies zeigt sich in einer negativen Steigung: $\beta_{\text{lautesDenken}} < 0$.*

*Hypothese X.2: Die **Aufgabendauer** ist signifikant kürzer in der Gruppe mit hoher Usability als in der Gruppe mit niedriger Usability. Dies zeigt sich in einer positiven Steigung: $\beta_{\text{Usability}} > 0$.*

*Hyp. X.9 (i) bis (iv): Die (i) **wahrgenommene Usability**, (ii) **wahrgenommene Attraktivität**, (iii) **Valenz** und das (iv) **Arousal** ist signifikant größer in der Gruppe mit hoher Usability als in der mit niedriger Usability. Dies zeigt sich in negativen Steigungen: $\beta_{\text{Usability}} 's < 0$.*

*Hypothese X.10: Die **geschätzte Aufgabendauer** ist signifikant kürzer in der Gruppe mit hoher Usability als in der Gruppe mit niedriger Usability. Dies zeigt sich in einer positiven Steigung: $\beta_{\text{Usability}} > 0$.*

Hypothese X.11: Die **PTRs** sind signifikant kleiner in der Gruppe mit lautem Denken als in der Gruppe ohne lautes Denken. Dies zeigt sich in einer positiven Steigung: $\beta_{\text{lautesDenken}} > 0$.

Hypothese X.12 (i) bis (ii): In der Gruppe mit hoher Usability zeigt sich ein signifikant geringerer Wert als in der Gruppe mit niedriger Usability in Bezug auf (i) die **Dauerbewertung**, (ii) die **Zeitflussbeurteilung**. Dies zeigt sich in positiven Steigungen: $\beta_{Usability} 's > 0$.

Hypothese X.14: Die **Dauerschätzung** ist signifikant kleiner in den Durchgängen mit kürzeren objektiven Dauern als in den Durchgängen mit längeren objektiven Dauern: Dies zeigt sich in einer positiven Steigung: $\beta_{Dauer} > 0$.

C.0.2 Zusammenhangshypothesen

Hypothese X.3 (i) bis (ii): Die Aufgabendauer ist umso größer, je größer die (i) geschätzte Aufgabendauer, (ii) Dauerbewertung, (iii) Zeitflussbeurteilung ist. Dies zeigt sich in positiven Korrelationkoeffizienten: $r's > 0$.

Hypothese X.4 (i) und (ii): Die geschätzte Aufgabendauer ist umso größer, je größer die (i) Dauerbewertung, (ii) Zeitflussbewertung ist. Dies zeigt sich in positiven Korrelationkoeffizienten: $r's > 0$.

Hypothese X.5 (i) bis (iv): Die wahrgenommene Usability ist umso geringer, je größer die zeitbezogene Variable (i) Aufgabendauer, (ii) geschätzte Aufgabendauer, (iii) Dauerbewertung, (iv) Zeitflussbeurteilung ist. Dies zeigt sich in negativen Korrelationkoeffizienten: $r's < 0$.

Hypothese X.6 (i) bis (iv): Die wahrgenommene Attraktivität ist umso geringer, je größer die zeitbezogene Variable (i) Aufgabendauer, (ii) geschätzte Aufgabendauer, (iii) Dauerbewertung, (iv) Zeitflussbeurteilung ist. Dies zeigt sich in negativen Korrelationkoeffizienten: $r's < 0$.

Hypothese X.7 (i) bis (iv): Die Valenz ist umso geringer, je größer die zeitbezogene Variable (i) Aufgabendauer, (ii) geschätzte Aufgabendauer, (iii) Dauerbewertung, (iv) Zeitflussbeurteilung ist. Dies zeigt sich in negativen Korrelationkoeffizienten: $r's < 0$.

Hypothese X.8 (i) bis (iv): Das Arousal korreliert mit der zeitbezogenen Variable (i) Aufgabendauer, (ii) geschätzte Aufgabendauer, (iii) Dauerbewertung, (iv) Zeitflussbeurteilung ist. Dies zeigt sich in Korrelationkoeffizienten ungleich Null: $r's \neq 0$.

Hypothese X.13: Die Zeitflussbeurteilung zeigt umso größere Werte, je größer die Werte der Dauerbewertung sind. Dies zeigt sich in einem positiven Korrelationkoeffizienten: $r > 0$.

Anhang D

Anhang Studie I

D.1 Versuchsmaterialien

D.1.1 Darstellung der Ausbalancierung

TABELLE D.1: Ausbalancierung Studie I.

Usability	lautes Denken	E A B C	E B C A	E C A B	Σ
hoch	mit	N = 7	N = 6	N = 4	17
hoch	ohne	N = 4	N = 4	N = 6	14
niedrig	mit	N = 6	N = 5	N = 4	15
niedrig	ohne	N = 4	N = 5	N = 6	15
Σ		21	20	20	61

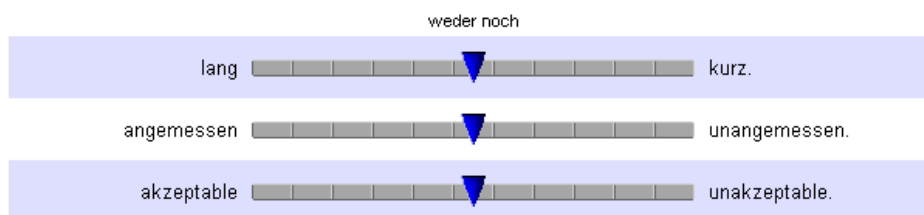
D.1.2 Fragebogen

Zeitschätzung

Bitte schätzen Sie die Zeit, die sie für die letzte Aufgabe gebraucht haben:

Minuten & Sekunden

Die Zeit, die ich für die Bearbeitung der letzten Aufgabe gebraucht habe, empfand ich als



Geben Sie bitte an, wie Sie die die Webseite nach der Durchführung der Aufgabe bewerten.

Bewegen Sie dazu bitte die Pfeilspitze an die Stelle, die Ihrer Gesamtbewertung der Webseite am ehesten entspricht.

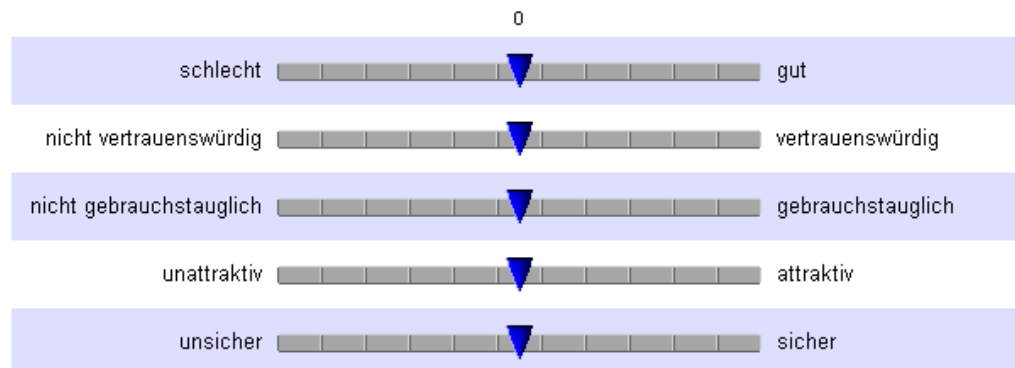


ABBILDUNG D.1: Selbstentwickelte Skalen in Studie I.

D.2 Poweranalysen

TABELLE D.2: Poweranalysen Studie I.

Hyp.	AV	UV	Power	95% Konf.
1	Aufg.dauer	laut. Denken	97%	[95%, 98%]
2	Aufg.dauer	Usability	100%	[100%, 100%]
–	Aufg.dauer	laut. D.:Usab.	16%	[14%, 19%]
–	gesch. Aufg.dauer	laut. Denken	7%	[6%, 9%]
–	gesch. Aufg.dauer	Usability	97%	[96%, 98%]
–	gesch. Aufg.dauer	laut. D.:Usab.	16%	[14%, 19%]
–	PTR	laut. Denken	65%	[62%, 68%]
–	PTR	Usability	6%	[5%, 8%]
–	PTR	laut. D.:Usab.	5%	[4%, 7%]
–	Dauerbewertung	laut. Denken	12%	[10%, 14%]
–	Dauerbewertung	Usability	100%	[100%, 100%]
–	Dauerbewertung	laut. D.:Usab.	11%	[9%, 13%]
–	wahrg. Usability	laut. Denken	6%	[4%, 7%]
9(i)	wahrg. Usability	Usability	100%	[100%, 100%]
–	wahrg. Usability	laut. D.:Usab.	5%	[3%, 6%]
–	wahrg. Attrakt.	laut. Denken	5%	[4%, 7%]
9(ii)	wahrg. Attrakt.	Usability	100%	[100%, 100%]
–	wahrg. Attrakt.	laut. D.:Usab.	7%	[5%, 9%]
–	Valenz	laut. Denken	6%	[5%, 8%]
9(iii)	Valenz	Usability	79%	[77%, 82%]
–	Valenz	laut. D.:Usab.	30%	[27%, 33%]
–	Arousal	laut. Denken	7%	[6%, 9%]
9(iv)	Arousal	Usability	40%	[37%, 43%]
–	Arousal	laut. D.:Usab.	55%	[51%, 58%]

ANMERKUNG: Schätzung der Power und Konfidenzintervalle auf Basis von $n = 1000$ Simulationen.

D.3 Hauptkomponentenanalyse der Items zur Dauerbewertung für die Messzeitpunkte 2. und 3. Aufgabe und die Aufgaben A bis C

TABELLE D.3: Zusammenfassung der Komponentenladungen beim zweiten und dritten Messzeitpunkt nach der Exploration.

zweiter Messzeitpunkt		
Bartlett-Test	$\chi^2 (3) = 81.80, p < .001$	
KMO	.70	
		Ladung
wahrgenommene Länge		0.82
wahrgenommene Unangemessenheit		0.90
wahrgenommene Inakzeptabilität		0.91
Eigenwert		2.30
aufgeklärte Varianz		77%
dritter Messzeitpunkt		
Bartlett-Test	$\chi^2 (3) = 170,71, p < .001$	
KMO	.68	
		Ladung
wahrgenommene Länge		0.86
wahrgenommene Unangemessenheit		0.95
wahrgenommene Inakzeptabilität		0.96
Eigenwert		2.58
aufgeklärte Varianz		86%

ANMERKUNG: $N = 61$

TABELLE D.4: Zusammenfassung der Komponentenladungen beim Aufgaben A, B und C.

Aufgabe A		
Bartlett-Test	$\chi^2 (3) = 99.00, p < .001$	
KMO	.62	
		Ladung
wahrgenommene Länge	0.71	
wahrgenommene Unangemessenheit	0.93	
wahrgenommene Inakzeptabilität	0.93	
Eigenwert	2.23	
aufgeklärte Varianz	74%	
Aufgabe B		
Bartlett-Test	$\chi^2 (3) = 136.00, p < .001$	
KMO	.69	
		Ladung
wahrgenommene Länge	0.85	
wahrgenommene Unangemessenheit	0.94	
wahrgenommene Inakzeptabilität	0.95	
Eigenwert	2.50	
aufgeklärte Varianz	83%	
Aufgabe C		
Bartlett-Test	$\chi^2 (3) = 162.33, p < .001$	
KMO	.71	
		Ladung
wahrgenommene Länge	0.91	
wahrgenommene Unangemessenheit	0.94	
wahrgenommene Inakzeptabilität	0.97	
Eigenwert	2.64	
aufgeklärte Varianz	88%	

ANMERKUNG: $N = 61$

D.4 Korrelationen der UX-bezogenen Variablen

TABELLE D.5: Innersubjektkorrelationen zwischen den UX-bezogenen Variablen.

	wahrg. At- traktivität	Valenz	Arousal
wahrg. Usability	.63 ***	.45 ***	-.05
wahrg. Attraktivität		.34 ***	-.02
Valenz			.01

ANMERKUNG: * bei $p < .05$, ** bei $p < .01$, *** bei $p < .001$, fett für $|r| \geq .3$

D.5 Analyse der Effekte der UVn auf die UX-bezogenen Variablen

TABELLE D.6: Effekte der unabhängigen Variablen auf UX-bezogene Variablen log-transformierte wahrgenommene Usability und log-transformierte wahrgenommene Attraktivität.

	β -Gewicht	95% Konf.	Std. Fehler	t -Wert	
Effekte auf log-transformierte wahrgenommene Usability					
Intercept	1.339	[1.22, 1.46]	0.059	22.868	*
laut. D. mit/ohne	0.006	[-0.23, 0.24]	0.117	0.051	
Usab. hoch/niedrig	-1.182	[-1.42, -0.95]	0.117	-10.092	*
laut. D.:Usab.	0.001	[-0.47, 0.47]	0.234	0.005	
Aufgabe Expl./A	-0.237	[-0.38, -0.09]	0.074	-3.180	*
Aufgabe A/B	-0.059	[-0.21, 0.09]	0.074	-0.798	
Aufgabe B/C	0.051	[-0.1, 0.2]	0.074	0.689	
Random Effect Struktur: (1 <i>Versuchsperson</i>)					
$R^2_{GLMM(m)}$: 52.1 %, $R^2_{GLMM(c)}$: 75.8 %					
Effekte auf log-transformierte wahrgenommene Attraktivität					
Intercept	1.240	[1.1, 1.38]	0.069	18.076	*
laut. D. mit/ohne	0.051	[-0.22, 0.32]	0.137	0.375	
Usab. hoch/niedrig	-0.974	[-1.25, -0.7]	0.137	-7.101	*
laut. D.:Usab.	0.131	[-0.42, 0.68]	0.274	0.477	
Aufgabe Expl./A	-0.235	[-0.38, -0.09]	0.073	-3.221	*
Aufgabe A/B	-0.117	[-0.26, 0.03]	0.073	-1.598	
Aufgabe B/C	0.110	[-0.03, 0.25]	0.073	1.505	
Random Effect Struktur: (1 <i>Versuchsperson</i>)					
$R^2_{GLMM(m)}$: 38.7 %, $R^2_{GLMM(c)}$: 75.5 %					

ANMERKUNG: Die Sterne geben an, ob der Effekt auf einem Signifikanzniveau von $\alpha = .05$ und zweiseitiger Testung signifikant ist.

TABELLE D.7: Effekte der unabhängigen Variablen auf Emotionen mit den Dimensionen Valenz und Arousal.

	β -Gewicht	95% Konf.	Std. Fehler	t -Wert	
Effekte auf Valenz					
Intercept	5.048	[4.69, 5.41]	0.181	27.960	*
laut. D. mit/ohne	0.129	[-0.59, 0.85]	0.361	0.356	
Usab. hoch/niedrig	-1.012	[-1.73, -0.29]	0.361	-2.803	*
laut. D.:Usab.	1.043	[-0.4, 2.48]	0.722	1.444	
Aufgabe Expl./A	-0.475	[-0.82, -0.13]	0.175	-2.714	*
Aufgabe A/B	-0.180	[-0.58, 0.22]	0.201	-0.896	
Aufgabe B/C	0.311	[-0.03, 0.66]	0.175	1.778	
Random Effect Struktur: $(1 + \text{Aufg. A vs. B} \text{Versuchsperson})$					
$R^2_{GLMM(m)}$: 12.3 %, $R^2_{GLMM(c)}$: 71 %					
Effekte auf Arousal					
Intercept	5.623	[5.3, 5.94]	0.160	35.091	*
laut. D. mit/ohne	-0.186	[-0.82, 0.45]	0.320	-0.580	
Usab. hoch/niedrig	0.588	[-0.05, 1.23]	0.320	1.834	
laut. D.:Usab.	1.372	[0.1, 2.65]	0.641	2.140	*
Aufgabe Expl./A	0.197	[-0.1, 0.49]	0.151	1.305	
Aufgabe A/B	-0.033	[-0.3, 0.23]	0.136	-0.242	
Aufgabe B/C	0.098	[-0.17, 0.37]	0.136	0.726	
Random Effect Struktur: $(1 + \text{Expl. vs. Aufg. A} \text{Versuchsperson})$					
$R^2_{GLMM(m)}$: 9.4 %, $R^2_{GLMM(c)}$: 75 %					

ANMERKUNG: Die Sterne geben an, ob der Effekt auf einem Signifikanzniveau von $\alpha = .05$ und zweiseitiger Testung signifikant ist.

Anhang E

Anhang Studie II

E.1 Versuchsmaterialien

E.1.1 Darstellung der Ausbalancierung

TABELLE E.1: Ausbalancierung Studie II.

Usability	lautes Denken	A B C D	B C D A	C D A B	D A B C	Σ
hoch	mit	N = 5	N = 5	N = 5	N = 5	20
hoch	ohne	N = 5	N = 5	N = 5	N = 5	20
niedrig	mit	N = 5	N = 5	N = 5	N = 5	20
niedrig	ohne	N = 5	N = 5	N = 6	N = 5	20
Σ		20	20	20	20	80

E.2 Bescheinigung der Ethikkommission



TU Berlin | Sekr. MAR 3-2 | Marchstraße 23 | 10587 Berlin

Antragsteller: Anna Katharina Trapp

Eingangsdatum des Antrages: 09.12.15

Antragsnummer: TRA_02_20151209

Datum der Beschlussfassung: 06.01.16

Fakultät V
Verkehrs- und Maschinensysteme
Institut für Psychologie und
Arbeitswissenschaft

Kognitionspsychologie und
Kognitive Ergonomie

Dr. Stefan Brandenburg
Vorsitzender Ethikkommission

Sekr. MAR 3-2
Raum MAR 3.080
Marchstraße 23
10587 Berlin

Telefon +49 (0)30 314-24838
Telefax +49 (0)30 314-25289
stefan.brandenburg@tu-berlin.de

Berlin, 06. Januar 2016

Stellungnahme der Ethikkommission zu einem Forschungsantrag

Die Ethikkommission des Instituts für Psychologie und Arbeitswissenschaft (IPA) der TU - Berlin, im Folgenden Kommission genannt, hat Ihr Forschungsvorhaben begutachtet.

Unter Einhaltung der von Ihnen vorgegebenen Richtlinien werden die wichtigsten Vorkehrungen zur Minimierung des Probandenrisikos getroffen. Es besteht ein angemessenes Verhältnis zwischen dem Nutzen und dem Risiko des Untersuchungsvorhabens.

Die Freiwilligkeit der Versuchsteilnahme wird im geprüften Forschungsvorhaben sichergestellt. Weiterhin werden die Bestimmungen zum Datenschutz eingehalten.

Als Ergebnis der Begutachtung durch die Kommission, wird die Untersuchung zum „Einfluss der Usability eines Cloudservice auf die Zeitwahrnehmung, die Vertrauenswürdigkeit des Service und die Stimmung der Nutzer“ als ethisch unbedenklich eingestuft.

Mit freundlichen Grüßen,

Stefan Brandenburg
- Vorsitzender -

> Seite 1/1 |

www.ipa.tu-berlin.de



ABBILDUNG E.1: Bescheinigung der Ethikkommission für StudieII.

E.3 Poweranalysen

TABELLE E.2: Poweranalysen Studie II.

Hyp.	AV	UV	Power	95% Konf.
1	Aufg.dauer	laut. Denken	100%	[100%, 100%]
2	Aufg.dauer	Usability	100%	[100%, 100%]
–	Aufg.dauer	laut. D.:Usab.	95%	[94%, 97%]
–	gesch. Aufg.dauer	laut. Denken	16%	[13%, 18%]
10	gesch. Aufg.dauer	Usability	99%	[98%, 100%]
–	gesch. Aufg.dauer	laut. D.:Usab.	80%	[77%, 82%]
11	PTR	laut. Denken	87%	[85%, 89%]
–	PTR	Usability	12%	[10%, 14%]
–	PTR	laut. D.:Usab.	9%	[7%, 11%]
–	Dauerbewertung	laut. Denken	6%	[5%, 8%]
12(i)	Dauerbewertung	Usability	100%	[100%, 100%]
–	Dauerbewertung	laut. D.:Usab.	67%	[64%, 70%]
–	Zeitfluss	laut. Denken	20%	[17%, 22%]
12(ii)	Zeitfluss	Usability	92%	[91%, 94%]
–	Zeitfluss	laut. D.:Usab.	36%	[33%, 39%]
–	wahrg. Usability	laut. Denken	10%	[8%, 12%]
9(i)	wahrg. Usability	Usability	100%	[100%, 100%]
–	wahrg. Usability	laut. D.:Usab.	19%	[16%, 21%]
–	wahrg. Attrakt.	laut. Denken	6%	[4%, 7%]
9(ii)	wahrg. Attrakt.	Usability	100%	[100%, 100%]
–	wahrg. Attrakt.	laut. D.:Usab.	27%	[24%, 30%]
–	Valenz	laut. Denken	24%	[21%, 26%]
9(iii)	Valenz	Usability	98%	[97%, 99%]
–	Valenz	laut. D.:Usab.	8%	[7%, 10%]
–	Arousal	laut. Denken	6%	[4%, 7%]
9(iv)	Arousal	Usability	26%	[24%, 29%]
–	Arousal	laut. D.:Usab.	42%	[39%, 45%]

ANMERKUNG: Schätzung der Power und Konfidenzintervalle auf Basis von $n = 1000$ Simulationen.

E.4 Hauptkomponentenanalyse der Items zur Dauerbewertung für die Messzeitpunkte 1 bis 4 sowie Aufgaben A bis D

TABELLE E.3: Zusammenfassung der Komponentenladungen beim ersten bis dritten Messzeitpunkt.

erster Messzeitpunkt		
		Ladung
wahrgenommene Länge		0.82
wahrgenommene Unangemessenheit		0.95
wahrgenommene Inakzeptabilität		0.93
Eigenwert		2.43
aufgeklärte Varianz		81%
zweiter Messzeitpunkt		
Bartlett-Test	$\chi^2 (3) = 260.39, p < .001$	
KMO	.71	
		Ladung
wahrgenommene Länge		0.89
wahrgenommene Unangemessenheit		0.96
wahrgenommene Inakzeptabilität		0.97
Eigenwert		2.68
aufgeklärte Varianz		89%
dritter Messzeitpunkt		
Bartlett-Test	$\chi^2 (3) = 181.62, p < .001$	
KMO	.72	
		Ladung
wahrgenommene Länge		0.89
wahrgenommene Unangemessenheit		0.94
wahrgenommene Inakzeptabilität		0.95
Eigenwert		2.57
aufgeklärte Varianz		86%

ANMERKUNG: $N = 80$

E.4. HAUPTKOMPONENTENANALYSEN ZUR DAUERBEWERTUNG

TABELLE E.4: Zusammenfassung der Komponentenladungen beim vierten Messzeitpunkt sowie bei Aufgaben A und B.

vierter Messzeitpunkt		
Bartlett-Test	$\chi^2 (3) = 157.43, p < .001$	
KMO	.69	
		Ladung
wahrgenommene Länge	0.82	
wahrgenommene Unangemessenheit	0.94	
wahrgenommene Inakzeptabilität	0.94	
Eigenwert	2.45	
aufgeklärte Varianz	82%	

Aufgabe A		
Bartlett-Test	$\chi^2 (3) = 186.05, p < .001$	
KMO	.69	
		Ladung
wahrgenommene Länge	0.83	
wahrgenommene Unangemessenheit	0.95	
wahrgenommene Inakzeptabilität	0.95	
Eigenwert	2.49	
aufgeklärte Varianz	82%	

Aufgabe B		
Bartlett-Test	$\chi^2 (3) = 208.65, p < .001$	
KMO	.68	
		Ladung
wahrgenommene Länge	0.85	
wahrgenommene Unangemessenheit	0.96	
wahrgenommene Inakzeptabilität	0.95	
Eigenwert	2.45	
aufgeklärte Varianz	85%	

ANMERKUNG: $N = 80$

TABELLE E.5: Zusammenfassung der Komponentenladungen bei Aufgaben C und D.

Aufgabe C		
Bartlett-Test	$\chi^2 (3) = 171.38, p < .001$	
KMO	.71	
		Ladung
wahrgenommene Länge		0.87
wahrgenommene Unangemessenheit		0.95
wahrgenommene Inakzeptabilität		0.93
Eigenwert		2.53
aufgeklärte Varianz		84%
Aufgabe D		
Bartlett-Test	$\chi^2 (3) = 157.13, p < .001$	
KMO	.69	
		Ladung
wahrgenommene Länge		0.87
wahrgenommene Unangemessenheit		0.91
wahrgenommene Inakzeptabilität		0.95
Eigenwert		2.48
aufgeklärte Varianz		83%

ANMERKUNG: $N = 80$

E.5 Korrelationen der UX-bezogenen Variablen Studie II

TABELLE E.6: Innersubjektkorrelationen zwischen den UX Variablen in Studie II.

	wahrg. At- traktivität	Valenz	Arousal
wahrg. Usability	.70 ***	.41 ***	-.02
wahrg. Attraktivität	–	.28 ***	-.07
Valenz	–	–	.06

ANMERKUNG: * bei $p < .05$, ** bei $p < .01$, *** bei $p < .001$, fett für $|r| \geq .3$

E.6 Analyse der Effekte der UVn auf die UX-bezogenen Variablen

TABELLE E.7: Effekte der unabhängigen Variablen auf UX-bezogene Variablen wahrgenommene Usability und wahrgenommene Attraktivität.

	β -Gewicht	95% Konf.	Std. Fehler	t -Wert	
Effekte auf wahrgenommene Usability					
Intercept	6.931	[6.53, 7.33]	0.201	34.523	*
laut. D. mit/ohne	0.262	[-0.53, 1.06]	0.402	0.654	
Usab. hoch/niedrig	-3.625	[-4.42, -2.83]	0.402	-9.028	*
laut. D.:Usab.	-0.800	[-2.39, 0.79]	0.803	-0.996	
Aufgabe A/B	-1.075	[-1.63, -0.52]	0.283	-3.795	*
Aufgabe B/C	1.287	[0.73, 1.84]	0.283	4.545	*
Aufgabe C/D	-0.125	[-0.68, 0.43]	0.283	-0.441	
Random Effect Struktur: (1 <i>Versuchsperson</i>)					
$R^2_{GLMM(m)}$: 39.1 %, $R^2_{GLMM(c)}$: 65.3 %					
Effekte auf wahrgenommene Attraktivität					
Intercept	6.791	[6.34, 7.24]	0.226	30.065	*
laut. D. mit/ohne	0.056	[-0.84, 0.95]	0.452	0.125	
Usab. hoch/niedrig	-2.444	[-3.34, -1.55]	0.452	-5.410	*
laut. D.:Usab.	-1.187	[-2.98, 0.6]	0.903	-1.314	
Aufgabe A/B	-1.062	[-1.53, -0.59]	0.239	-4.454	*
Aufgabe B/C	1.150	[0.68, 1.62]	0.239	4.820	*
Aufgabe C/D	-0.300	[-0.77, 0.17]	0.239	-1.257	
Random Effect Struktur: (1 <i>Versuchsperson</i>)					
$R^2_{GLMM(m)}$: 23.7 %, $R^2_{GLMM(c)}$: 70 %					

ANMERKUNG: Die Sterne geben an, ob der Effekt auf einem Signifikanzniveau von $\alpha = .05$ und zweiseitiger Testung signifikant ist.

TABELLE E.8: Effekte der unabhängigen Variablen auf Emotionen mit den Dimensionen Valenz und Arousal.

	β -Gewicht	95% Konf.	Std. Fehler	t -Wert	
Effekte auf Valenz					
Intercept	5.734	[5.47, 6]	0.132	43.373	*
laut. D. mit/ohne	-0.331	[-0.86, 0.19]	0.264	-1.253	
Usab. hoch/niedrig	-1.081	[-1.61, -0.56]	0.264	-4.089	*
laut. D.:Usab.	-0.263	[-1.31, 0.79]	0.529	-0.496	
Aufgabe A/B	0.188	[-0.14, 0.52]	0.167	1.123	
Aufgabe B/C	0.287	[-0.04, 0.62]	0.167	1.723	
Aufgabe C/D	-0.150	[-0.48, 0.18]	0.167	-0.899	
Random Effect Struktur: $(1 Versuchsperson)$					
$R^2_{GLMM(m)}$: 13.7 %, $R^2_{GLMM(c)}$: 57 %					
Effekte auf Arousal					
Intercept	5.719	[5.48, 5.96]	0.121	47.133	*
laut. D. mit/ohne	0.025	[-0.46, 0.51]	0.243	0.103	
Usab. hoch/niedrig	-0.312	[-0.79, 0.17]	0.243	-1.288	
laut. D.:Usab.	-0.850	[-1.81, 0.11]	0.485	-1.751	
Aufgabe A/B	0.088	[-0.1, 0.28]	0.096	0.911	
Aufgabe B/C	0.038	[-0.11, 0.19]	0.076	0.490	
Aufgabe C/D	-0.113	[-0.31, 0.08]	0.099	-1.142	
Random Effect Struktur: $(1 + Aufg.A/B + Aufg.C/D Versuchsperson)$					
$R^2_{GLMM(m)}$: 4.7 %, $R^2_{GLMM(c)}$: 84.8 %					

ANMERKUNG: Die Sterne geben an, ob der Effekt auf einem Signifikanzniveau von $\alpha = .05$ und zweiseitiger Testung signifikant ist.

Anhang F

Anhang Studie III

F.1 Versuchsmaterialien

F.1.1 Darstellung der Ausbalancierung

TABELLE F.1: Häufigkeit der Aufgaben an erster, zweiter, dritter oder vierter Stelle in Studie III.

	A	B	C	D
1.	N = 8	N = 7	N = 9	N = 9
2.	N = 10	N = 10	N = 8	N = 5
3.	N = 8	N = 8	N = 8	N = 9
4.	N = 7	N = 8	N = 8	N = 10

F.1.2 Fragebogen

1. Zeitschätzung

Bitte schätzen Sie, wie viel Zeit Sie für die Aufgabe benötigt haben!

	Minuten		Sekunden
----------------------	---------	----------------------	----------

2. Zeitbeurteilung

Die Zeit, die ich für die Bearbeitung der letzten Aufgabe gebraucht habe, empfand ich als

weder noch

lang		kurz.
------	----------------------	-------

akzeptabel unakzeptabel.

Bitte bewerten Sie die Bearbeitungszeit, die Sie mit der letzten Aufgabe verbracht haben:

Die Zeit verging während der
Aufgabenbearbeitung langsam.

Die Zeit verging während der
Aufgabenbearbeitung schnell.

Die Zeit hat sich gezogen.

Die Zeit ist verflogen.

ABBILDUNG F.1: Selbstentwickelte zeitbezogeneSkalen in Studie III.

F.2 Bescheinigung der Ethikkommission



TU Berlin | Sekr. MAR 3-2 | Marchstraße 23 | 10587 Berlin

Fakultät V
Verkehrs- und Maschinensysteme
Institut für Psychologie und
Arbeitswissenschaft

Einreichende/r: Anna Steinbach

Eingangsdatum: 31.01.2017

Fast Track-Bearbeitungsnummer: FT_2017_05

Datum des Beschlusses: 08.02.2017

Ethik-Kommission

Elke Paetzold-Teske
Skr. MAR 3-2
Marchstraße 23
10587 Berlin

Telefon +49 (0)30 314-22370
Telefax +49 (0)30 314-24374
ethik@ipa.tu-berlin.de

Berlin, 08. Februar 2017

Vorsitzender

Dr. Stefan Brandenburg

Stellungnahme der Ethik-Kommission zu einer Anfrage im Fast Track-Verfahren

Skr. MAR 3-2
Raum MAR 3.080
Marchstraße 23
10587 Berlin

Für die von Frau Anna Steinbach gestellte Anfrage für die Studie mit dem Titel „Zusammenhang zwischen dem bewerteten Zeitempfinden und Usability von Webseiten“ wird eine positive Evaluation im informellen Fast Track-Verfahren durch die Ethik-Kommission ausgesprochen. Unter einem Fast Track-Verfahren ist zu verstehen, dass ein von der Ethik-Kommission vorgelegter Fragenkatalog (www.tu-berlin.de, Direktzugang: 172129) in Selbsteinschätzung beantwortet wurde und dass alle Antworten durch die Kommission als unbedenklich eingestuft wurden. Es hat keine Begutachtung eines umfassenden Antrags stattgefunden.

Telefon +49 (0)30 314-24838
Telefax +49 (0)30 314-25289
stefan.brandenburg@tu-berlin.de

Aus der Sicht der Ethik-Kommission bestehen auf Grundlage dieser Antworten in Bezug auf Fragen zum Datenschutz, zur Anonymisierung der erhobenen Daten und zu grundlegenden Aspekten des ethischen Handelns in der Wissenschaft keine Bedenken.

Dieses Votum setzt die wahrheitsgemäße Beantwortung der Fragen im Antwort-Wahl-Verfahren sowie die gewissenhafte Berücksichtigung sämtlicher Angaben bei der Umsetzung des Forschungsvorhabens durch die Antragstellerin bzw. den Antragsteller voraus.

Mit freundlichen Grüßen

Janna Protzak

> Seite 1/1

www.ipa.tu-berlin.de



ABBILDUNG F.2: Bescheinigung der Ethikkommission für Studie III.

F.3 Poweranalysen

TABELLE F.2: Poweranalysen Studie III.

Hyp.	AV	UV	Power	95% Konf.
MC(i)	wahrg. Usability	Aufg. A/B	88%	[86%, 90%]
MC(ii)	wahrg. Usability	Aufg. B/C	100%	[100%, 100%]
MC(iii)	wahrg. Usability	Aufg. C/D	67%	[64%, 70%]
–	Aufg.dauer	Aufg. A/B	100%	[100%, 100%]
–	Aufg.dauer	Aufg. B/C	100%	[100%, 100%]
–	Aufg.dauer	Aufg. C/D	100%	[99%, 100%]
–	PTR	Aufg. A/B	19%	[16%, 21%]
–	PTR	Aufg. B/C	42%	[39%, 45%]
–	PTR	Aufg. C/D	55%	[52%, 58%]

ANMERKUNG: Schätzung der Power und Konfidenzintervalle auf Basis von $n = 1000$ Simulationen.

F.4 Korrelationen der UX-bezogenen Variablen Studie III

TABELLE F.3: Innersubjektkorrelationen zwischen den UX-bezogenen Variablen.

	wahrg. At- traktivität	Valenz	Arousal
wahrg. Usability	.75 ***	.42 ***	-.18, n.s.
wahrg. Attraktivität		.23 *	-.23 *
Valenz			.18, n.s.

ANMERKUNG: * bei $p < .05$, ** bei $p < .01$, *** bei $p < .001$, fett für $|r| \geq .3$

Anhang G

Anhang Studie IV

G.1 Versuchsmaterialien

G.1.1 Darstellung der Ausbalancierung

TABELLE G.1: Ausbalancierung Studie IV.

Usability	Dauer	
	3 min	6 min
hoch	N = 30	
niedrig	N = 30	
Σ	N = 60	

G.1.2 Fragebogen

Zeitschätzung

Bitte geben Sie eine Zeitschätzung ab, wie viel Zeit Sie Ihrer Meinung nach mit der Aufgabe verbracht haben.

Minuten und Sekunden

Wie war Ihr Zeitempfinden während der Bearbeitungszeit?

Die Bearbeitungszeit kam mir kurz vor.		Die Bearbeitungszeit kam mir lang vor.
Die Bearbeitungszeit war akzeptabel.		Die Bearbeitungszeit war unakzeptabel.
Die Zeit ist verfliegen.		Die Zeit hat sich gezogen.
Die Zeit schien während der Aufgabenbearbeitung schnell zu vergehen.		Die Zeit schien während der Aufgabenbearbeitung langsam zu vergehen.

Bitte geben Sie nun an, wie Sie das System nach der Bearbeitung der Aufgabe bewerten würdest.

nicht gebrauchstauglich		gebrauchstauglich
unattraktiv		attraktiv
schön		hässlich
einfach zu bedienen		schwierig zu bedienen

ABBILDUNG G.1: Selbstentwickelte Skalen in Studie IV.

G.2 Bescheinigung der Ethikkommission



TU Berlin | Sekr. MAR 3-2 | Marchstraße 23 | 10587 Berlin

Anna Katharina Trapp
Technische Universität Berlin
Institut für Psychologie und Arbeitswissenschaft
Fachgebiet Kognitionspsychologie und Kognitive Ergonomie
Marchstraße 23, Sekr. MAR 3-2
10587 Berlin

Eingangsdatum des Antrages: 28.08.2017
Antragsnummer: TRA_05_20170828

Datum der Beschlussfassung: 04.09.2017

Berlin, den 04.09.2017

Stellungnahme der Ethik-Kommission zu einem Forschungsantrag

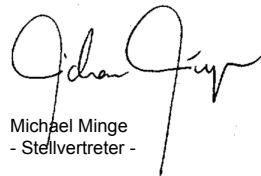
Die Ethik-Kommission des Instituts für Psychologie und Arbeitswissenschaft (IPA) der TU Berlin, im Folgenden Kommission genannt, hat Ihr Forschungsvorhaben begutachtet.

Unter Einhaltung der von Ihnen vorgegebenen Richtlinien werden die wichtigsten Vorkehrungen zur Minimierung des Probandenrisikos getroffen. Es besteht ein angemessenes Verhältnis zwischen dem Nutzen und dem Risiko des Untersuchungsvorhabens.

Die Freiwilligkeit der Versuchsteilnahme wird im geprüften Forschungsvorhaben sichergestellt. Weiterhin werden die Bestimmungen zum Datenschutz eingehalten.

Als Ergebnis der Begutachtung durch die Kommission, wird die Untersuchung „Manipulierte Passage of Time Judgments und UX“ als **ethisch unbedenklich** eingestuft.

Mit freundlichen Grüßen,



Michael Minge
- Stellvertreter -

> Seite 1/1 |

www.ipa.tu-berlin.de

Fakultät V
Verkehrs- und Maschinensysteme
Institut für Psychologie und
Arbeitswissenschaft

Ethik-Kommission

Elke Paetzold-Teske
Skr. MAR 3-2
Marchstraße 23
10587 Berlin

Telefon +49 (0)30 314-22370
Telefax +49 (0)30 314-24374
ethik@ipa.tu-berlin.de

Vorsitzender

Dr. Stefan Brandenburg

Skr. MAR 3-2
Raum MAR 3.080
Marchstraße 23
10587 Berlin

Telefon +49 (0)30 314-24838
Telefax +49 (0)30 314-25289
stefan.brandenburg@tu-berlin.de



ABBILDUNG G.2: Bescheinigung der Ethikkommission für Studie IV.

G.3 Poweranalysen

TABELLE G.2: Poweranalysen Studie IV.

Hyp.	AV	UV	Power	95% Konf.
14	Dauerschaetzung	Dauer	100%	[100%, 100%]
–	Dauerschaetzung	Usability	29%	[27%, 32%]
–	Dauerschaetzung	Dauer:Usab.	82%	[80%, 85%]
–	wahrg. Usability	Dauer	6%	[5%, 8%]
9(i)	wahrg. Usability	Usability	79%	[76%, 81%]
–	wahrg. Usability	Dauer:Usab.	4%	[3%, 5%]
–	wahrg. Attrakt.	Dauer	48%	[44%, 51%]
9(ii)	wahrg. Attrakt.	Usability	40%	[37%, 43%]
–	wahrg. Attrakt.	Dauer:Usab.	13%	[11%, 15%]
–	Valenz	Dauer	63%	[60%, 66%]
9(iii)	Valenz	Usability	89%	[87%, 91%]
–	Valenz	Dauer:Usab.	7%	[5%, 9%]
–	Dauerbewertung	Dauer	100%	[100%, 100%]
–	Dauerbewertung	Usability	15%	[13%, 17%]
–	Dauerbewertung	Dauer:Usab.	7%	[5%, 9%]
–	Dauerbewertung	Dauerschaetzung	62%	[58%, 65%]
–	Dauerbewertung	Valenz	96%	[94%, 97%]
–	Zeitfluss	Dauer	13%	[11%, 16%]
–	Zeitfluss	Usability	10%	[8%, 12%]
–	Zeitfluss	Dauer:Usab.	43%	[40%, 46%]

ANMERKUNG: Schätzung der Power und Konfidenzintervalle auf Basis von $n = 1000$ Simulationen.

G.4 Korrelationen der UX-bezogenen Variablen

Studie IV

TABELLE G.3: Korrelationen zwischen den UX-bezogenen Variablen.

Korrelationen innerhalb der Subjekte			
	wahrg. At- traktivität	Valenz	Arousal
wahrg. Usability	.59 ***	.31 *	.18, n.s.
wahrg. Attraktivität		.39 **	.34 **
Valenz			.39, **
Korrelationen zwischen den Subjekten			
	wahrg. At- traktivität	Valenz	Arousal
wahrg. Usability	.57 ***	.57 *	.08, n.s.
wahrg. Attraktivität		.55 **	.15, n.s.
Valenz			.28, *

ANMERKUNG: * bei $p < .05$, ** bei $p < .01$, *** bei $p < .001$, fett für $|r| \geq .3$