Smart@load? Modeling interruption while using a Smartphone-app in alternating workload conditions

Masterarbeit

im Studiengang Human Factors (M.Sc.)

Technische Universität Berlin

Fakultät V Verkehrs- und Maschinensysteme

Institut für Psychologie und Arbeitswissenschaft

Fachgebiet Kognitive Modellierung in dynamischen Mensch-Maschine-Systemen

eingereicht von: Maria Wirzberger

Matrikelnummer: 346577

Erstgutachterin: Prof. Dr.-Ing. Nele Rußwinkel

Zweitgutachter: Prof. Dr.-Ing. Sebastian Möller

eingereicht beim Prüfungsamt am: 01.12.2014

Eidesstattliche Versicherung

Hiermit erkläre ich, Maria Wirzberger, dass ich die vorliegende Masterarbeit selbstständig und eigenhändig sowie ohne unerlaubte fremde Hilfe und ausschließlich unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe. Ich versichere ebenfalls, dass ich bisher keine entsprechende Arbeit mit gleichem oder ähnlichem Thema an der Technischen Universität Berlin oder einer anderen Hochschule eingereicht habe.

Berlin, den 01.12.2014 Unterschrift: ________________________________

Abstract

Based on a time course model of interruption and resumption, the current thesis aims to inspect cognitive processes after being interrupted by product advertisements while performing a shopping task with a smartphone application. In doing so, different levels of mental workload, which are assumed to influence human performance as well as resumption strategy choice in this context, are taken into account. Within the applied research approach, cognitive modeling in the framework of the cognitive architecture ACT-R is combined with the development of a corresponding experimental design. The derived model predictions are validated with a 2x3-factorial design that includes repeated measures upon the second factor, and consists of 62 human participants. In detail, the influence of mental workload (high vs. low) and interruption (no vs. low vs. high) on various aspects of task-related performance and the applied resumption strategy is assessed. While the inspected performance parameters and resumption strategy choice usually point towards the expected direction for the model data, a converse pattern for the human data shows up in most cases. Comparing model and human data for each level of workload displays rather mixed results that are discussed afterwards. An outline of potential expansions and toeholds for future research within and beyond the mobile sector forms the completion of the thesis.

Zusammenfassung

Auf Basis eines Modells zum zeitlichen Verlauf der Unterbrechung und Wiederaufnahme einer Aufgabe, untersucht die vorliegende Arbeit kognitive Prozesse nach der Unterbrechung durch Produktwerbung im Rahmen einer Einkaufsaufgabe am Smartphone. Dabei werden verschiedene Ausprägungen mentaler Beanspruchung berücksichtigt aufgrund der Annahme, dass diese die aufgabenbezogene Leistung sowie die Wahl der jeweiligen Strategie zur Wiederaufnahme der Aufgabe beeinflussen. Im Rahmen des verwendeten Forschungsansatzes wird die Erstellung eines kognitiven Modells innerhalb der kognitiven Architektur ACT-R mit der Entwicklung eines korrespondierenden experimentellen Designs kombiniert. Die abgeleiteten Modellvorhersagen werden mit einem 2x3-faktoriellen Design mit Messwiederholung auf dem zweiten Faktor verglichen, welches 62 Versuchsteilnehmer umfasst. In diesem Zusammenhang wird der Einfluss von mentaler Beanspruchung (hoch vs. niedrig) und Unterbrechung (keine vs. wenig vs. viel) auf verschiedene Aspekte der aufgabenbezogenen Leistung und die jeweils genutzten Wiederaufnahmestrategien erhoben. Während die untersuchten Leistungsparameter und die Wahl der Wiederaufnahmestrategie in den Modelldaten überwiegend in die erwartete Richtung weisen, zeigt sich in den Experimentaldaten in vielen Fällen ein gegenläufiges Muster. Ein direkter Vergleich der Modell- und Experimentaldaten für jede der beiden Ausprägungen mentaler Beanspruchung bringt eher gemischte Resultate, welche im Anschluss daran diskutiert werden. Den Abschluss der Arbeit bildet eine Betrachtung möglicher Erweiterungen und Ansatzpunkte für zukünftige Forschungsarbeiten innerhalb und außerhalb des mobilen Sektors.

Acknowledgements

A project like that conducted within the current thesis cannot be realized without a bunch of helping hands. On this account, I am very grateful for all the kind people supporting me with words and deeds during the last few months.

In particular, I give my thanks to Marc Halbrügge for his advice in statistical data analysis and model-related issues, and to my advisor Prof. Dr.-Ing. Nele Rußwinkel for acquainting me with ACT-R by means of coursework and discussions, and in this vein bringing forward my ideas. Moreover, I really appreciate the support of Nikolaus Rötting regarding the technical as well as organizational implementation of the experimental part, and the dedicated aid of Lisa Dörr in app programming and conducting experiments. Beyond that, I thank Fabian Joeres and the team of this year's ACT-R Spring School and Master Class in Groningen for improving my knowledge of LISP-related issues, and all the colleagues across the Department of Psychology and Ergonomics, questioned by me on various subjects, for their kind explanations. Last but not least, many thanks to my proofreaders Marika Nürnberg, Elizabeth Morris and Daniel Wirzberger for their invaluable feedback and the patient correction of my mistakes in grammar and spelling.

Above all, I am deeply grateful to my family, for extensively supporting me on financial, organizational, emotional, motivational and technical accounts, not only during this thesis, but also the study as a whole, previous studies and my entire life. Mom, Dad and Daniel – this thesis is dedicated to you, thank you for being part of my life!

Table of Contents

Eidesstattliche Versicherung **1**

Abstract **2**

Zusammenfassung **3**

Acknowledgements **4**

List of tables **9**

List of figures **9**

1 Introduction **10**

2 Theoretical background **11**

2.1 Interruption 11

2.1.1 Definition and constituting aspects 11

2.1.2 Time course model of interruption and resumption 12

2.1.3 Resumption strategies 13

2.2 Mental workload 14

2.2.1 Definition and measurement 14

2.2.2 Conjunction to working memory 15

2.3 Cognitive Modeling 16

2.3.1 ACT-R core features 17

2.3.2 Memory-related issues 19

2.4 Thesis characteristics 19

2.5 Hypotheses 21

3 Methods **22**

3.1 Task 22

3.1.1 Shopping list application 22

3.1.2 Shopping task 23

3.1.3 Interruption 23

3.1.4 Workload variation 24

3.1.5 Performance parameters 25

3.2 Model 26
3.2.1 ACT-R experimental GUI 26
3.2.2 Chunk-types and chunks 28
3.2.3 Task processing 29
3.2.3.1 Read and remember 29
3.2.3.2 Navigate and select 30
3.2.3.3 Interruption and resumption 31
3.2.3.4 Final recall 33
3.2.4 Adjusted subsymbolic parameters 34
3.2.5 Model assumptions 34
3.3 Experiment 35
3.3.1 Participants 35
3.3.2 Design 35
3.3.3 Material 36
3.3.3.1 Shopping task 36
3.3.3.2 Structured interview 37
3.3.3.3 Counting Span task (CSPAN) 38
3.3.3.4 Questionnaire on affinity for technology (TA-EG) 39
3.3.3.5 NASA Task Load Index (NASA-TLX) 40
3.3.4 Procedure 40
3.3.5 Scoring 40
4 Results 42
4.1 Model predictions 42
4.1.1 Hypothesis 1: Main effect of interruption 43
4.1.2 Hypothesis 2: Main effect of mental workload 43
4.1.3 Hypothesis 3: Interaction between interruption and mental workload 44
4.1.4 Hypothesis 4: Difference in resumption strategies 45
4.2 Experimental results 46

4.2.1 Hypothesis 1: Main effect of interruption 47

4.2.2 Hypothesis 2: Main effect of mental workload 48

4.2.3 Hypothesis 3: Interaction between interruption and mental workload 49

4.2.4 Hypothesis 4: Difference in resumption strategies 50

4.3 Evaluation of the model fit 52

4.3.1 Applied goodness-of-fit indices 52

4.3.2 High workload variation 52

4.3.3 Low workload variation 54

5 Discussion 56

5.1 Interpretation 56

5.1.1 Interruption 57

5.1.2 Mental workload 58

5.2 Implications 59

5.2.1 Theoretical implications 59

5.2.2 Practical implications 60

5.3 Limitations 61

5.3.1 Model complexity 61

5.3.2 Sample size 61

5.3.3 Experimental setting 62

5.4 Prospect 62

5.4.1 Extending the model 62

5.4.2 Extending the focus 63

6 Conclusion 64

References 65

Appendix 70

A Experiment material 70

A1.1 Demographic questionnaire 71

A1.2 Instructions for the shopping task 73

A1.2 Questions within the structured interview 76

B Digital appendix.. 77

C Proof of subject hours 78

List of tables

Table 1. Descriptive statistical values of the performance parameters product selection duration, number of selected products, interruption time, resumption time, and finally recalled products in model runs with different intensity of interruption, divided by high and low workload and overall …………………… 42

Table 2. Descriptive statistical values of the performance parameters product selection duration, number of selected products, interruption time, resumption time, and finally recalled products in human runs with different intensity of interruption, divided by high and low workload and overall …………………… 46

List of figures

Figure 1. Time course of interruption and resumption during a main task. ……………………………………... 12

Figure 2. Relationship between primary task demand, resources supplied and performance. …………………... 15

Figure 3. Overview of modules contained in ACT-R 6.0 with duty and corresponding brain region. ………….. 17

Figure 4. Main menu, store menu and product menu for drugstore of the shopping list application. …………… 22

Figure 5. Example of the product advertisement “body lotion” appearing within the drugstore. ………………... 24

Figure 6. Implementation of main menu, shop menu and an example product menu in the ACT-R GUI. ……... 27

Figure 7. Example for a product advertisement implementation in the ACT-R GUI. …………………………... 27

Figure 8. Reading and remembering process implementation in ACT-R. ………………………………………. 30

Figure 9. Navigation and selection process implementation in ACT-R. ………………………………………… 31

Figure 10. Interruption and resumption process implementation in ACT-R. ……………………………………. 32

Figure 11. Final recall process implementation in ACT-R. ……………………………………………………... 33

Figure 12. Example display of the CSPAN task. ………………………………………………………………... 38

Figure 13. Model data on final recall in high and low workload variation. ……………………………………... 43

Figure 14. Model data on task performance under high and low workload. …………………………………….. 44

Figure 15. Model data on resumption strategies for high and low workload. …………………………………… 45

Figure 16. Scores on the NASA-TLX subscales and the overall sum score. ……………………………………. 48

Figure 17. Human data on final recall under high and low workload. …………………………………………... 49

Figure 18. Human data on task performance parameters for high and low workload. …………………………... 50

Figure 19. Human data on resumption strategies in high and low workload condition. ………………………… 51

Figure 20. Comparison between model and human data on task performance under high workload. …………... 53

Figure 21. Comparison between model and human data on final recall for high and low workload. …………… 54

Figure 22. Comparison between model and human data on task performance under low workload. …………... 55

1 Introduction

According to statistical information, currently more than 40 million people in Germany use a smartphone (Statista, 2014a). A core feature of this kind of device depicts the fact that its possibilities go far beyond making phone calls. In fact, users are offered a variety of functions, mainly organized in more or less considerable, self-contained applications – briefly called "apps" – serving specific purposes. They can be expanded in any order and are selected by just one touch. Despite all convenience, using a smartphone entails some trouble. Besides others, interruption depicts a frequently appearing phenomenon in interaction with mobile technical systems. Potential distractors while interacting with a smartphone application can be induced by the system itself (e.g., advertisement, updates, system crash) or caused due to the mobile context (e.g., motion, road traffic). The key challenge after facing such an interruption consists in successfully resuming the main task.

Especially advertisements constitute an omnipresent type of interruption in this setting, obvious by means of predicted sales amounting to € 107 million in the sector of mobile display advertisement in 2014 (Statista, 2014b). This implies an increase in sales of about 65% compared to the previous year. Advertising messages often appear triggered by a certain kind of previous user behavior, and in some cases are even unavoidable, e.g., appearing as pop-up windows with a closing button which is rather unobvious or emerges delayed, thus forcing the user to inspect the advertisement more closely. In general, the stronger an advertisement is related to the given context, the more likely it is to receive the attention of the potential customer (Yi, 1990). Whenever a decision is demanded for or against an offered product, additional cognitive demands are placed on the user, since he or she has to put substantial effort into information processing and decision making. Moreover, since smartphones are claimed to be designed particularly for mobile settings, their use is embedded into various situational contexts. On this account, due to today's busy lifestyle, demands on users might already be enhanced in some cases, putting additional constraints on the available cognitive capacity. Hence, they experience an increased level of mental workload, since information processing is employed to a broader extent. By this means, interruption would be perceived as more critical, providing the urgent necessity for designing interfaces able to support the user in such cases via fostering successful resumption.

The current thesis aims to inspect cognitive processes after being confronted with an interruption by advertisement while using a smartphone app, taking into account various levels of mental workload. In particular, the examined research question queries how task-related strategies and processes change due to inducing interruption and manipulating mental workload in a mobile task setting, and by this means influence the resulting task performance. Serving this purpose, an important part of the chosen cognitive scientific approach depicts establishing a user model within a cognitive architecture, for its strength in analyzing basic cognitive mechanisms. In the field of human-machine interaction this method becomes increasingly popular, since apart from a solid theoretical background it offers a computational implementation as well for testing the model. Nevertheless, to assess the adequacy of such a user model in terms of actual human behavior, a validation with human data is essential. For this reason, another core part of the thesis comprises the development of a corresponding experimental design with human participants performing the same task.

2 Theoretical background

Before outlining characteristics of the current thesis and in this vein the examined hypotheses as well, the core concepts of the topic should be elucidated, ensuring a solid theoretical base for the derived assumptions. On this account, first of all interruption and mental workload are discussed broadly by considering related research. As the cognitive modeling approach depicts a main methodological focus of this work, certain aspects bearing relevance within the given context are explained later on.

2.1 Interruption

When approaching the matter of interruption, the first emerging issues consist of what exactly characterizes interruptions and which aspects influence their disruptiveness. Next, relevant theories applicable for the given problem have to be inspected.

2.1.1 Definition and constituting aspects

Following Brixey et al. (2007), this kind of human experience is usually neither planned nor expected, and depicts a cognitive break with the task performed up to that time. It can be induced by internal or external sources ("self-interruption" vs. "external interruption"), resides within a certain situational context, and indicates a delay in finishing the previous activity. The main goal after facing an interruption comprises to successfully return the mental resources to the actual focus of attention, commonly denoted as resumption. Particularly dedicated to the context of human-computer interaction, McFarlane additionally considers aspects like the method of interruption coordination – immediate, negotiated, mediated or scheduled – or the modality used for their expression (McFarlane & Latorella, 2002). Apart from the described unexpected interruptions, those occurring expected or even planned exist as well, commonly referred to as multitasking (Salvucci & Taatgen, 2010). However, this aspect should not be part of the thesis, which examines a certain kind of unexpected interruption, and in this vein sticks to the definition of Brixey et al. (2007). Interruptions are known to impair the main task performance particularly due to a set of disruptive aspects. Besides others, a high complexity in terms of processing or memory demands (Gillie & Broadbent, 1989), a great similarity to the main task (Gillie & Broadbent, 1989), the appearance at inappropriate moments within the respective activity (Adamczyk & Bailey, 2004), and an immediate occurrence (Trafton, Altmann, Brock, & Mintz, 2003) are qualified to foster its significant decrease. Additionally, if the user has no opportunity to refuse or delay the interruption ("forced interruption"), its impairing effects usually increase (Salvucci & Taatgen, 2010) compared to interruptions with higher potential of control in timing ("deferrable interruptions"). Tying in with McFarlane's theory mentioned earlier, only in the latter case people have the choice to handle interruptions negotiated, mediated or scheduled, whereas forced interruptions always bear the necessity to immediately receive attention.

2.1.2 Time course model of interruption and resumption

Cognitive processes in face of an external interruption can be described by means of a time course model of interruption and resumption by Altmann, Trafton and colleagues (Trafton et al., 2003).

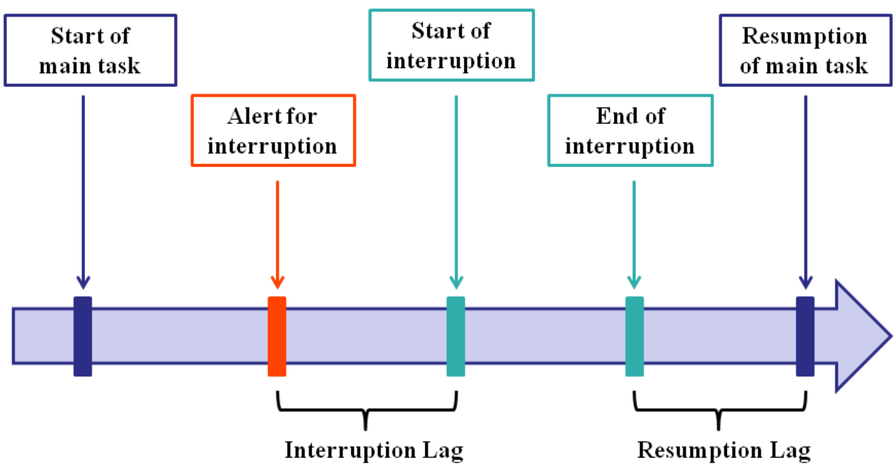

Figure 1. Time course of interruption and resumption during a main task. Adapted from Trafton et al. (2003) to the wording used within this thesis.

As shown in *Figure 1*, after starting the main task (originally referred to as “primary task”) and performing it for some time, an alert appears, e.g., a telephone ring, announcing the interruption (originally referred to as “secondary task”) before it actually occurs, e.g., answering the telephone. Although an alert like a telephone ring itself already causes a break within task execution, along with Trafton et al. (2003) in the context of this thesis it is not regarded as interruption, since there still is the possibility to deny the interruption, i.e. refuse to answer the phone. The time span between the alert for and start of the interruption is called the *interruption lag*, while the *resumption lag* specifies the interval between ending the interruption, e.g., finishing the telephone call, and successfully resuming the main task, indicated by performing the first main task-related action. Both periods play a crucial role within the discussed process: on one hand, the interruption lag is reckoned to give the opportunity to prepare a quick and effective return to the main task later on. Otherwise, the resumption lag comprises an authentic measure for the extent of disruptiveness of the interruption, with longer time spans indicating stronger disruption effects. Iqbal and Horvitz (2007) discuss a similar approach, stating a pre-interruption, preparation, diversion, and resumption phase sequently performed within an *"interruption lifecycle*" (Iqbal & Horvitz, 2007, p. 679).

In theoretical accounts, the time course depicted in *Figure 1* rests upon the memory for goals theory described by Altmann and Trafton (2002). In brief, it assumes a decay of the cognitive representation facilitating the main task, i.e. its goal, knowledge necessary to solve it, and already performed steps, in aid of the cognitive representation supporting the interruption. Nevertheless, there are two ways to reduce such a decay. At first, the *rehearsal* of core aspects related to the main task can be performed ("strengthening constraint") either retrospective with focus on the last, or prospective with focus on the next task-related step. Amongst others, Cades, Boehm-Davis, Trafton, and Monk (2007) show the facilitating role of the ability to rehearse during an interruption for a successful resumption. Secondly, *environmental cues* can be defined and directly linked to certain aspects of the main task ("priming constraint"). As outlined by Trafton, Altmann, and Brock (2005), such cues entail strong effects, especially when they are quite obvious for the user. For both techniques, the interruption lag introduced above is of great importance, as it offers the time needed for applying them, and in this vein fosters effective resumption.

2.1.3 Resumption strategies

Derived from those issues, two main approaches that bear high relevance within this thesis can be distinguished in terms of applicable resumption strategies. While the *memory-based* strategy simply consists of trying to remember information on previous actions, the *reconstruction-based* strategy relies on environmental context for recreating the prior task setting (Salvucci & Taatgen, 2010). That distinction reminds us of the concepts of *knowledge* *in the head* and *knowledge in the world* (Norman, 1988). Referring to the first aspect, he also outlines the rehearsal as being of high relevance for memorizing things. However, although the application might be highly efficient, potential problems concerning this kind of knowledge arise from the fact that it needs to be learned adequately beforehand. Moreover, the retrieval in critical situations may fail or require costly memory search, resulting in decreased task performance. On the other hand, the author specifies the world as an opportunity to put memory load out of the person. This corresponds to the premise from embodiment research that people can *"off-load cognitive work onto the environment"* (Wilson, 2002, p. 626) to relieve the limited information processing capacity, particularly in demanding situations. A great advantage of such world-based knowledge depicts that it does not require extensive learning processes but can be used forthright. Nevertheless, it requires people to find and interpret information first, taking additional time and on this account potentially impairing task performance.

2.2 Mental workload

Approaching the construct of mental workload implies giving a definition first, and in this vein discussing potential ways of assessment. Due to the fact that working memory plays a crucial role in this context, an elucidation of related theoretical issues completes this section.

2.2.1 Definition and measurement

As discussed by Gopher and Donchin (1986), mental workload depicts a concept enfolding various dimensions and facets. Although it has been broadly inspected, deriving a clear definition forms a rather difficult matter. Nevertheless, there are two constituting aspects commonly agreed on in most cases. While *task difficulty* results from the demands required to successfully solve a task, *resource supply* points to the information processing capacity available for this purpose. In this vein,

> *"mental workload may be viewed as the difference between the capacities of the information processing system that are required for task performance to satisfy performance expectations and the capacity available at any given time"* (Gopher & Donchin, 1986, p. 41-3).

Task difficulty can be enhanced by inducing an additional task, e.g., related to motor, perceptual or memory demands. Those secondary tasks might stand on their own, or even be natural part of the actual task, referred to as embedded secondary tasks. When trying to measure mental workload in this context, a widely used approach consists of inspecting aspects of primary task performance facing such increased demands (O'Donnel & Eggemeier, 1986).

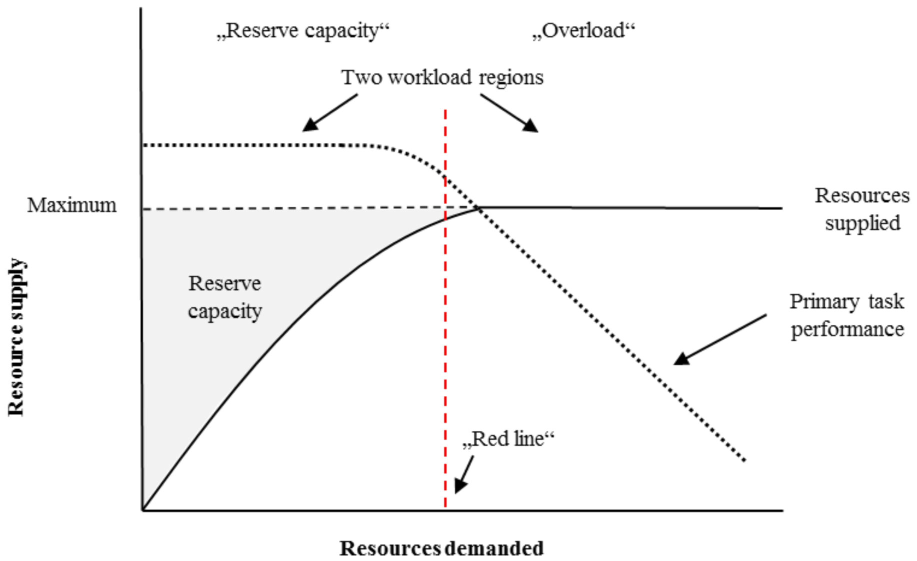

Figure 2. Relationship between primary task demand, resources supplied and performance. The "red line" marks the boarder to workload overload. Adapted from Wickens, Hollands, Banbury, & Parasuraman (2013, p. 348).

A combined focus on speed and accuracy depicts a frequently applied measure addressing different ways of inducing workload as well as diverse levels of load. Based on the assumption that tasks with increased difficulty require additional resources, a significant decrease in performance due to the lack of resources should appear as soon as resource demands cross the "red line", just as shown in *Figure 2*.

2.2.2 Conjunction to working memory

One important source of constraint in information processing exists due to working memory limitations, both in terms of duration and capacity (Wickens et al., 2013). The first aspect refers to the fact that information in working memory decays after a certain time. In order to extend such period of availability, people can rehearse relevant information. In contrast, the matter of capacity indicates that not more than a defined amount of information can be stored at the same time. According to Miller's prominent paper, it should reside between five and nine items (Miller, 1956), although more recent research proposes smaller numbers. Again, rehearsing information depicts a way to increase this span. In general, when performing a memory-related task, memory load has to be maintained by means of working memory (Anderson, Reder, & Lebiere, 1996). On this account, increasing load on working memory effects task performance, and may result in difficulties to retrieve the necessary information.

A further aspect linked to working memory capacity depicts the process of working memory updating, which is inevitable as changing working memory content should be represented correctly over a certain time. As a result of having examined the construct, Ecker, Lewandowsky, Oberauer, and Chee (2010) postulate three constituting features of working memory updating, described as *retrieval*, *transformation*, and *substitution*. While the first one consists of extracting relevant information from memory, the second can be identified by adjusting this information according to situational changes. Finally, substitution results in replacing the previous informational state by the current one, entailing an updated content representation in working memory. All described components have been applied in working memory update tasks to various extents, and according to Ecker et al. (2010) independently contribute to the respective updating performance.

2.3 Cognitive Modeling

As mentioned at the outset, besides collecting human experimental data, this thesis employs cognitive modeling to inspect the underlying research question, how the manipulation of interruption and mental workload might change the resulting task-related behavior. Such a decision was made due to several relevant characteristics of the cognitive modeling approach, corresponding well with the chosen research focus. Generally, cognitive modeling aims to understand and predict constraints, errors or interference in human behavior by inspecting the cognitive processes behind them. For this purpose, cognitive architectures as certain way to apply cognitive modeling have proven of value, providing a theoretical framework to explain basic and constant mechanisms of human cognition behind a variety of tasks (Gray, Young, & Kirschenbaum, 1997). Since they offer a computational platform for model execution as well, there is the opportunity to directly link the model to other devices, e.g., for usability evaluation (Rußwinkel & Prezenski, 2014), or even to artificial cognitive agents (Trafton, Jacobs, & Harrison, 2012). Especially the cognitive architecture ACT-R (Adaptive Control of Thought – Rational), developed by John R. Anderson and colleagues (Anderson & Lebiere, 1998), is and has been used actively within a vibrant and growing research community, to address plenty of subjects. Besides its successful application in basic cognitive psychology research, ACT-R is utilized as well in more applied domains, like the field of human-computer interaction. In this area, it provides a useful theoretical foundation, and offers the chance to analyze cognitive processes while interacting with an interface, already in very early stages of development. This applies by creating user models able to conduct predefined tasks without the need for providing physical mock-ups. In this vein, the impact of devices on user's behavior can be tested without costly building equipment, and on the long term, as soon as there are broadly validated user models, even without the need to search for adequate human participants.

2.3.1 ACT-R core features

A key characteristic of the cognitive architecture ACT-R, operating on the list-based programming language LISP, depicts the assumption of different modules occupying defined duties and interacting in certain ways to create cognitive processing (Anderson, 2007). On this account, they form the foundation of any task-related behavior. *Figure 3* gives an outline of the modules comprised within the currently existing ACT-R 6.0 version.

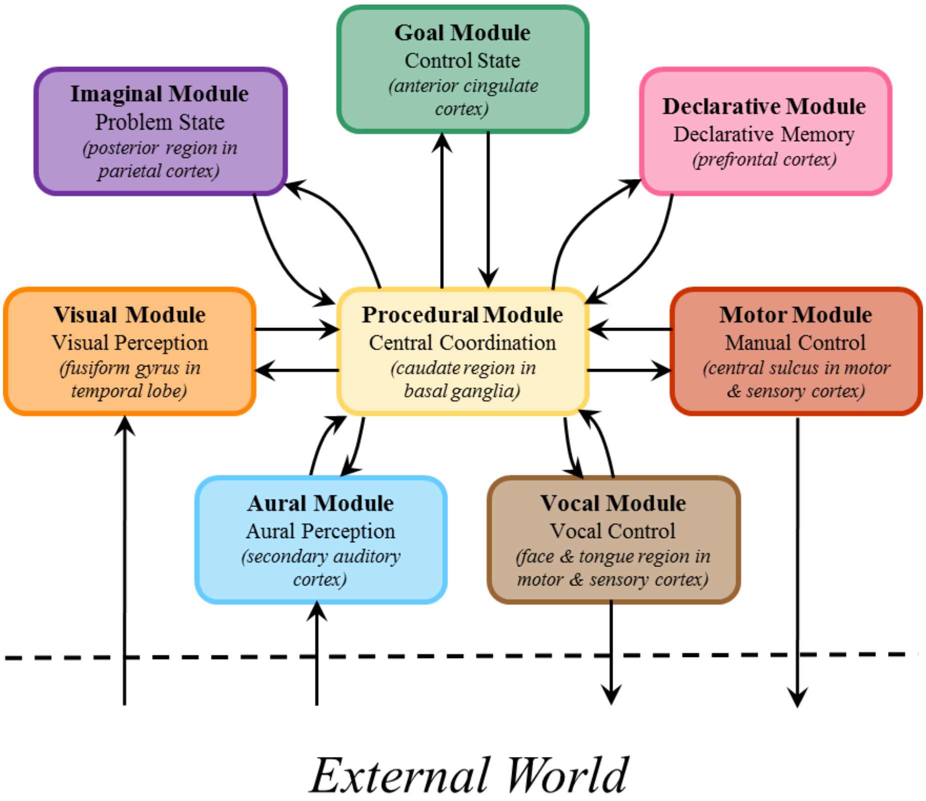

Figure 3. Overview of modules contained in ACT-R 6.0 with duty and corresponding brain region. Adapted from Borst & Anderson (in press) and Anderson (2007).

There are two modules responsible for collecting information from the respective environment, a *visual module* dealing with visual perception, and an *aural module* managing aural input. Whereas the *motor module* performs a manual response (e.g., via mouse click or key press), the *vocal module* is able to react verbally. Altogether, those modules serve as an interface to the external world, while the remaining ones are concerned with various aspects of central processing. The *goal module* maintains focusing on the actual problem and in this vein controls its solution. In contrast, the *imaginal module,* related to the current problem as well, is concerned with mentally representing transient stages of problem processing, e.g., intermediate results when performing a complex arithmetic task. Within the declarative memory all kinds of factual knowledge can be stored, hence the *declarative module* is occupied with retrieving information relevant to the respective task from memory. Finally, the *procedural module* coordinates this information provided by the other modules and selects, based on the resulting patterns, production rules that bring forth the desired behavior. Besides validation by human behavioral data, the described modules hold a vested biological background as well, since fMRI studies (Anderson, 2007; Borst & Anderson, in press) indicate the association of each module with a brain region relevant to the respective duty. Those neurophysiological areas are specified as well in *Figure 3*.

Each module holds a buffer, serving as interface to enable communication with the procedural module and by this means amongst all modules. In some cases, the buffer is simply named after the related module, but in others there are discrepancies. Thus, the buffer belonging to the declarative module is called the *retrieval buffer*, while the *manual buffer* is part of the motor module. Visual and aural system actually break up to represent the distinction between the ventral path associated with object recognition ("what system"), and the dorsal path linked to action affordance ("where system"). In contrast to the *visual/aural buffer* in the former case, in the latter case they are named *visual-/aural-location buffer*. Information processing within the outlined structure occurs via *chunks*, small units encoding relevant elements of knowledge, affiliated with a certain category (*chunk-type*) and containing specific attributes (*slots*). It incorporates a duality of parallel and serial features, since although processes in different modules can be executed in parallel, each buffer can hold just one chunk at the same time. This bottleneck intends to represent the existing limitations in information processing resources. As already mentioned, interaction between modules happens by means of *production rules*. They consist of both a condition and an action part, and depict a main duty of the procedural module. It scans the buffer's contents and, based on the resulting pattern, selects a suitable production rule that initiates the related action. However, an important constraint persists in the fact that just one production rule can be conducted at the same time, even if more than one would fit. In such cases, subsymbolic mechanisms apply, and a cost-benefit function (*utility*) decides which production rule is selected.

2.3.2 Memory-related issues

If and how fast a chunk can be retrieved from declarative memory depends on another subsymbolic mechanism called *activation*. It holds relations to working memory, for it reflects the availability of information, and is determined by the respective context and history of use:

$$A_i = B_i + \varepsilon. \tag{1}$$

As depicted in *Equation 1*, the activation value is computed by summing up the base-level activation B_i that reflects how recent and frequent a chunk has been used, and a noise value ε. The latter one is composed both of permanent noise associated with each chunk, and instantaneous nose computed in the course of any retrieval request. In the case that the activation of a requested chunk exceeds a defined threshold, its retrieval will succeed. Base-level activation itself rests upon the calculation shown in *Equation 2*:

$$B_i = ln\left(\sum_{j=1}^{n} t_j^{-d}\right). \tag{2}$$

It comprises the number of presentations n for the respective chunk, the time t_j since the jth presentation, and a decay parameter d. Each time a chunk is used, its base-level activation is increased, whereas it decays by means of a power function of time since presentation. To identify the respective base-level activation, those decay effects are accumulated and then logarithmically transformed. On this account, a possibility for increasing a chunk's activation could consist of rehearsing this information, and in this vein maintaining its presence.

2.4 Thesis characteristics

As mentioned initially, the research focus examined within this thesis is located in the applied context of smartphone use. In particular, the explored task consists of performing a shopping task by means of an application suitable to meet this demand. Although the shopping list application has been developed just on research purposes and is not used commercially, the task is claimed to hold a strong proximity to daily-life situations. The same applies to the induced interruption via product advertisements (*section 3.1.3*) and the enhanced workload by the demand to deal with additional information (*section 3.1.4*). On this account, an improved external validity of the inspected mobile setting is assumed.

Without any doubt, advertising comprises an externally induced interruption, and is hardly ignorable in most cases, thus depicting a forced interruption. In contrast to the time course model of Trafton et al. (2003), stated in *section 2.1.2*, such kind of disruption is usually characterized by the absence of an alert announcing it, implicating a missing interruption lag as well. However, without an interruption lag a user lacks the opportunity to explicitly create environmental cues or apply rehearsal before turning to the interrupting task. In consequence, naturally existing cues from memory or environment have to be used for resumption in this case.

The prominent role of information rehearsal in terms of memory has already been discussed in *section 2.1.2*. Within the current task, there is the opportunity to rehearse information while performing the product selection. However, the ability to rehearse the content of the main task while facing an interruption depends on its cognitive demands. According to Salvucci, Taatgen and Borst (2009), interrupting tasks can be classified by the difficulty of the respectively following subtask. When abruptly being confronted with an advertisement while performing a shopping task, it would consist of an information recall on the previously performed selection, depicting a medium level of difficulty. Nevertheless, since reacting towards the advertised offer requires decision making, the interruption is regarded as too demanding of cognitive resources to enable rehearsal. On this account, an extended resumption process results.

Regarding resumption strategies, those described in *section 2.1.3*, based on either knowledge in the head or knowledge in the world, are regarded as applicable as well, despite the missing interruption lag. As stated above, already existing memory or environmental contents are used instead of explicitly creating new cues. In the following, the strategy applying memory retrieval is referred to as a *head-based strategy*, whereas the strategy utilizing the appearing selection mark as environmental cue is called a *world-based strategy*. On the subject of their application, differences in terms of the actual workload demands are assumed, influencing strategy choice. Without additional demands, both resumption strategies are assumed to be chosen with equal frequency whereas increased workload determines preferring the world-based strategy. This resides upon the assumption that people try to offload as many cognitive demands as possible into their environment in case their cognitive capacity is already claimed, just as stated in *section 2.1.3*.

On methodological accounts, to shed light on the stated research question, this thesis employs cognitive modeling within a cognitive architecture as well as a related experimental design for testing the derived hypotheses. Besides validating the model performance, the human experimental setting contains additional measures relevant to further inform the chosen research focus. Nevertheless, due to the effort coming along with such an approach, there are certain limitations under conditions of such a thesis, broadly discussed in *section 5.3*.

2.5 Hypotheses

To examine the initially outlined research question, based on the discussed theoretical background for interruption (*section 2.1*) and mental workload (*section 2.2*), and the characteristics of the current thesis depicted in *section 2.4*, the following hypotheses are derived. They will serve as framework for determining the model behavior as well as inspecting the human data generated within the experimental setting.

As stated in *section 2.1.1*, interruptions impair the main task performance, especially when there is no possibility to delay or at least prepare for this cognitive break. On this account, the *first hypothesis* assumes the induction of product advertisement as forced interruption without interruption lag to significantly decrease the performance within the shopping task.

In terms of mental workload, *section 2.2.1* already outlined the negative effect of increased mental demands on the respective task performance. Such increased demands might result from an enhanced task difficulty, e.g., the necessity to deal with an additional part of the task, requiring further cognitive resources. Based on this assumption, within the *second hypothesis*, it is stated that increasing the level of mental workload by extending the scope of the task leads towards a decreased task performance as well.

Apart from the discussed impairment of the task performance due to separately inducing interruption or mental workload – just as examined in the previous hypotheses – Iqbal and Horvitz (2007) claim an increased difficulty of resource reallocation when combining both aspects. So the *third hypothesis* predicts a further decrease in task performance when interruption appears under constraints of enhanced mental workload.

In *section 2.1.3*, two strategies for resuming the main task after facing an interruption were outlined. Whereas the first one relies solely on memory content (head-based strategy), the second one deals with cues from the respective environment (world-based strategy). As already discussed, under conditions of enhanced mental workload, the environment might serve as additional cognitive resource to handle such demands. Based on this assumption, the *fourth hypothesis* postulates that users being interrupted tend to prefer the world-based resumption strategy when facing increased mental workload. In contrast, without raised cognitive demands, head-based and world-based resumption strategy should be applied to comparable extents.

3 Methods

Developing and testing an ACT-R model and validating it by human experimental data at the same time always entails substantial redundancy in describing task procedure, used application, assessed behavior and so on. On this account, consistent aspects between both parts will be outlined first, before distinct features of model respective experiment are discussed separately.

3.1 Task

The task refers to a shopping list application on an Android smartphone, already used in previous usability research (Rußwinkel & Prezenski, 2014). Compared to the originally described version, the one used within the current thesis embraces additional features explained in detail subsequently.

3.1.1 Shopping list application

Overall, the used shopping list application is composed of a simple structure of relevant menus. In detail, there is a main menu, containing “overview”, “shops”, and “my list”, a shop menu, consisting of a set of the seven shops – “bakery”, “drugstore”, “fresh & gourmet food”, “greengrocer”, “beverage shop”, “stationery shop”, and “tuck store” –, and a product menu for each shop, comprising an amount of 49 products for each shop. They are depicted in *Figure 4*. As displayed there, a back button within the upper left corner of each menu, showing a left facing arrow, a shopping cart, and the menu name, enables the transition back to the previous menu. Additionally, an overview menu, containing links to alphabetically sorted product menus, grouped in two to three letters, as well as a list menu, inclosing the preliminary selected products, are part of the application, but not used within the current task.

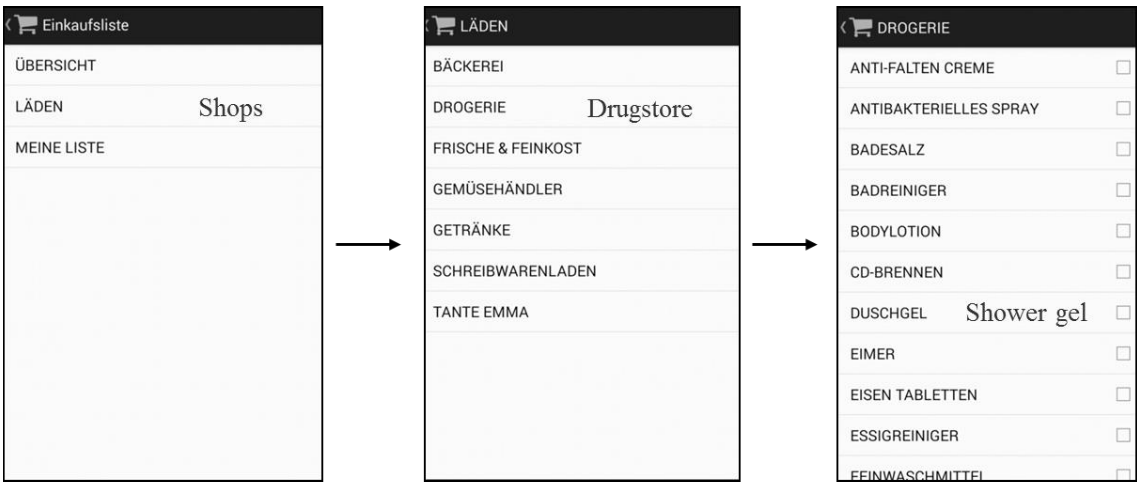

Figure 4. Main menu, store menu, and product menu for drugstore of the shopping list application.

3.1.2 Shopping task

The shopping task consists of encoding, remembering, searching for and selecting a set of 12 predefined products within the described shopping list application – shower gel, blueberries, canned pineapples, pencil, Edam cheese, farmhouse bread, iceberg lettuce, coke, apple pie, blood orange juice, sea bream, and white button mushrooms – , divided into groups of four within three runs. Products appear in a fixed sequence to minimize irrelevant sources of variance. Each run starts with the four products to be remembered listed on the screen for 30 sec. This period of time is regarded to be sufficient for an adult with average cognitive abilities to read and remember such a short set of products without difficulties. Although not explicitly announced at the outset of the task, after performing all runs, the products still remembered from the previous selections have to be recalled. As outlined, the current work involves the inspection of task-related memory processes, but nevertheless its focus is not put on the process of acquiring knowledge in using the shopping list application itself. For this reason, already existing previous knowledge about how to use the application is assumed, especially on which shop category relates to a certain product. This decision was made to reduce the complexity of the established cognitive model as well as that of the related experimental setting. Both aim to shed light on the underlying research question, particularly dedicated to cognitive processes after being interrupted while performing the task.

3.1.3 Interruption

Different to the originally developed shopping list application (Rußwinkel & Prezenski, 2014), interruptions in terms of product advertisements occur during two of the three runs, announcing a special offer. Those interruptions differ in frequency: within a run with low interruption frequency (*low ad*) an interruption is displayed after the second selected product, whereas in a run with high interruption frequency (*high ad*) an interruption appears after the first and third selected product. There are runs without an interruption as well (*no ad*). In order to avoid unrequested confounding effects, no ad, low ad and high ad runs appear once with random sequence each time the task is conducted. The occurrence of an interruption is always triggered by a certain user behavior, i.e. successfully selecting a defined amount of products within the respective run. In this vein, it affects comparable stages in human information processing each time it happens, avoiding further confounding effects (Adamczyk & Bailey, 2004). As stated in *section 2.4*, the interruption itself requires a substantial amount of cognitive effort, as it forces an encoding and afterwards decision making process to get back to the shopping task.

Each product advertisement is related to the shop the previously selected product resides in. *Figure 5* shows a product advertisement related to products within the drugstore, another product advertisement in this store offers fabric softener. There are two different product advertisements for each shop, varying randomly in appearance. Those related to the remaining shops offer bread baked in a wood-fired oven and a nut cake within the bakery, a trout and a filet of beef within the shop for fresh and gourmet food, cherries as well as romaine lettuce at the greengrocer, non-alcoholic beer and ice tea within the beverage shop, A4 size folders and yellow highlighters at the stationery shop, and pretzel sticks as well as chocolate cookies in the tuck store.

Figure 5. Example of the product advertisement “body lotion” appearing within the drugstore.

All product advertisements share a steady structure: the header “!!! SPECIAL OFFER!!!” is followed by a prominent picture of the offered product and a short description, e.g., “Today’s offer: summer body lotion with a tropically-fresh fragrance. Indulging, moisturizing care with that exotic holiday feeling!” as with the body lotion displayed above. Finally, it contains the offer to buy the product and two selection buttons for “Yes” and “No”.

3.1.4 Workload variation

As already outlined in *section 2.2.1*, mental workload is strongly related to human information processing, in particular its limitation in capacity. To increase the level of mental workload, the requirement to deal with a further aspect of the task is therefore regarded as sufficient to raise the overall task difficulty, and in this vein demand more resources to maintain an adequate task performance. Therefore, a task variation with enhanced mental workload enfolds to encode, memorize, and retrieve an additional piece of information, i.e. the respective person – Diana, Fiona or Norbert – the product has to be bought for. In detail, the list consists of shower gel for Diana, blueberries for Fiona, canned pineapples for Norbert, a pencil for Diana, Edam cheese for Fiona, farmhouse bread for Diana, iceberg lettuce for Norbert, coke for Fiona, apple pie for Norbert, a sea bream for Diana, blood orange juice for Fiona, and white button mushrooms for Diana. Dealing with this additional part of information can be regarded as a kind of embedded secondary task, previously explained in *section 2.2.1*, for it needs to be encoded as well while reading the items, remembered during the product search and holds high relevance within the final product recall, as for each product it has to be recalled for whom it was bought. Having in mind the concept of working memory updating, described in *section 2.2.2*, adding further information affects all three postulated aspects. Within the retrieval stage, besides the respective product, also the product-related person has to be retrieved, whereas in the transformation and substitution stages the target person has to be adjusted as well. On this account, all steps are assumed to require a higher amount of cognitive effort to be effectively performed.

3.1.5 Performance parameters

Task related behavior is assessed in terms of several performance parameters. At first, the mean duration needed to successfully select a product (*product selection time*) is computed as the time difference between the successful selection of a product and the transition back to the related shop menu. The latter occurs by pressing the back button after finishing a selection, and marks the starting point of the product selection. However, for the first product in each run, pressing the "SHOPS" button depicts the product selection onset. The offset of the product selection time consists in the already mentioned successful completion of the product selection process.

In order to calculate the amount of selected products (*selected products*), all correctly selected products in each run are summed up. In the case of errors, within the current work just errors of omission (Hollnagel, 1998), i.e. products missing due to the lack of ability to remember them within the product search and selection, are considered. Another type of error, errors of commission, i.e. selecting a similar product instead of the actual target product, are not included. This decision was made for reasons of complexity, as the focus of the work consists of examining the process of interruption and resumption, and not in how similar certain products are recognized.

Further parameters inspected are the times needed for interruption (*interruption time*) and resumption (*resumption time*), computed in low and high ad runs. Interruption time starts with the onset of the interruption and ends with the reaction to the offered product, i.e. pressing “YES” or “NO” to close the ad. Although this part aspect not comprise the main focus of the inspected process, it is included on explorative accounts. Resumption time consists of the difference between the offset of the interruption and the transition to the shop menu by pressing the back button. In high ad runs, interruption and resumption times are calculated as the mean of both appearances.

As mentioned in *section 3.1.2*, after completing the selection part of the task all previously handled products have to be named. This final product recall (*final recall*) serves as a measure for a longer memory span capacity. It is computed as the sum of all correctly recalled products in the low workload respectively the correctly recalled products with the related person in the high workload variation.

3.2 Model

Based on the illustrated task requirements, an ACT-R model is devised which is able to perform such a task. The development occurs with ACT-R 6 version 1.5 [r1451s] and the Clozure Common Lisp Version 1.8-r15286M (WindowsX8664). Model characteristics related to the features described in *section 3.1* are outlined within the following subsections.

3.2.1 ACT-R experimental GUI

As already mentioned in *section 2.3*, one big advantage of using ACT-R models in the field of human-technology interaction research comprises the fact that it is not mandatory to have a physical mockup, but a virtual device can be used instead. Nevertheless, due to an already existing visual implementation of parts of the shopping list application in the ACT-R experimental GUI from previous coursework, the decision was made not to create a new virtual device. Rather the existing implementation was improved and adjusted to better meet the task requirements, although the GUI is quite limited in possibilities. However, the benefit of this approach consists of the opportunity to use the visual interface for debugging as well as demonstrating the model behavior to people without detailed knowledge of the ACT-R framework. *Figure 6* depicts the implementation of the main, shop and product menu (exemplarily shown for drugstore). As obvious, creating a model always implies a reduction to the core features of an application or task. Since scrolling processes should not be part of the task, just the first 11 products of each shop are included. The back button in the upper left corner of each menu shows a rather plain design and that on the main menu page serves as a transition into the following run for getting the next four products displayed. To make it easier for the model to encode the products and navigate within the application, umlauts, blanks, and parentheses were excluded in the menus, but limited in appearance to the product advertisements.

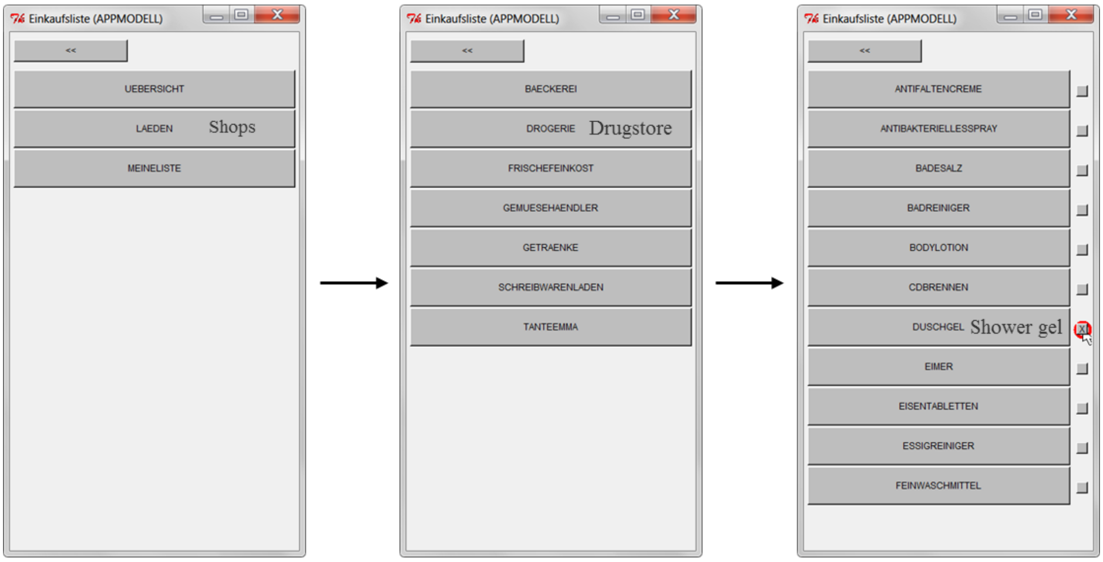

Figure 6. Implementation of main menu, shop menu and an example product menu in the ACT-R GUI.

The appearing product advertisements are distinguished by their remarkable yellow color, representing the high salience of the interruption. They always contain a comparable advertisement message – limited to just one line due to GUI constraints – and the “YES” respective “NO” button necessary to indicate the decision for or against the offered product. *Figure 7* exemplarily shows the product advertisement for body lotion.

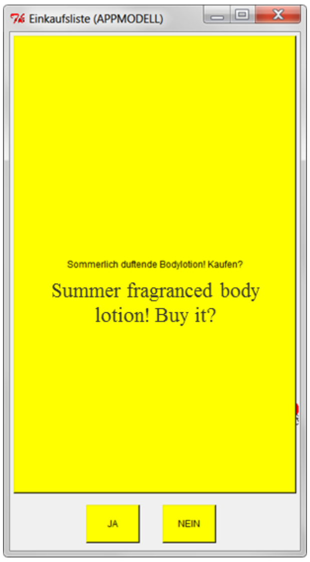

Figure 7. Example of a product advertisement implementation in the ACT-R GUI.

3.2.2 Chunk-types and chunks

As described in *section 2.2.1*, relevant information in ACT-R models is stored in chunks of different types. Within the currently specified model, a total of nine chunk-types, serving certain purposes, can be distinguished. In the *product* chunk, essential information about the relation of a product to the shop it can be bought in can be found, since it holds slots for the product name and category. An example chunk of this type would be *duschgel isa product name "DUSCHGEL" category "DROGERIE"*. In the low workload version of the model, for each product to be remembered, searched for, and selected, a *remember* chunk, containing name and current run, is created from reading the products in the beginning of a run. By contrast, in the high workload version of the model the information to be remembered is split up in two kinds of chunks. On one hand, the *remember-product* chunk contains information on the product name, the line it is located in, and the current run. On the other hand, the *remember-person* chunk comprises slots for the respective person, the line, and the current run. The decision for choosing such an implementation was made to emphasize the embedded secondary task character of dealing with this additional piece of information. Moreover, participants might forget the person-related information, but are nevertheless able to keep at least the product in mind, constituting the necessity for related but distinct chunks. The *selected* chunk serves as intermediate storage for the already selected products, preventing the model from retrieving already selected products again, and therefore includes a slot for each of the four target products. A comparable chunk-type exists for performing the final product recall after finishing the search and selection process. Such a *recalled* chunk features a slot for each of the possibly remembered 12 products, and was established for the same reason as the selected chunk. Although storing 12 products clearly extends the postulated seven plus or minus two pieces of information (Miller, 1956), it was assumed that people are able to remember even this amount of products for a rather limited time span, enabling them not to repeat already recalled products all the time.

After all, several chunk-types indicating the current task focus are part of the model. For the search and selection of a product, a *maintask* chunk exists, holding information on the product currently to be selected, the related person in the high workload version, the shop category the product can be found in, the actual number of already selected products, the current run, and a state slot announcing what step the model performs. An example of a chunk of this type in the beginning of a model run would be *maintask isa maintask product nil person nil category nil selected nil run nil state search-text*. In the case an interruption occurs, the *interruption* chunk becomes of relevance, consisting of the respective state. Eventually, the final recall exhibits its own *finalrecall* chunk as well, comprising the number of already recalled products and a state for low workload, and the person related to a product as well in the high workload variation of the model.

While the products chunks are already added to the declarative memory in advance, most of the chunks of the other types are created while the model is actually performing the task and added to the declarative memory over the course. They were chosen in the described way to be able to maintain the respective goal-state and problem-state as well as memory related content. On this account, the desired model behavior should be created, i.e. a loss in performance on the previously described performance parameters (*section 3.1.5*) when inducing interruption and enhancing workload.

3.2.3 Task processing

As mentioned in *section 3.1.2,* it is assumed that the modeled user already gained substantial experience with the application. Within the model, such kind of previous knowledge is installed by setting the base levels of all product chunks to 50 and their creation time to -100, simulating the model having used the application a lot of times and having started to do so a long time ago.

3.2.3.1 Read and remember

Processing the task always starts with searching for, finding, reading and remembering the first four products, appearing as separate lines of text on the screen – in the high workload version accompanied by the related person. The explained display duration of 30 sec for the four target products is established in a particular way within the model. As shown in *Figure 8*, it reads word by word, and at the same time creates a remember chunk for low workload respective remember-product and remember-person chunks for high workload, just as described previously. Thereby, the link between person and product is established by encoding the person in relation to the respective line in the high workload version, causing a link between both remember chunks by means of the value of the line slot. After inspecting each word, a short sleep time of four seconds is added to simulate a “read-again-and-remember” process – although it actually does not affect model time, but gives a better impression of model demonstration – before entering the navigation and selection process. This decision was made, since the information to remember has already been stored in chunks by the first inspection of each word. Moreover, developing a cognitive model always involves reducing reality, but focus on the core aspects of the task instead. In the given context, the latter comprises the process of interruption and resumption, not the reading and remembering part.

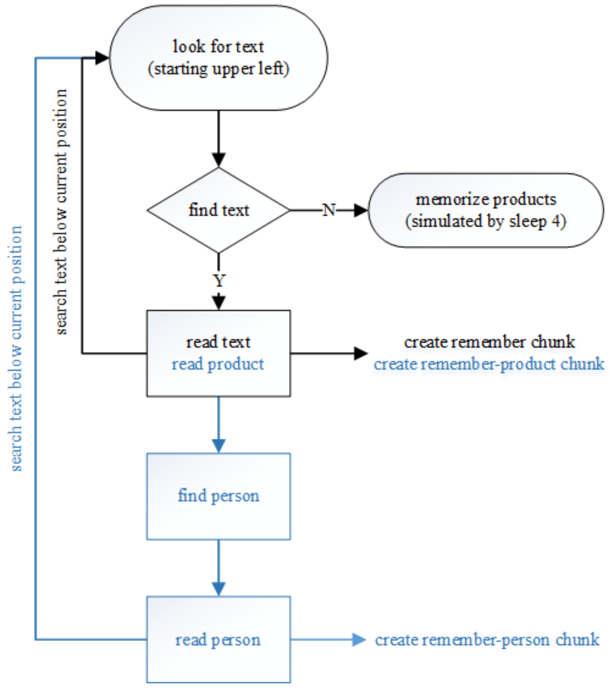

Figure 8. Reading and remembering process implementation in ACT-R. The blue color indicates distinct or additional features within the high workload variation of the model.

3.2.3.2 Navigate and select

The actual product search and selection process is depicted in *Figure 9*. It starts with retrieving a remember chunk of a previously not selected product as well as the respective chunk for the related person in the high workload condition. To be able to perform the selection process, a product chunk has to be retrieved as well, revealing the shop related to the current product. Within the main menu, navigation to and selection of the "SHOPS" button is performed randomly with or without subvocalizing the product (in both conditions) or alternatively the related person (in the high workload variation). As participants usually do not subvocalize all the time during a task, this is regarded as feasible model behavior. After successfully navigating through the shop menu, finding and selecting the correct shop – again randomly with our without performing a subvocalizing procedure as stated above – navigation within the product menu occurs. By the time the correct product is found and it is checked that the product has not been selected before, it is selected and the selection validated visually afterwards. The navigation back to the shop menu finishes the product selection and the already selected products are saved in the respective selected chunk in the imaginal buffer. A navigation back to the main menu ends the current run after either successfully remembering and selecting all four target products or lacking the ability to remember the next product, while using the back button in the main menu starts the next run. The described procedure, occupying the goal, retrieval, imaginal, visual, visual-location, aural, aural-location, vocal, and motor buffer, is repeated until all target products have been presented and attempted to be remembered, searched for, and selected.

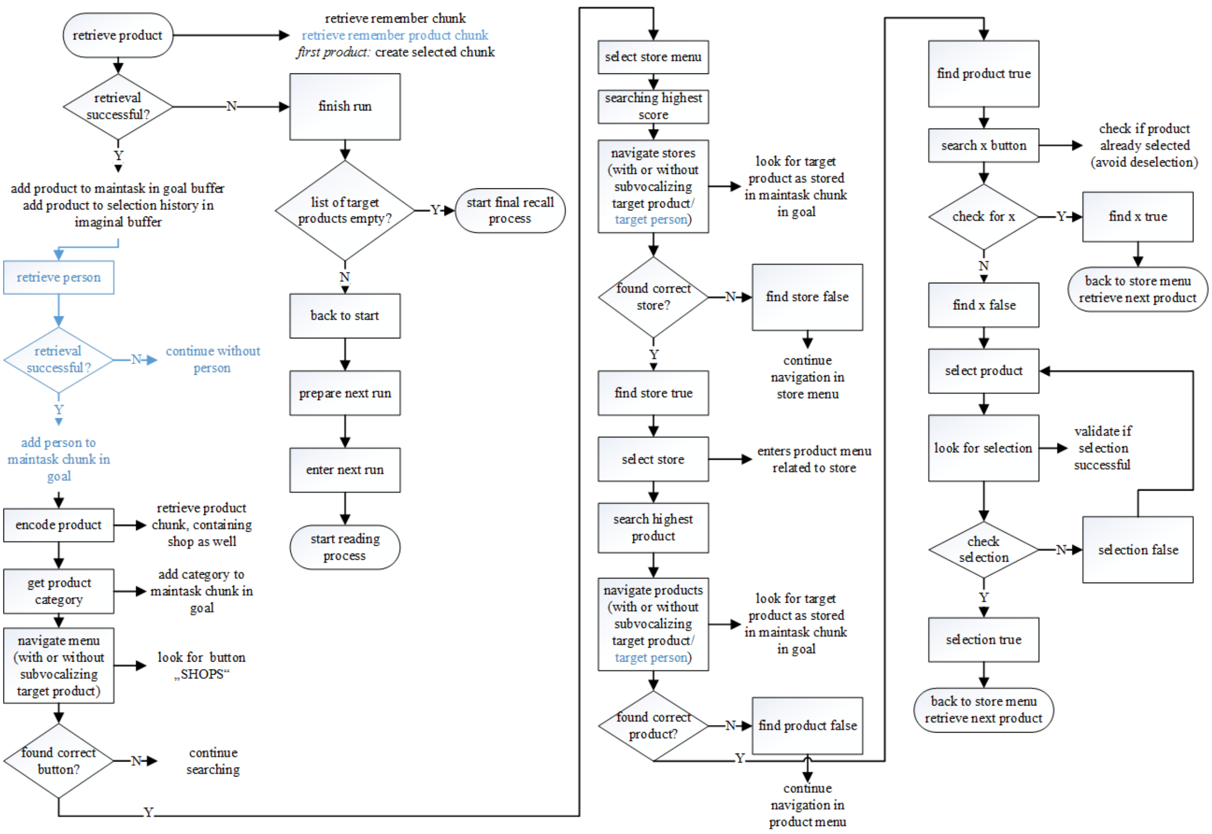

Figure 9. Navigation and selection process implementation in ACT-R. The blue color indicates distinct or additional features within the high workload variation of the model.

3.2.3.3 Interruption and resumption

In case an interruption occurs, it is immediately detected within a bottom-up process, i.e. cognitive processing is directly triggered by a certain perceptual input (Städtler, 2003), due to its remarkable salience. In terms of coding, such behavior is implemented by setting the default visual location to yellow colored objects (*set-visloc-default isa visual-location color yellow*). In consequence, objects meeting this requirement are added to the visual-object buffer as soon as they appear. The advertisement is detected, and at the same moment the goal buffer, previously filled with information on the main task, gets emptied and now comprises the interruption instead. According to Trafton, Altmann and Ratawni (2011), whose model *"clears out all state information from the primary task"* in line with the changing screen content, retrieval and imaginal buffer are cleared as well. After reading the advertisement message word by word and a short sleep period of five seconds, representing a period of decision making, "YES" or "NO" as a reaction to the offered product is chosen randomly, and the appropriate button searched for and selected. This action leads back into the previous shop menu and forces the application of one of the potential resumption strategies described in *section 2.1.3*.

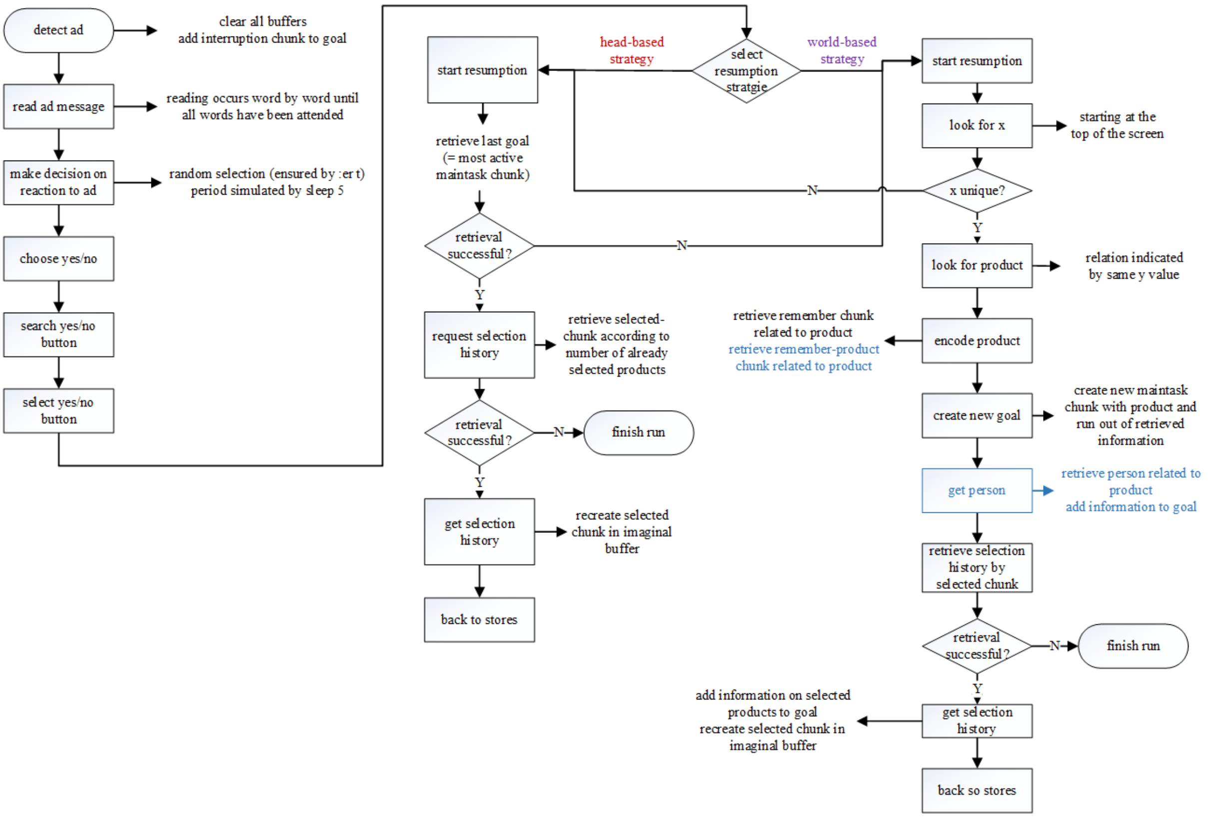

Figure 10. Interruption and resumption process implementation in ACT-R. The blue color indicates distinct or additional features within the high workload variation of the model.

As depicted in *Figure 10*, the head-based strategy tries to retrieve the lastly used goal chunk, and in this vein gets the correct history of already selected products. On the contrary, the world-based strategy consists in searching for the last selection mark within the product menu, encoding the related product and trying to reconstruct the last goal by retrieving the current run, and in the high workload version the related person as well. Based on this information, the opportunity arises to retrieve the previous selection history. By this means, in both cases a procedure of problem-state recall occurs, just as described by Salvucci and Taatgen (2010). While the head-based strategy can be applied without constraints as long as the retrieval of the last goal succeeds, the latter one is applicable only in the case of unique world-based knowledge, i.e. just a single selection mark within the product menu. If there is more than one selection mark, a switch towards the head-based strategy occurs. Within the low workload condition, both resumption strategies are chosen with equal frequency within a random selection process. In contrast, in the high workload condition the world-based strategy is applied more often for the reasons outlined in *section 2.1.3*. On this account, the utility of the production responsible for starting the world-based strategy is set to the value five *(spp world-start-resumption :u 5)* to ensure its preferred application in the high workload condition.

3.2.3.4 Final recall

A visual impression of the final recall procedure can be found in *Figure 11*. Obviously, after completing the last run, i.e. the list of products to be searched for and selected is empty, the final recall starts by changing the goal into the final recall chunk, and retrieving the remember chunk (respectively remember-product chunk for high workload) for the product with the highest activation and vocalizing the product name. In the high workload condition, a retrieval of the remember-person chunk that indicates the person related to the product – accessible by comparing the line and run slots – hooks up, again followed by verbal feedback.

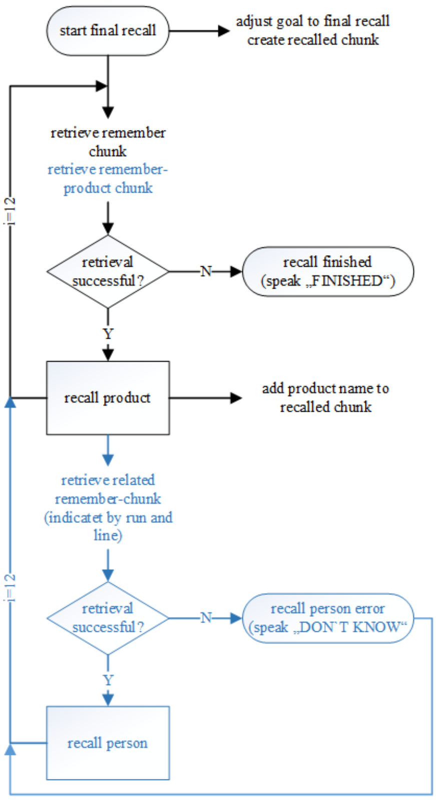

Figure 11. Final recall process implementation in ACT-R. The blue color indicates distinct or additional features within the high workload variation of the model.

Thereby, the final recall process is assumed to be controlled by the product in the high workload condition as well, i.e. the recall of the product is followed by that of the related person, not the other way round. In consequence, the products for different people are recalled in mixed sequence, regarded as being closer to actual human behavior. Comparable to the product search and selection process, the already recalled products are saved in a prepared chunk in the imaginal buffer, holding a slot for each selected product. As soon as there is no product related chunk activated highly enough to be retrieved left, the final recall and as well the task ends by speaking “FINISHED”.

3.2.4 Adjusted subsymbolic parameters

As the goal of the current work consisted of designing a model, able to perform the task without errors, i.e. able to remember all target products, under conditions of low workload and without the occurrence of an interruption, a few parameters related to memory retrieval were adjusted. First of all, the retrieval threshold *:rt*, which influences the accessibility of information in declarative memory, was set to -1.0 and the retrieval latency *:lf*, which determines how fast a chunk can be retrieved from declarative memory, to 0.75. Moreover, the base-level learning parameter *:bll* received the value 0.5, according to personal communication with Nils Taatgen and the ACT-R tutorial a default value, inspected and broadly validated in related research, e.g., by Trafton et al. (2011). Finally, the parameter to enable randomness *:er* was set true, permitting random choices between productions working on the same constraints, e.g., applied for navigating with or without subvocalizing or reacting to the product offer with “YES” or “NO”.

3.2.5 Model assumptions

Regarding the task related behavior described in *section 3.1.5*, and with reference to the hypotheses outlined in *section 2.5*, a substantial decrease in performance due to the appearance of interruption is expected. In detail, this should result in longer product selection times, fewer selected products and extended resumption times. Regarding interruption times, no explicit assumptions are made since potential differences in the reading process of the advertising message are not part of the model. In the case of an enhanced mental workload, the task performance should be worse in terms of prolonged product selection times, less selected products, increased resumption times, and fewer products remembered within the final recall. Again, interruption times are not inspected closer for the previously stated reasons. Moreover, an interaction between interruption and mental workload is anticipated, i.e. task performance should be worst in high ad runs under conditions of high workload. Concerning resumption strategies, the already explained difference in terms of workload is expected, stating a strong preference for the world-based strategy under high workload, and an equal application of head-based and world-based strategy in the low workload variation.

3.3 Experiment

As already mentioned, a human experimental setting serves to validate the predictions derived from the data generated by the model. It is mainly based upon the task setting described in *section 3.1*, but entails some additional aspects, exemplified subsequently.

3.3.1 Participants

The study "Shopping with the smartphone" [Original German title: "Einkaufen mit dem Smartphone"] was conducted with 62 human participants at the Institute of Psychology and Ergonomics' laboratory of the Technische Universität Berlin. They received either an allowance of € 10 or one experimental subject hour, and were recruited via the participant tool of the former graduate school Prometei[1] as well as personal contacts. About 66% of the participants were female, and 71% stated that they were students. As no specific assumptions were made regarding age, an ordinary adult sample aged 20 to 49 years (*M* = 28.53, *SD* = 7.16 years) was tested. To rule out errors due to misunderstanding the presented instructions, only native German speaking participants or those with close to native German speaking skills were included.

3.3.2 Design

Hypotheses were tested using a 2x3-factorial, multivariate design with mental workload (high vs. low workload) and frequency of interruption (no vs. low vs. high ad) as independent variables. Whereas the former aspect was assessed between-subjects, the latter one depicted a within-subjects factor. The aspects of task performance described in *section 3.1.5* – product selection time, selected products, interruption time, resumption time and final recall – as well as resumption strategies served as dependent variables. Additionally, working memory capacity and affinity for technology were regarded as potentially confounding variables.

The shopping list application described in *section 3.1.1* was used for the product search and selection task. The entire task was embedded into a short story to make the situation more plausible and realistic. Within this context, the level of mental workload was varied by inducing further information the participants had to remember, just as outlined in *section 3.1.4*. For manipulation check, the level of experienced workload was assessed with the NASA Task Load Index (NASA-TLX), developed by Hart and Staveland (1988). Interruption was operationalized via product advertisements, appearing as explained in *section 3.1.3*. The induction of

[1]Accessible via https://proband.prometei.de/

interruption as well as the induction of the shopping situation as a whole were checked by several questions within a structured interview.

By means of the generated log files during the smartphone use, the participant's task performance could be analyzed regarding the stated performance parameters. The extent of finally recalled products was accessible via comparing a prepared product list with the orally performed product recall. Finally, resumption strategies were addressed within the structured interview.

Affinity for technology and working memory span as potentially confounding sources of variance were controlled using standardized means of measurement. For the former the Questionnaire on Affinity for Technology (TA-EG), developed by Karrer, Glaser, Clemens, and Bruder (2009) was applied while the latter aspect was addressed with a translated and slightly modified version of the Counting Span task (CSPAN), reported by Engle, Tuholski, Laughlin, and Conway (1999).

3.3.3 Material

Apart from the described shopping task, several standardized tests and questionnaires as well as a specifically developed interview were used within the experimental setting. In order to reduce complexity, all measures and their characteristics are described separately.

3.3.3.1 Shopping task

The shopping task was conducted with the previously described shopping list application (see *Figure 3*), using a LG Google Nexus 4 smartphone with a screen size of 4.7", a display resolution of 1280 x 768 pxl, pixel density of 319 ppi and Android 4.4.2 (KitKat) serving as operating system. Instructions were presented via Microsoft PowerPoint 2007 on a Desktop Computer using Microsoft Windows XP Professional ServicePack 3 with a maximum display resolution of 1280 x 1024 pxl.

In order that participants spend cognitive effort on the interruption, a respective task scenario was created. It asked the user to imagine being the virtual person Diana – in case of male participants changed to Dennis to foster identification – who conducts shopping by using a shopping list application related to a shopping center close to the University Campus. Participants are provided with some information on their character, a 24-year-old student of architecture, who plays the clarinet, participates in a neighborhood care project and loves to cook with friends. To stimulate involvement in the interruption, information on the intended shopping behavior – buying special offers in about half of the cases – was given. In order to induce various levels of workload, shopping is done either just for Diana/Dennis herself/himself in the case of low workload, or additionally as a favor for the old neighbor Norbert and the ill friend Fiona in the case of high workload. In the latter case, participants are provided with some information on Norbert, a 70-year-old, retired teacher, with a dog and who suffers from a slight walking impairment, and Fiona, a 26-year-old fellow student, who takes part in shared exam preparation, loves to cook as well but is currently suffering from a severe flu. A detailed outline of the used instruction material can be found in *Appendix A1.2*.

The task itself consisted of accompanying Diana respectively Dennis during a usual day in life and conducting the search and selection of the target products as described in *section 3.1.2*. Since already existing experience in using the shopping list application was assumed, participants performed an additional run with four products without interruption. In doing so they had to buy the newspaper Berliner Morgenpost (for Norbert), iron pills (for Fiona), soured mild (for Norbert), and cornflakes (for Diana/Dennis). This was regarded as sufficient to familiarize participants with the product categories and the general handling of the application. For each participant a log file was generated during task execution, recording relevant events, i.e. button press, menu change, product selection, onset and offset of the product advertisement, with their respective times of occurrence. Those data were used to compute the performance parameters described in *section 3.1.5*. The recall of all previously selected products in the end of the shopping task was based on the instruction that the smartphone of Diana/Dennis, containing the list of previously selected products, ran out of battery. Correctly recalled products – in the high workload conditions just those with correctly indicating of the related person as well – were marked by the experimenter on a prepared list.

3.3.3.2 Structured interview

A structured interview served as a qualitative measure for assessing resumption strategies, as well as manipulation check for spending cognitive engagement on the task and especially the product advertisement. It consisted of 11 questions in the low, respectively 13 questions in the high workload condition, one to three questions on the visualization of Diana/Dennis, Norbert and Fiona, four questions on the disruptiveness, plausibility and handling of the interruption, four questions dealing with the previously assumed head-based respective world-based resumption strategy, and one question on how participants tried to remember the target products. A detailed list of all questions can be found in *Appendix A1.3*.

3.3.3.3 Counting Span task (CSPAN)

The CSPAN task was developed for measuring the individual working memory span. It was applied within the experimental setting on purpose of controlling the participant's working memory capacity to rule out potential confounding effects. An already existing German translation by Tobias Staudigl, suitable for E-Prime version 1.1, was used on the previously mentioned Desktop Computer with E-Run version 1.1.4.6 and an average display refresh rate of 60.31 Hz (*SD* = 0.01 Hz). The used CSPAN version comprises 60 test and six practice screens in 15 test and three practice trials, with a randomly arranged set of three to nine dark blue circles as target shapes, and one to nine dark blue squares as well as one to five light blue circles as distractors (Conway et al., 2005). The latter ones share either shape or color with the targets, requiring conjunctive search while counting the targets. An example screen showing target shapes as well as both kinds of distractor shapes can be found in *Figure 12*.

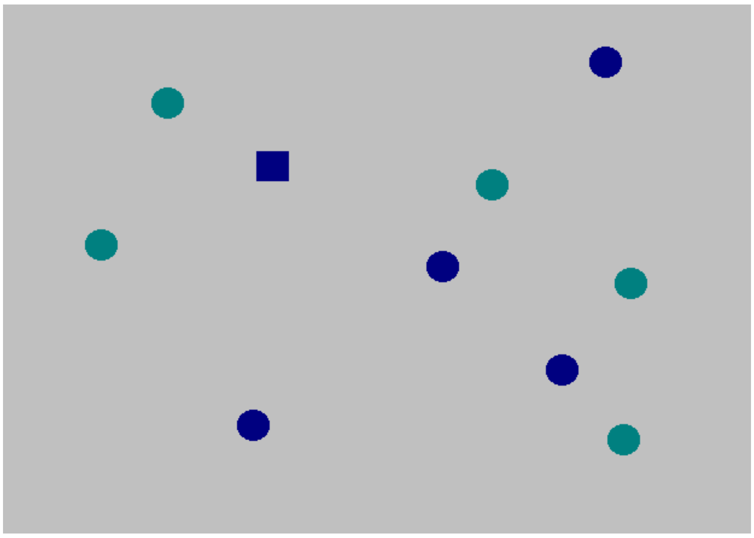

Figure 12. Example display of the CSPAN task.

Each trial contains a set of two to six screens, each set size appearing three times, completed by a screen showing three question marks in the center. Following detailed screens of instructions with self-paced navigation, the participant's task consists of counting the target shapes, i.e. the dark blue circles, speaking each number as well as the final count aloud, e.g., saying "one, two, three, four, four" in the case of the screen displayed above. After telling and remembering the final number, the experimenter releases the next screen. By the time the three question marks appear, indicating the end of the trial, the remembered numbers of the respective screens have to be written down on a prepared answer sheet in their serial order of occurrence.

Unlike the procedure described by Engle et al. (1999), in the current context the screen onset and offset was controlled by the participants via pressing the space key, due to technical constraints. Nevertheless, they were clearly instructed to proceed immediately after stating the final count and reproved by the experimenter if they tried to violate this instruction. But in most cases participants observed rules and stuck straight to the instructions. Regarding reliability and validity of the CSPAN task, Kane et al. (2004) report an internal consistency (coefficient alpha) of α = .77 within a tested sample size of 236 participants, whereas Conway et al. (2005) describe substantial correlations to other measures of working memory capacity, e.g., r = .66 between CSPAN and reading span (RSPAN) task or r = .60 between CSPAN and different transformation span tasks.

3.3.3.4 Questionnaire on affinity for technology (TA-EG)

The TA-EG assesses affinity for technology with 19 items, merged into the four subscales “enthusiasm for technology” (five items), “competence in dealing with technology” (four items), “positive impacts of technology” (five items), and “negative impacts of technology” (five items). In the given context it served as control measure as well, to rule out differences in participant’s task performance just due to discrepancies in attraction to the technical device. Items are presented in mixed sequence, and have to be rated by the participant regarding the respective strength of application on a scale ranging from one (“applies not at all”) to five (“applies completely”). Within the subscale “enthusiasm for technology” the items deal with information about, trying and buying new technical devices, whereas the subscale on “competence in dealing with technology” asks for knowledge about functions and handling of technical devices as well as understanding appropriate magazines. “Positive impacts of technology” regards electronic devices besides others as helpful in searching for information or fostering security, while “negative impacts of technology” blames electronic devices, e.g., for causing stress, mental depletion and illness. According to Karrer et al. (2009), the subscale “enthusiasm for technology” shows an internal consistency of α = .842 and “competence in dealing with technology” a coefficient α of .789. “Positive impacts of technology” is reported as possessing an internal consistency of α = .722, whereas the respective coefficient for “negative impacts of technology amounts to α = .747. Regarding validity, the authors report significant correlations with scales on competency with control beliefs in handling technology (Beier, 1999) and enthusiasm in domain specific innovativeness (Goldsmith & Hofacker, 1991).

3.3.3.5 NASA Task Load Index (NASA-TLX)

The multi-dimensional NASA-TLX aims to address the experienced mental workload while performing a task with a set of six subscales related to task, behavioral and subjective characteristics (Hart & Staveland, 1988). Within the conducted experiment, it was mainly used as manipulation check to the mental workload induction described in *section 3.1.4*. In detail, it comprises “mental demand”, i.e. the extent of mental and perceptual activity required to solve the task, “physical demand”, i.e. the extent of physical activity required to solve the task, “temporal demand”, i.e. the extent of time pressure experienced during the task, “performance”, i.e. the individual satisfaction in accomplishing the goals of the task, “effort”, i.e. the energy necessary to solve the task, and “frustration level”, i.e. the extent of stress, irritation, annoyance and so on experienced while dealing with the task. Participants have to mark the individual extent of application on a bipolar scale, ranging from low to high extent respectively good to poor in the case of “performance”. An unweighted sum score is computed out of the subscales, commonly used in various settings and stated to be highly correlated with the weighted score (Cao, Chintamani, Pandya, & Ellis, 2009). According to Battiste and Bortolussi (1988), the NASA-TLX achieves a reliability for repeated measures of $r = .77$ and the authors also report high correlations with other measures of workload, e.g. the Subjective Workload Assessment Technique (SWAT) by Reid and Nygren (1988), indicating high convergent validity.

3.3.4 Procedure

Experimental data were generated within individual testing sessions with an average session duration of 43 min (SD = 6 min), ranging from 30 min in the fastest to 65 min in the slowest case. After being welcomed and signing the consent form, participants answered a written questionnaire on demographic details. Then the CSPAN task was presented at the computer, whereas the following TA-EG questionnaire was conducted in paper-pencil form. Afterwards participants were introduced to the shopping task with the provided smartphone by means of a computer-based presentation. They completed that task without time constraints and subsequently rated the experienced workload with the NASA-TLX. Information on resumption strategies as well as manipulation checks on task behavior were collected then within a structured interview. Finally, participants received their allowance, were thanked, and approved.

3.3.5 Scoring

The performance parameters were computed based on the log files generated within the shopping task, just in the way described in *section 3.1.5*.

Analysis of the structured interview was conducted by means of an analysis of content (Diekmann, 2007, p. 576 ff.), creating categories out of the participant's answers and assigning each answer to one of such categories. In the following, frequencies of the respective categories – "Yes", "No", "Don't know" in the easiest case, and options like "reconstruction of previous selection", "focus on next product", "don't regard advertisement as interrupting" if more complex categories were required – were calculated, both for resumption strategies and manipulation checks.

The CSPAN score was computed using an all-or-nothing scoring procedure (Conway et al., 2005), based on the *"cumulative number of digits recalled from perfectly recalled trials"* (Engle et al., 1999, p. 316). According to Conway and colleagues, such a scoring procedure depicts a frequently applied approach when inspecting working memory span measures (Conway et al., 2005).

Scores for the subscales of the TA-EG were computed by summing up the raw values of the corresponding items and calculating the respective mean. According to Karrer et al. (2009), one item within the scale on "competence for technology" and the entire scale "negative impacts of technology" have to be reversed in polarity before conducting this calculation.

As mentioned above, the NASA-TLX score was computed as unweighted sum of the six subscale ratings. Those were measured with millimeter accuracy according to the participants marking on the scale, and the resulting value was adjusted to the original scale length of 100 mm.

4 Results

Model as well as participants generated the behavioral data that are subsequently analyzed. The individual depiction of model and participant data will be followed by a visual and numerical comparison of both.

4.1 Model predictions

As stated in *section 2.5*, the postulated hypotheses served as framework for developing the ACT-R model. Based on the obtained model behavior, *Table 1* shows the results of the descriptive analyses for the extracted performance parameters.

Table 1

Descriptive statistical values of the performance parameters product selection duration, number of selected products, interruption time, resumption time, and finally recalled products in model runs with different intensity of interruption, divided by high and low workload and overall

	Level of workload	**No ad**		**Low ad**		**High ad**		**Overall**	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Product selection time (in sec)	H	7.42	0.64	7.31	0.73	7.58	0.77	7.41	0.28
	L	6.09	0.64	6.09	0.69	6.10	0.72	6.05	0.20
	-	6.76	0.93	6.70	0.94	6.84	1.05	6.73	0.73
Selected products (sum)	H	3.67	0.48	3.03	0.18	3.00	0.00	9.70	0.47
	L	4.00	0.00	3.77	0.43	3.30	0.47	11.07	0.52
	-	3.83	0.38	3.40	0.49	3.15	0.36	10.38	0.85
Resumption time (in sec)	H			3.08	0.68	3.95	0.36	3.66	0.29
	L			2.65	0.27	2.72	0.26	2.69	0.20
	-			2.86	0.56	3.33	0.69	3.18	0.54
Interruption time (in sec)	H			1.74	0.08	1.72	0.10	1.72	0.07
	L			1.71	0.10	1.71	0.10	1.71	0.06
	-			1.73	0.09	1.71	0.97	1.72	0.06
Final recall (in %)	H							60.56	9.01
	L							84.17	6.32

Note. H: high workload (data based on n = 30 model runs), L: low workload (data based on n = 30 model runs), -: no separation by workload (data based on N = 60 model runs).

Altogether, the model performance seems to be sensitive for the induction of interruption and workload. In the case an interruption occurs, fewer products are selected, and product selection takes slightly longer. Such effects show up especially with the increasing frequency of interruption. Moreover, without enhanced workload the model clearly performs better across all performance-related measures. Additionally, it prefers the world-based strategy when resuming under high workload. A more detailed outline of the attained results with respect to the initially specified hypotheses is done subsequently.

4.1.1 Hypothesis 1: Main effect of interruption

Regarding product selection time values without considering workload, there seems to be a slight difference for the amount of interruption. At least in the case of high ad trials, *Table 1* points towards the assumed direction. The sum of selected products indicates a precise negative linear trend, i.e. decreasing scores with increasing amount of interruption, exactly as expected. For resumption times, again the statistical values seem to support the hypothesized tendency with a substantially higher duration in trials with more interruptions. Finally, since there was no clear assumption on differences in interruption time, no attempt at creating such behavior was made. In consequence, the model shows equal durations for both amounts of interruption.

In summary, model data indicate the assumed loss in performance due to rising interruption in the case of the number of selected products and the resumption time. Weaker evidence persists for product selection time, whereas interruption time was not considered to be different.

4.1.2 Hypothesis 2: Main effect of mental workload

Descriptive values for the overall comparison between both workload conditions are shown as well in *Table 1*. Obviously, the high workload variation is characterized by a considerably longer product selection time, fewer selected products, and a longer resumption time. Again, no difference in the case of the interruption time can be found for the already explained reasons.

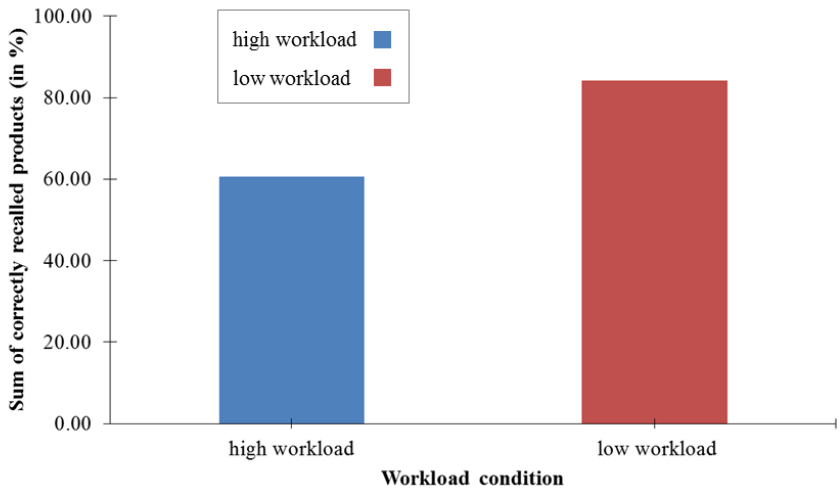

Figure 13. Model data on final recall in high and low workload variation.

For the amount of finally recalled products, a noticeable difference is evident as well, graphically presented in *Figure 13*. As depicted, with an enhanced level of workload just about 60% of the previously selected products can be recalled. In contrast, in the low workload variation a successful recall for more than 80% of the target products occurs.

In conclusion, model data point towards the hypothesized direction for all performance-related measures apart from interruption time.

4.1.3 Hypothesis 3: Interaction between interruption and mental workload

In order to shed light on the interaction between the amount of interruption and the level of workload, *Table 1* receives graphical support by *Figure 14*. Apparently, for the product selection time in the high workload variation exists a slight increase towards highly interrupted trials, whereas values stay at a comparable, and at the same time lower level all the time in the low workload variation.

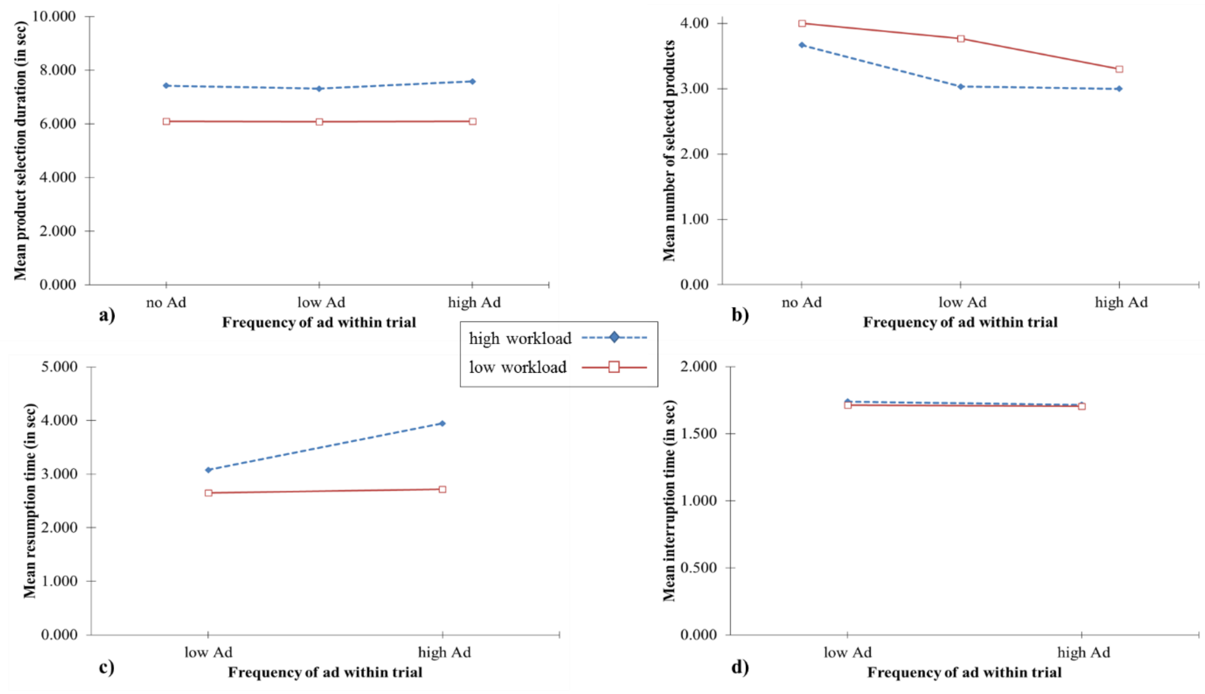

Figure 14. Model data on task performance under high and low workload. **a)** Product selection time, **b)** Selected products, **c)** Resumption time, **d)** Interruption time.

The sum of selected products forms a decreasing line with rising interruption frequency in both workload conditions. As expected, it holds a considerably higher level for the low workload variation, and graphs develop distinctively respective to their level of workload. In the case of low workload, trials without and with a low frequency of interruption bear just a slight difference, whereas the difference towards high frequency of interruption increases. In contrast, for high workload, the main difference exists between trials without and with interruptions, regardless of the frequency. Taking a look at the resumption time, there is no noteworthy distinction in the case of a low and high amount of interruption in the event of low workload, but for high workload there is a noticeable increase towards the higher frequency of interruption. As already mentioned, there is no difference regarding workload for interruption time. Therefore, both lines within the graph rest on each other.

In summary, evidence for an interruption of both factors becomes obvious especially in the case of the number of selected products and resumption time.

4.1.4 Hypothesis 4: Difference in resumption strategies

As indicated by *Figure 15*, compared to the high workload variation, head-based and world-based strategy use is suggestive as being more balanced in the low workload variation. However, the head-based strategy seems the preferable one, used twice or more in 70.33% of the model runs. In contrast, the world-based strategy depicts a usage of just once or even less in 63.33% of the model runs.

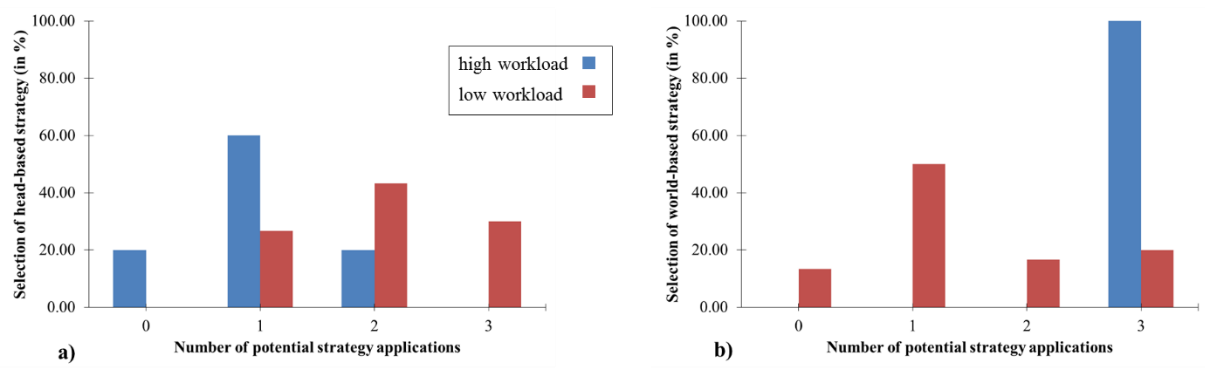

Figure 15. Model data on resumption strategies for high and low workload. **a)** Application of the head-based strategy. **b)** Application of the world-based strategy.

On the contrary, within the high workload variation the world-based strategy is preferred all the time and the head-based strategy is only applied in the case the former one is not applicable, i.e. knowledge in the world is not unique. As can clearly be seen, in 60% of the performed resumption procedures, a change towards the head-based strategy occurs at least once.

To summarize, there is indeed a difference in strategy use related to workload variation in the assumed direction.

4.2 Experimental results

Nearly 84% of the participants were smartphone users with an average usage time of 2.91 hours daily (*SD* = 2.60 hours) respectively *M* = 20.06 hours weekly (*SD* = 18.12 hours) and *M* = 95.13 hours monthly (*SD* = 131.72 hours). Regarding the purpose of use, mainly communicative intentions, i.e. checking and sending emails and short messages or using instant messenger services, were reported (88.7%), followed by searching for information (e.g., internet browsing, reading newspapers) with 35.8% and navigation respective using applications related to public transport (24.5%).

Table 2

Descriptive statistical values of the performance parameters product selection duration, number of selected products, interruption time, resumption time, and finally recalled products in human runs with different intensity of interruption, divided by high and low workload and overall

	Level of workload	No ad			Low ad			High ad			Overall		
		N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
Product selection time (in sec)	H	31	9.32	5.72	31	9.58	5.89	31	10.07	7.44	31	9.61	4.63
	L	31	9.28	4.35	31	8.55	3.79	31	9.71	8.68	31	9.21	4.39
	-	62	9.30	5.04	62	9.06	4.94	62	9.89	8.02	62	9.41	4.48
Selected products (sum)	H	31	3.81	0.48	31	3.81	0.48	31	3.77	0.50	31	11.39	0.88
	L	31	3.90	0.30	31	3.77	0.50	31	3.81	0.48	31	11.48	0.85
	-	62	3.85	0.40	62	3.79	0.48	62	3.79	0.48	62	11.44	0.86
Resumption time (in sec)	H				26	3.45	2.63	26	4.26	2.35	29	3.88	1.83
	L				28	4.41	3.20	28	4.30	3.18	31	4.47	2.79
	-				54	3.95	2.95	54	4.28	2.79	60	4.19	2.37
Interruption time (in sec)	H				29	6.46	3.69	30	6.95	3.88	31	6.75	3.05
	L				30	5.81	3.39	30	6.90	2.97	31	6.54	2.36
	-				59	6.13	3.53	60	6.92	3.42	62	6.64	2.71
Final recall (in %)	H										31	51.01	25.01
	L										31	73.59	14.50

Note. H: high workload, L: low workload, -: no separation by workload

The majority of participants seemed to successfully put themselves in the position of the created task scenario, as 90.3% in the high respective 80.6% in the low workload condition reported a visualization of the main character Diana/Dennis in either characteristics or appearance or both. Similarly, nearly all participants in the high workload variation succeeded in visualizing Norbert (100.0%) and Fiona (93.5%) in one or both ways. Only two participants within each condition (6.5% in each case) stated to have had severe difficulties in performing the task, while the others reported no or just slight difficulties. Reported difficulties were mainly due to non-intuitive shop categories (25.8% in the high and 29.0% in the low workload variation), the demand to finally recall the selected products (19.4% in the high and 22.6% in the low workload variation), and the requirement to additionally remember the person related to the product in the high workload variation (16.1%).

Descriptive results regarding the performance related dependent variables are shown in *Table 2*. As displayed, for interruption and resumption time, smaller samples had to be used due to missing values. Moreover, in the case of resumption time, two participants within the high workload variation had to be excluded for reasons of outliers, i.e. values beyond three SD from the mean (Rey, 2009).

Working memory span and affinity for technology were inspected as well, but found to hold comparable levels in the high and low workload condition, $-1.589 < t(60) < .601$, $.115 < p < .966$. Moreover, no systematic correlations with the dependent variables occurred. Following Bortz (2005), both aspects are mandatory for including a covariate, therefore those measures were expelled in the subsequent analyses.

4.2.1 Hypothesis 1: Main effect of interruption

Taking into account the structured interview on purpose of manipulation check, most participants regarded the induced interruptions as severe or at least slightly disruptive (71.0% in the high and 61.3% in low workload variation), and as suitable or at least partly suitable within the given context (83.9% in both workload variations). Additionally, the majority of the participants reported handling the product advertisement in a different way to everyday situations (67.7% in the high and 58.1% in the low workload condition), either regarding the recognition of the offered product, the resulting shopping behavior, or both.

Although the descriptive values indicate a slight difference between runs with and without interruption, the Greenhouse-Geisser corrected result of the ANOVA for product selection time, $F(1.554, 93.255) = .395$, $p = .623$, $\eta_p^2 = .007$, does not indicate the existence of a main effect of interruption. A Greenhouse-Geisser correction was performed in this case due to the missing sphericity, pointed out by the Mauchly-test. For the amount of selected products, both descriptive and ANOVA results state no significant difference, $F(2,120) = .448$, $p = .640$, $\eta_p^2 = .007$. In the case of resumption as well as interruption time, there seems to be an effect towards the expected direction, at least on a descriptive view, but the ANOVAs do not indicate an adequate statistical significance with $F(1,52) = .474$, $p = .494$, $\eta_p^2 = .009$ for resumption time, and $F(1,55) = 1.419$, $p = .239$, $\eta_p^2 = .025$ for interruption time.

In sum, the first hypothesis stays unsupported for all four performance related measures.

4.2.2 Hypothesis 2: Main effect of mental workload

For the workload induction, the NASA-TLX served as manipulation check. Although *Figure 16* displays a slightly higher score for nearly all scales in the high workload variation, especially for the overall sum score, no statistical significance for those differences was revealed, $-.620 < t(60) < 1.619$, $.111 < p < .605$.

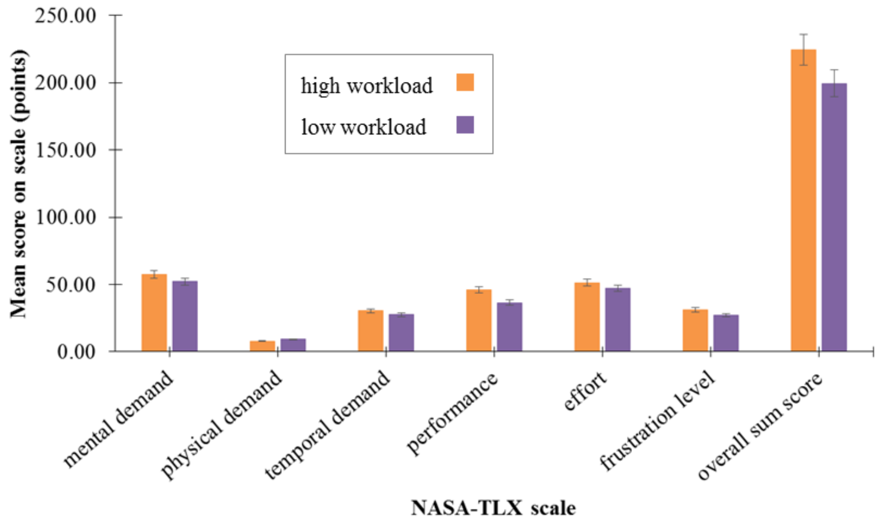

Figure 16. Scores on the NASA-TLX subscales and overall sum score. Error bars represent 95% confidence intervals on human data.

As is obvious in *Table 2*, product selection time shows a slight difference between both variations, but it yields no statistical significance, $F(1,60) = .176$, $p = .676$, $\eta_p^2 = .003$. In the case of selected products, high and low workload variation seems to achieve nearly comparable levels. Again, no statistical significance can be revealed, $F(1,60) = .193$, $p = .662$, $\eta_p^2 = .003$. Results for resumption and interruption times contradict expectancies as well, even assumable by means of the visual impression, but statistically indicated within the results of the ANOVAs. In the latter one, for resumption time, $F(1,52) = .698$, $p = .407$, $\eta_p^2 = .013$, and interruption time, $F(1,55) = .278$, $p = .600$, $\eta_p^2 = .005$, no significant result is achieved.

For the final recall performance, the descriptive comparison is shown in *Figure 17*. Obviously, without an enhanced level of workload, participants were able to recall on average more than 70% of the previously selected products. In contrast, in terms of increased workload this number amounts to just about 50%. The descriptively indicated difference between high and low workload variation holds statistical significance as well, $t(48.122) = -4.350$, $p < .001$, $d = -1.41$. As the Levene test claims the absence of homogeneity of variances, the corrected *df* for the t-test value are reported here.

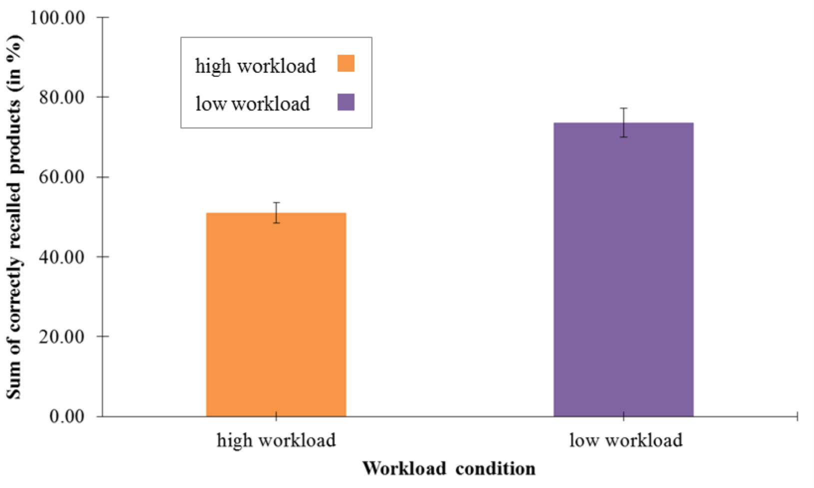

Figure 17. Human data on final recall under high and low workload. Error bars represent 95% confidence intervals on human data.

In sum, the second hypothesis can be regarded as failed for all performance measures apart from the finally recalled products. For the latter one, there indeed seems to be a difference in the expected direction.

4.2.3 Hypothesis 3: Interaction between interruption and mental workload

Figure 18 gives an impression on changes in human task performance due to varying the intensity of interruption and mental workload. Regarding product selection time, values within the high workload variation form a slightly ascending line with increasing interruption frequency. On the contrary, for low workload an initial decrease towards the low ad run is followed by an increase towards the high ad run. In terms of selected products, both lines stay at a comparable level all the time. Approaching resumption time, values marginally drop with higher interruption frequency for low workload, but indeed display the expected rising trend for high workload, though participants resumed faster in the low ad run. Finally, inspecting the graph for interruption time reveals marginally rising lines for both levels of workload, starting with a slightly higher value for high workload. Although *Figure 18* as well as *Table 2* could support the assumption of an existing interaction between interruption and workload on this account, at least for resumption and interruption times, such an effect is not found for any of the performance measures within the ANOVAs. Statistical values for product selection time are

$F(1.554, 93.255) = .141$, $p = .815$, $\eta_p^2 = .002$, again Greenhouse-Geisser corrected due to the missing sphericity within the Mauchly-test. The number of selected products holds a value of $F(2,120) = .336$, $p = .715$, $\eta_p^2 = .006$, whereas for resumption time $F(1,52) = .474$, $p = .365$, $\eta_p^2 = .016$, and for interruption time $F(1,55) = .177$, $p = 676$, $\eta_p^2 = .003$ are achieved.

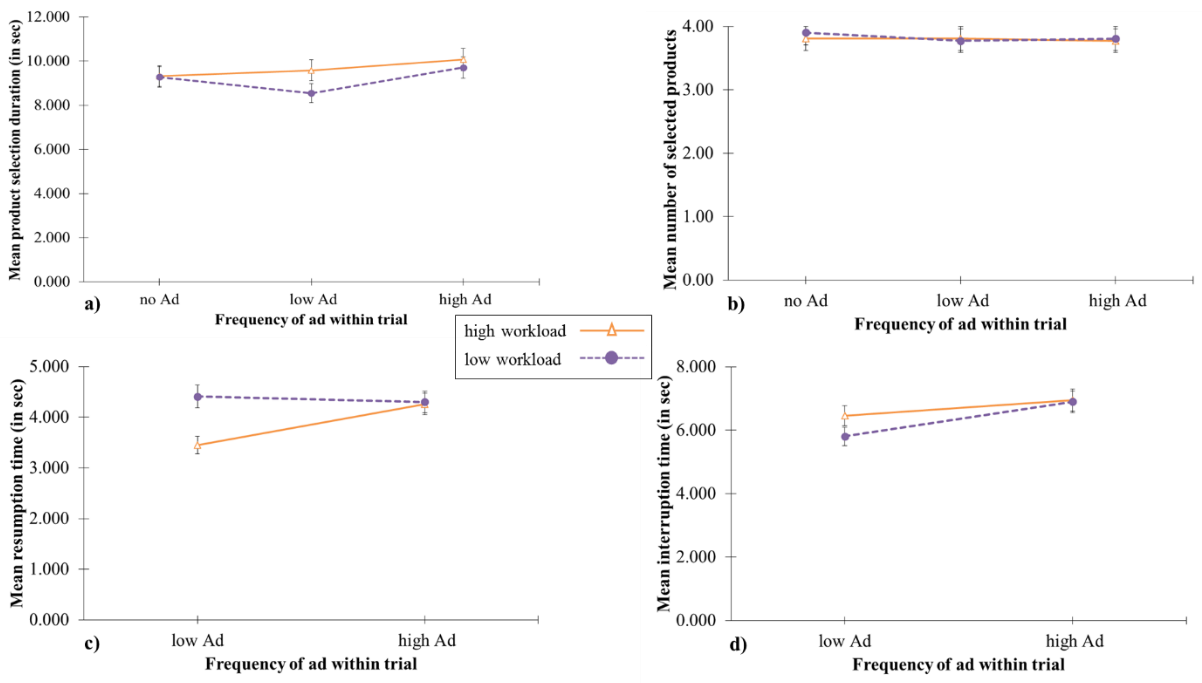

Figure 18. Human data on task performance for high and low workload. **a)** Product selection time, **b)** Selected products, **c)** Resumption time, **d)** Interruption time. Error bars represent 95% confidence intervals on human data.

To summarize, the third hypothesis ought to be regarded as not supported taking into account the results previously described.

4.2.4 Hypothesis 4: Difference in resumption strategies

As the last hypothesis yields a more explorative character, analyses are just done descriptively. They are based on the participant's answers within the structured interview at the end of the experimental session. Being confronted with the more open question *"How did you get back to the shopping task after being interrupted by the product advertisement?"*, 25.8% of the participants in the high and 38.7% of the participants in the low workload variation reported the application of any kind of cognitive strategies. Taking a closer look at the kind of cognitive strategies, 22.6% of the participants in the high and 12.9% of the participants in the low workload variation said that they tried to reconstruct their previous selection.

The values displayed in *Figure 19* are based on the more directed questions on applying the head-based and world-based strategy just as done by the ACT-R model. Obviously, those strategies are used comparably within both variations of mental workload, and in both conditions the head-based strategy seems to be preferred over the world-based one. In detail, 45.2% of the participants in the high and 41.9% of the participants in the low workload variation state that they tried to remember the product last selected when resuming the main task. This behavior was regarded as an indicator for using the strategy based on knowledge in the head. In contrast, 16.1% of the participants in the high and 12.9% of the participants in the low workload variation report the performance of a visual search on the display – pointing towards the application of the strategy based on knowledge in the world. The use of the history of successfully selected products within the current run was assumed to be part of both resumption strategies, and depicts a similar distribution in both workload variations (19.4% for high and 25.8% for low workload).

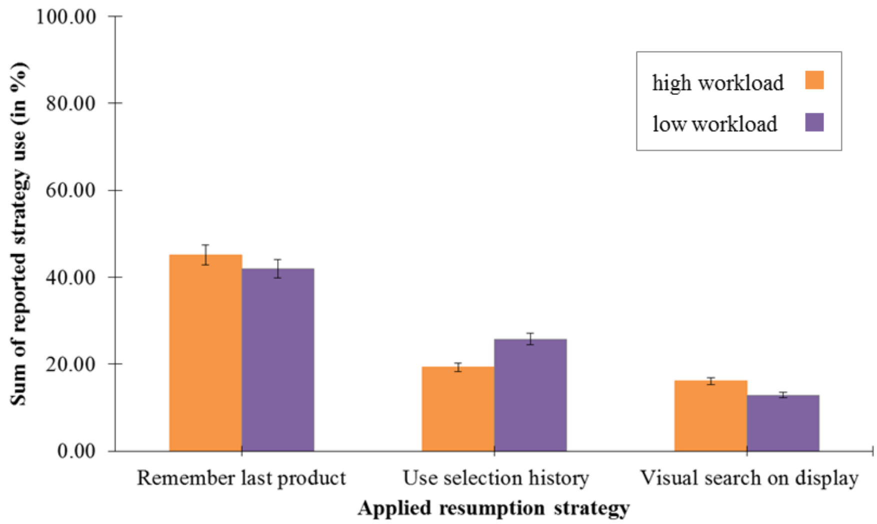

Figure 19. Human data on resumption strategies in the high and low workload condition. Error bars indicate 95% confidence intervals on human data.

Having the reported results in mind, the fourth hypothesis could be regarded as unsupported as well.

4.3 Evaluation of the model fit

To assess how good a model is qualified to represent the desired human behavior, performance indices generated during the model runs should be compared with those obvious from actual human data from a corresponding experimental setting.

4.3.1 Applied goodness-of-fit indices

Besides a graphical comparison of the model and human data, Schunn and Wallach (2005) recommend a combination of numerical goodness-of-fit measures displaying how well the relative trend magnitude is captured and those measuring the deviation from the exact location. In particular, they approve r^2 as a measure of the relative magnitude, for it relates directly to the accounted proportion of variance and better separates models with strong correlations with the data. It bears high similarity to effect size computations widely used in experimental psychological research (Cohen, 1988).

In order to assess the deviation from the exact location, the *RMSD* (root mean squared deviation) constitutes a commonly applied measure, as it can be applied to a variety of research foci and is already used regularly in corresponding research:

$$RMSD = \sqrt{\frac{\sum_{i=1}^{k}(m_i - d_i)^2}{k}}. \tag{3}$$

As displayed in *Equation 3*, it is computed as the root of the summed squared deviation between m_i, the model mean for each point *i*, and d_i, the data mean for each point *i*, divided by *k*, being the number of points compared.

In the following, comparisons of the high and low workload variation of the modeled task are done separately – with the exception of the final recall performance – since similar but distinct model code files were used for generating the data. Moreover, the focal point will mainly consist of the performance parameters, covered in hypotheses one to three, due to the more qualitative character of the strategy assessment treated in hypothesis four.

4.3.2 High workload variation

Figure 20 depicts a graphical comparison as well as the related numerical goodness-of-fit indices. Obviously, human data reside on a continuously higher level for all displayed parameters, indicating a model performance better than the performance of the human participants, except for the amount of selected products. In detail, for the product selection time, both lines bear a similar direction, marginally increasing with increasing frequency of interruption. According to Cohen (1988), the numerical comparison r^2 indicates a medium fit level for the relative trend. For the absolute location, the *RMSD* depicts a rather small value. Regarding the sum of selected products, taking a look at the graph already reveals a substantial difference between model and human data. Whereas human data stay nearly at the same level within all interruption levels, model data noticeably decrease with increasing interruption frequency. The numerical comparison depicts a poor fit in terms of the relative trend but a quite good fit for the absolute location. Considering resumption time, a graphical as well as a numerical comparison indicate a high similarity between model and human data. Nevertheless, the informative value should be treated carefully especially for r^2, as Schunn and Wallach (2005) claim to use at least three points to compare data in terms of relative fit. Otherwise *"the correlation is necessarily equal to 1.000"* (Schunn & Wallach, 2005, p. 20). For interruption time, interpretation is restricted by the fact that no definite adjustment occurred for the model. Therefore, the considerable deviation of the absolute location is unsurprising, and the perfect fit in the case of r^2 could be regarded as an artifact for the previously stated reason.

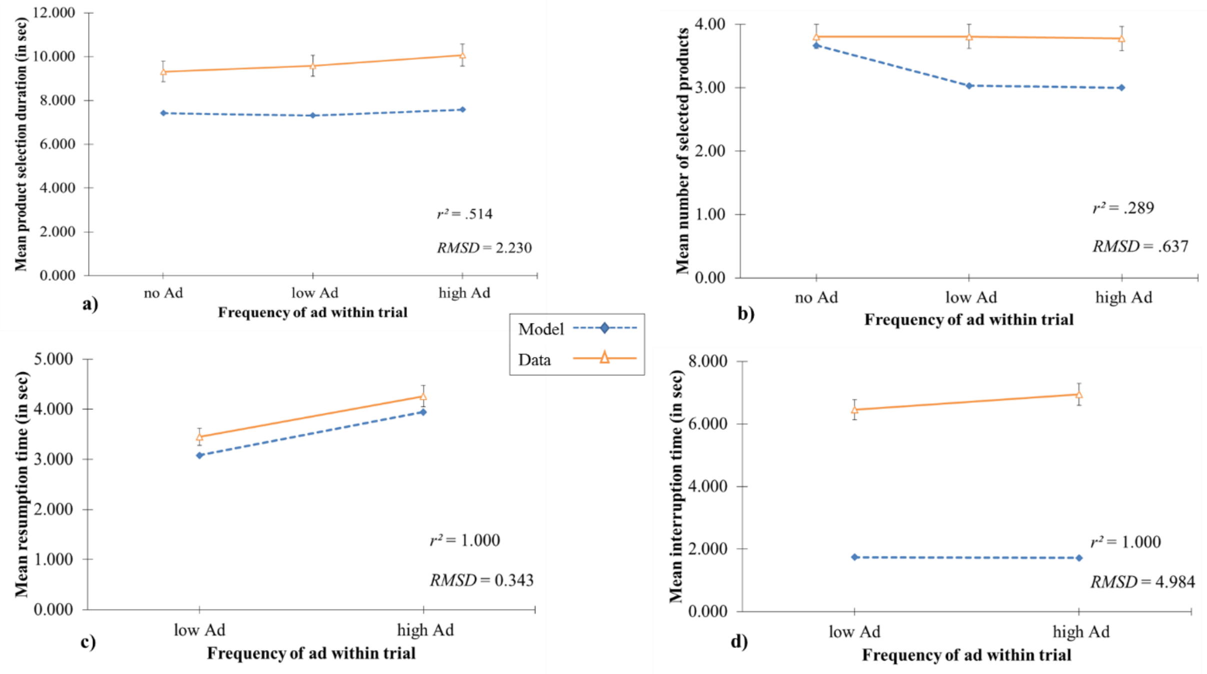

Figure 20. Comparison between model and human data on task performance under high workload. **a)** Product selection time, **b)** Selected products, **c)** Resumption time, **d)** Interruption time. Error bars represent 95% confidence intervals on human data.

In contrast to the other performance-related parameters, the final recall performance forms a single value, comparable just between high and low workload variation, and not in terms of interruption frequency. On this account, a separate report for each level of workload would not be reasonable, thus *Figure 21* shows the graphical trend as well as the numerical indices for both workload variations. Apparently, model and human data exhibit an equivalent trend with a substantially higher amount of recalled products in the low workload variation. However, the model performs considerably better in both variations, althought deviation between both sets of data is quite huge. Again, interpretation should be handled with care for the small amount of data points.

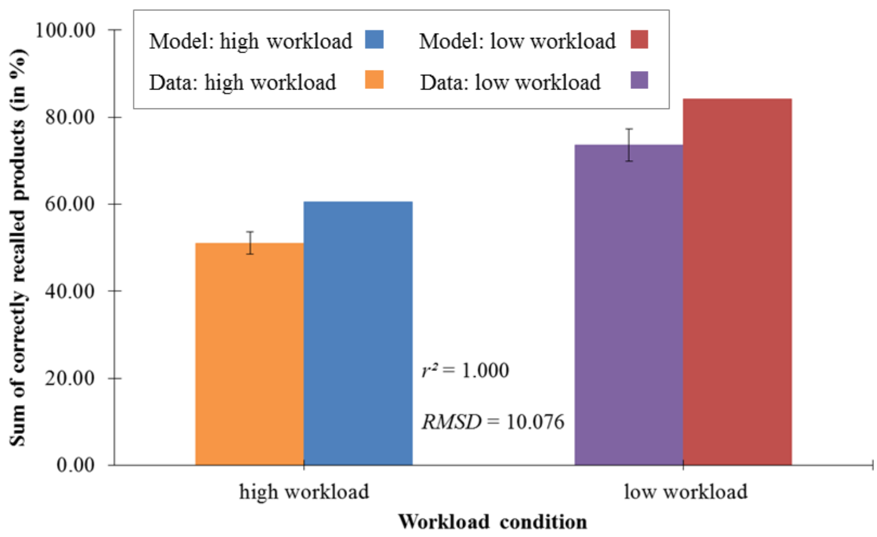

Figure 21. Comparison between model and human data on final recall for high and low workload. Error bars represent 95% confidence intervals on human data.

Regarding the difference in resumption strategies stated in hypothesis four, the modeled preference for the world-based strategy with increased workload does not become obvious within the human data. On the contrary, participants show a slight preference for the head-based strategy in the high workload variation as well.

4.3.3 Low workload variation

Graphical as well as numerical comparisons for the performance parameters in the low workload variation are displayed in *Figure 22*. Obviously, human data occupy a higher level as well within this condition, except for the amount of selected products. Having a look at the product selection time, the graph depicts a substantial difference between human and model data. Whereas model data form a nearly straight line, human data depict a considerable difference due to increasing advertisement frequency. For the numerical fit indices there is a medium fit in the case of the relative trend but a rather poor fit in the case of the absolute location. In contrast, regarding the number of selected products, both lines share nearly the same location besides the missing decrease towards the highly interrupted trials in the human data. On this account, the r^2 value is rather small, however, the *RMSD* reveals just a minor deviation. The resumption time graph shows again a comparable trend for both datasets, although the deviation is somewhat higher compared to the high workload variation. For interruption time the same applies as already explained above, apparent in graphical as well as numerical manner. Moreover, the earlier mentioned limitations due to the small amount of data points operate in both cases.

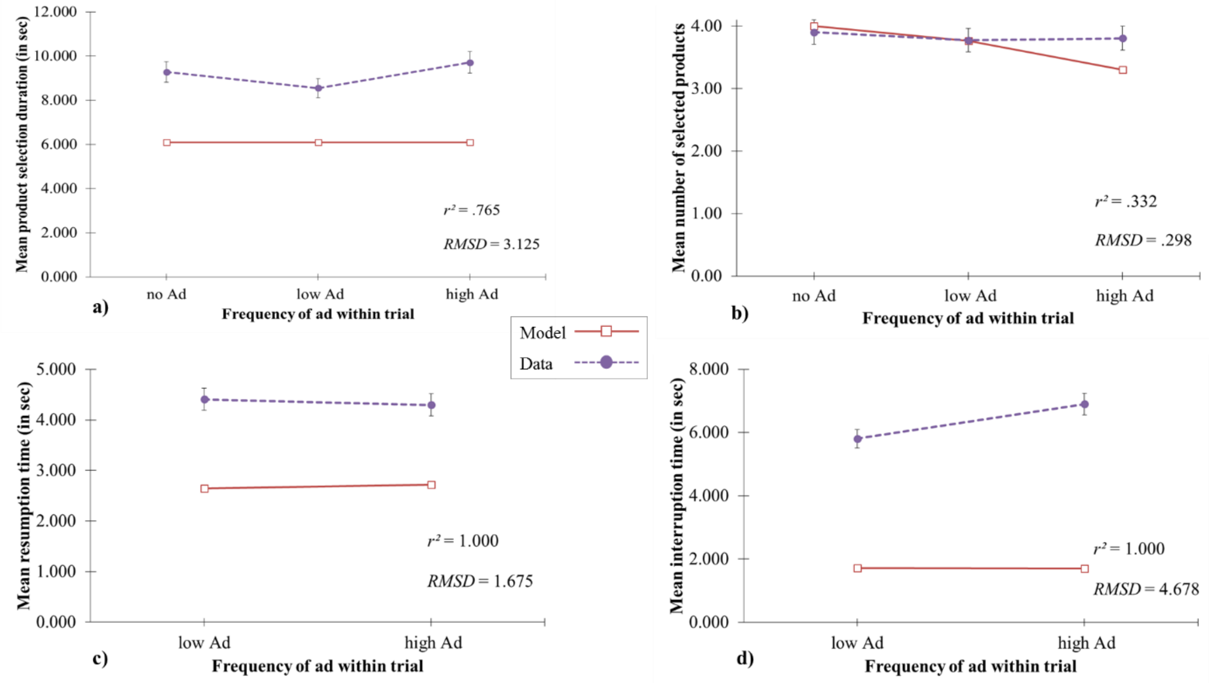

Figure 22. Comparison between model and human data on task performance under low workload. **a)** Product selection time, **b)** Selected products, **c)** Resumption time, **d)** Interruption time. Error bars represent 95% confidence intervals on human data.

Taking a look at the applied resumption strategies, model as well as human participants prefer the head-based over the world-based strategy in the low workload condition.

5 Discussion

The current thesis inspects cognitive processes related to interruption due to product advertisements within a shopping task. Taking into account situational characteristics, the effect of an enhanced mental workload induced by additional demands on working memory depicts a further aspect to be considered. On methodical accounts, the applied approach employs the development of a computational model within the cognitive architecture ACT-R as well as the introduction of a related experimental setting. A decay in task performance is expected by means of interruption (*first hypothesis*), workload (*second hypothesis*), and particularly the combined appearance of both aspects (*third hypothesis*). In this context, task performance is assessed via product selection time, the amount of selected products, resumption time, interruption time, and the amount of finally recalled products, all regarded as dependent variables. Additionally, two potential strategies used within the resumption process are inspected closer, a memory-based strategy ("head-based strategy") and a strategy mediated by information from the environment ("world-based strategy"). Regarding their application within the given task, a distinction in preference due to the level of workload is assumed (*fourth hypothesis*).

Looking at the results, obviously there are substantial differences between model and human performance. On one hand, the model data support the postulated decay in performance with increasing interruption, especially for selection and resumption times. In the case of enhanced workload, the same is true for all measures but interruption time. An interaction of the factors shows up particularly for the amount of selected products and resumption time. On the other hand, apart from the amount of finally recalled products, experimental data do not support the hypotheses for interruption, workload or their interaction. Moreover, the assumed difference in resumption strategy application does not show up either. Comparing model and human data by visual as well as numerical means, results are rather mixed. A good fit is indicated especially for the final recall performance. The separate inspection of the high and low workload variation of the task points towards comparable results in the former case especially for selection and resumption time. In contrast, a good fit within the low workload condition is indicated in particular in terms of resumption strategies.

5.1 Interpretation

Approaching potential explanations for the substantial deviation between model and human performance, there might be a variety of approaches. Those discussed subsequently depict just a selection of the most obvious ones in terms of the initially outlined theoretical background.

5.1.1 Interruption

First of all, one might consider a lack of disruptiveness of the chosen interruption responsible for failing to achieve the desired effect. On one hand, more than 80% of the tested participants reported to be familiar with smartphone use, and on this account may deal with advertisements on a regularly base. A common strategy when being confronted with such an interruption comprises ignoring it by closing the message as soon as possible. Within the structured interview, more than 90% of the participants stated that they usually behave like this. On this account, a median split was performed, separating those people not actually involved in the interruption, indicated by either stating explicitly that the product advertisement has not been experienced as disturbing or that the task has not affected their behavioral strategy shown in daily life. The resulting group of 34 people was compared with those involved in the interruption, but results did not indicate systematic differences. However, a median split depicts a rather criticized method, since it just accounts for differences between a high and low value of a certain parameter, but neglects the potential influence of a medium level (Rey, 2009). Therefore it might not depict the adequate method to inspect the data, but was nevertheless used due to the exploratory nature of the conducted analysis and existing time constraints. In future research, the application of more complex procedures like moderated regression analyzes might contribute to enlightening the existing relationships.

Further reasons for failing to interrupt could consist of training effects, i.e. people gain expertise in how to apply effective resumption strategies when they are frequently exposed to interruptions in mobile settings. This assumption directly corresponds to results reported by Trafon and Monk (2008) on task-related training with or without additionally practicing resuming after being interrupted. Obviously, the more experience people gained with the resumption process itself, the less negative effects on task performance due to the induced interruption arose. Above all, interruption might have occurred at stages with lower cognitive involvement (Adamczyk & Bailey, 2004), i.e. between already finished subtasks, additionally reducing its impairing effect.

Beyond that, another reason may consist of the short duration of the used interruption. In this vein, it might have not been able to prevent people from rehearsing the content of the main task during its appearance. Alternatively, as participants conducted the task predominantly at their own pace, some might have performed a short cognitive break of a few seconds to create a mental cue before actually reading the advertisement (McFarlane & Latorella, 2002). Therefore, such a rehearsal immediately before or during the interruption should be part of the model as well, to improve its fit to the human data. Moreover, the advertised product was related to the respective shop and, based on this fact, could have served as a resumption cue itself. Although this was explicitly reported by just about 11% of the participants, this strategy has possibly been used implicitly by a broader number of participants.

Finally, as people were mainly recruited within a database, there was a high variance in the resulting sample. On this account, a bunch of individual differences might have exerted influence on how people dealt with the interruption, just to name a few: arousal, anxiety, coordination ability, motivation, or the ability to deal with the stress that interruptions are prone to induce.

5.1.2 Mental workload

Regarding the induction of mental workload, the reasons for its unsuccessful induction may be twofold. On one hand, individual differences could be of relevance in this case as well, in particular those referring to the amount of available cognitive resources. According to Gopher and Donchin (1986), the main problem when inspecting the effects of workload by primary task performance consists of the fact that it usually does not reflect the amount of allocated resources, and therefore might remain unchanged. Wickens et al. (2013) describe different kinds of resource allocation facing enhanced workload, namely allowing the task performance to degrade, performing it more efficiently, optimizing task performance by focusing on the most important aspects, or dis-optimizing it by focusing on those of lower priority instead. However, those are difficult to inspect by just looking at the primary task performance.

Additionally, as already described in *section 3.3*, the CSPAN task was used as a control for the existing working memory capacity. Taking into account the memory demands arising from the shopping task, maybe the CSPAN has been too different from those. At least some participants explicitly stated this after completing the experimental session, which might be the reason for the lack of influence of the CSPAN score on the performance parameters. As reported above, it was not included as covariate due to this fact. Nevertheless, a median split on this variable was performed as well, separately for both workload conditions, in order to further enlighten potential relationships. Indeed, the obtained results indicate differences, at least on a descriptive level, for product selection time and resumption time. Obviously, a high CSPAN score seems to “protect” task performance by providing additional resources under conditions of high workload. This assumption is supported by the fact that interruptions were inspected longer in terms of higher working memory capacity.

On the other hand, how close participants stuck to the instruction and how much attention they paid to the person the product should be bought for depicts a second explanation for the obtained results. In order to shed light on this aspect, an index from the final recall performance was computed for participants in the high workload condition. In detail, the sum of correctly recalled products with the respective person was subtracted from that without considering the person-related information. Again, a median split was performed, and the descriptive values indicated differences between people highly engaged in remembering the additional information and those not doing so. They appeared in particular for product selection time, interruption time, and resumption time, with people less involved in the instruction achieving better results.

5.2 Implications

Based on the conceptual, methodological, and result-related issues discussed in previous sections, implications for research and development work can be derived, both in terms of theory and practice.

5.2.1 Theoretical implications

On theoretical accounts, when studying cognitive processes in the face of interruption and resumption, those have to be rather obvious to allow a deduction of potentially influencing aspects. In order to achieve stronger effects, interruptions should be less "cognitive avoidable" but demanding instead. Among other ways, this is achievable by content as well as timing, i.e. decrease its relations to the main task and foster an occurrence at inopportune moments in cognitive processing (Adamczyk & Bailey, 2004). Moreover, the choice of an interrupting task less familiar to the participants could strengthen the effects of an interruption as well, but on the other hand it bears the risk of provoking a more artificial setting. In the given context, a more demanding task and at the same time less ignorable interruption might have occurred by including prices for each product and advertised offer, accompanied by instructing the participants not to exceed a limited shopping budget but nevertheless buy some additional products. This idea would offer various extensions in terms of workload as well, e.g., prices requiring more complex arithmetic operations, or separate budgets for different protagonists.

Additionally, the focus on working memory might be extended above its capacity to entail aspects like the strength of connection between related constructs, or the rapidness of access to certain information. Within the model, such features could be included by adjusting parameters such as spreading activation in the former, and retrieval latency in the latter case. Spreading activation is determined by adding a source activation weighting parameter W_{kj} and a strength of association parameter S_{ji} to the activation equation, depicting the influence of the relationship of the respective chunk to slots of other chunks in the declarative memory. In contrast, retrieval latency, mentioned previously in *section 3.2.4*, determines how fast a chunk can be retrieved from declarative memory, therefore influencing its accessibility.

5.2.2 Practical implications

From a practical perspective, results and ideas might be of benefit for designers, engineers, and other industrial professionals, for they can be linked to the development of interfaces that are able to support users in successfully dealing with interruptions. As already outlined, the ability to rehearse memory content and the availability of adequate environmental cues depict critical issues in this context. The opportunity to rehearse may be fostered by inducing an alert, even in cases like the outlined interruption by product advertisement. Such warning periods can be very brief, e.g., the screen freezes and turn slightly grey before the actual advertising message is loaded, or a pop-up announcing a special offer related to the selected product with the question “Do you want to read it?” appears. This might enable a user to explicitly create cues or perform rehearsal. Especially in tasks entailing high memory load, this kind of increased user control could enhance the predictability of interruptions.

Beyond that, environmental cues linked to the last action should be rather blatant (Trafton et al., 2005) and uniquely linked to the last action, to avoid conflicting knowledge in the world. Given these prerequisites, such *“external mnemonic[s]”* (Trafton & Monk, 2008, p. 121) might be appropriate to show the assumed off-loading effect in cases of increased workload. In particular, those extending the implicit memory but directly relate to the current target instead have proven of value (Nelson & Goodmon, 2003). In addition, preserving situational awareness for the initial task might be helpful as well (McFarlane & Latorella, 2002).

In the case of advertisements, interruption usually depicts a desired effect, but nevertheless the existing trade-off between receiving attention and causing annoyance has to be maintained. Within the structured interview, participants sometimes reported feeling more attracted to the product offer since it was related to the respective product category. For this reason, a successfully inspected advertisement might be that linked to the content of the actual task. Moreover, besides an “agreeable” frequency of occurrence, it consists of short and simple messages to avoid long encoding times and user irritation attaching too many cognitive capacities. In optimal cases, an advertisement appears at opportune moments (Adamczyk & Bailey, 2004), within short interaction chains that lead to a goal or sub-goal (Trafton & Monk, 2008).

5.3 Limitations

The previously outlined topic depicts a promising field of work, but since a thesis like this always has to deal with limited resources, it faces certain boundaries. Some of the most important issues that might be considered in in future work are discussed in the following.

5.3.1 Model complexity

At first, when developing the ACT-R model, its complexity had to be restricted in certain ways. As already explained, it mainly focused on potential strategies applied with the purpose of resuming the main task, whereas processes while dealing with the interruption appeared in a rather simplified manner. In detail, the inspection of the advertising message occurred word by word, followed by making the decision to accept or refuse the offer. The latter was simulated by including a waiting period, however, there is a variety of ACT-R related research on more or less complex strategies and heuristics used for decision-making. Additionally, the model did not take into account individual differences in working memory, an aspect previously discussed in terms of workload. In ACT-R models, it can be included by means of the amount of source activation, reflecting *"a limitation on the amount of attention one can distribute over source objects"* (Anderson, Reder, & Lebiere, 1996, p. 225). It can be determined through adjusting the activation-related parameter W_j. Including such issues when further using the model opens a widespread field of research, regarded as highly valuable within the given context.

5.3.2 Sample size

Another core limitation relates to the inspected human sample, both in size and characteristics. Regarding sample size, according to Bortz and Döring (2009), the application of a 2x3-factorial design that inspects main effects as well as potential interactions with a power of $1\text{-}\beta = 0.80$, and a level of significance of $\alpha = 0.05$, demands 27 participants within each of the resulting six cells to reveal a medium-sized effect, leading up to an overall sample size of 162 people. Taking into account that interruption operates on repeated measures, the required amount decreases considerably, but still consists of 14 participants per cell, in total 84 people. Nevertheless, the latter amount would not have the strength to detect a workload main effect of medium size, since the mandatory sample size amounts to 128 participants for both conditions in this case. Inspecting the recommended quantities for high effects, although the current sample of 62 participants is adequate to detect a main effect of interruption, demanding 36 people, it slightly fails to reach that number required to cover a main effect for workload, demanding a sum of 66 people. In consequence, if a high to very high effect had existed, it might have been detected within the conducted experiment, but the non-significant results point towards an effect size of an at most medium level or even smaller. Having a look at the characteristics of the sample, using a database for recruitment might have increased its overall variance above that of ordinary student samples, maybe resulting in a lack of comparability to other studies within similar contexts. On this account, testing a bigger, more homogeneous sample could depict an opportunity to further inspect the assumed relationships.

5.3.3 Experimental setting

Within the human experimental setting, previous knowledge was generated by just performing one additional product selection run. However, this might have been not adequate to achieve a level of expertise comparable to that of the model. This assumption receives support from the fact that nearly one third of the participants reported difficulties in relating products to their respective categories. In consequence, a high variance on the inspected performance parameters occurred, as well as a certain amount of outliers distributed within all conditions of the experimental factors, i.e. people showing extremely long navigation times and in this vein extended search processes.

The difficulty concerning resumption strategies may exist due to the retrospective assessment with the structured interview and a rather limited set of questions. Those could have been too plain, vague or confounding for the participants, since at least in some cases people needed a more comprehensive explanation of the subject or totally failed to get an idea of the answer. Additionally, such self-reported information is rather prone to be biased or even inaccessible for the respective person. On this account, a more objective inspection of the used strategies via psychophysiological measures would be the method of choice to enlighten this matter. Indeed, the recording of gaze movements and pupil dilation happened in the reported experimental setting, but those data have not been analyzed and included within the thesis due to capacity and time constraints.

5.4 Prospect

As obvious by means of the previously discussed matters, there still are a lot of unanswered questions and toeholds for future research. In the following, some of the most apparent issues should be outlined in brief.

5.4.1 Extending the model

When trying to better represent the observed human behavior within the model, a first opportunity might consist of modeling novice users instead of experts. Thus, the model would have to learn the respective shop related to a product while performing the initial set of runs. In terms of modeling, this aspect could be included by giving a reward when gaining knowledge about the shop linked to a product. This growing relationship enables a faster retrieval of the correct information, and in this vein decreases selection times and errors. Another extension may comprise of shedding light on the second type of error as well, i.e. confusing a product with a similar one, commonly referred to as errors of commission. Within the modeling framework this aspect can be included by applying partial matching, defined by further extending the activation equation, already containing spreading activation (see *section 5.2.1*), with an additional component. It comprises the match scale parameter P, determining the strength of influence of the similarity values on the activation of the chunks, and the match similarities parameter M_{li}, referring to the actual similarity values between the chunks. The latter aspect defined by the modeler at the outset as well as the extent of instantaneous noise (*section 2.3.2*), enable the retrieval of incorrect chunks in the case the activation of the correct chunks resides beneath the retrieval threshold. Beyond that, more complex methods to adjust the used subsymbolic parameters might be applied to improve the model performance respective to the human data. Finally, when using a smartphone, certain device-related motor movements like scrolling or swiping occur. The ordinary ACT-R framework does not comprise such processes, but is limited to mouse movement and key press instead. However, Greene and Tamborello (2013) outlined an ACT-R extension for modeling the use of modern touchscreen devices, briefly called ACT-Touch, that may be embedded in future work as well to get closer to the actual smartphone use scenario.

5.4.2 Extending the focus

Another, more conceptual scope of extension might consist of applying the discussed theoretical and methodological framework to other kinds of interruption within the mobile context. At the beginning, further interrupting events linked to this setting were already mentioned, e.g., receiving an update or facing a system crash. While the first one is characterized by potentially arising learning requirements due to new features, in the latter case usually a loss of cues linked to the last state of action occurs. This might result in resumption procedures demanding more time and capacity resources. In contrast, interruptions by external events like motion or road traffic involve dealing with additional, and completely different situational constraints that may increase negative interruption effects. Besides interruptions without an alert, there are those which are announced before actually occurring, enabling the user to handle them in a more self-determined manner. Prominent examples could be receiving a phone call while working on a task with an application, or intentionally using more than one application at the same time. As explained by McFarlane and Latorella (2002), such negotiated interruptions offer different options to deal with them, ranging from immediate handling to completely ignoring. Thus having in mind the research conducted by Adamczyk and Bailey (2004), a core characteristic of those interruptions consists of the possibility of their delayed inspection at more opportune moments of cognitive involvement. Apart from its use in the mobile sector, the approach depicted in *section 2.1.2* can be applied in other settings as well. For instance Trafton et al. (2012) conducted research based on the outlined theoretical background in the field of human-robot interaction. In brief, a model was developed and implemented in an embodied robot to support a human storyteller in continuing after being interrupted. This occurred by giving reminding cues to the last event before the interruption. Overall, the authors contribute to the arising connection of predictions on human cognitive processing and their implementation in artificial intelligence platforms.

6 Conclusion

Based on the previously discussed issues, when looking at either model behavior or experimental results, one could ask which bears greater responsibility for the observed pattern. Did the model fail to adequately picture task-related human cognitive processing? Or did the experimental setting fail to evoke the desired effects? To shed light on these questions, it may be wise to remember the initial reasons for choosing the selected approach. As mentioned at the outset, the developed ACT-R model was established to support the comprehension of the user's cognitive processes while dealing with the inspected task. Although the collected experimental data face the discussed limitations, they give useful hints for improving the model in order to get closer to actual human cognition. Therefore, the attempt at a satisfying answer to the stated questions in the first instance leads back towards reflecting model development to explain the obtained results.

To finally draw conclusions, being interrupted while performing a certain task depicts a natural and sometimes intended part in our technologized society. Nevertheless, its negative effects can be mitigated by using effective strategies, even when the cognitive system already has to cope with enhanced demands. Therefore, an important goal for developers consists of designing interfaces that are able to support the successful application of such strategies. Referring to the question posed at the title outset, asking *"Smart@load?"*, the conclusion derived from this thesis should thus be as following: *"You can act smart under load – provided you know the right strategy!"*

References

Adamczyk P. D., & Bailey B. P. (2004). If not now, when?: The effects of interruption at different moments within task execution. In *Proceedings of the Conference on Human Factors in Computing Systems, CHI'04* (pp. 271-278), New York: ACM Press.

Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: an activation-based model, *Cognitive Science, 26*, 39-83.

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York: Oxford University Press.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval, *Cognitive Psychology, 30*, 221-256.

Battiste, V., & Bortolussi, M. (1988). Transport pilot workload - A comparison between two subjective techniques. In *Proceedings of the Human Factors and Ergonomics Society 32nd Annual Meeting* (pp. 150-154). Santa Monica, CA: Human Factors & Ergonomics Society.

Borst, J. P., & Anderson, J. R. (in press). Using the ACT-R Cognitive Architecture in combination with fMRI data. In B. U. Forstmann, & E. - J. Wagenmakers (Eds.), *An Introduction to Model-Based Cognitive Neuroscience*. Springer: New York.

Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler* [Statistics for human and social scientists] (6th ed.). Berlin: Springer.

Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* [Research methods and evaluation for human and social scientists] (4th ed.). Berlin: Springer.

Brixey, J. J., Robinson, D. J., Johnson, C. W., Johnson, T. R., Turley, J. P., & Zhang, J. (2007). A concept analysis of the phenomenon interruption. *Advances in Nursing Science, 30*(1), E26-E42.

Beier, G. (1999). Kontrollüberzeugungen im Umgang mit Technik [Locus of control in handling technology]. *Report Psychologie, 9*, 684-693.

Cades, D. M., Boehm-Davis, D. A., Trafton, J. G., & Monk, C. A. (2007). Does the difficulty of an interruption affect our ability to resume? In *Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting* (pp. 234-238). Baltimore, Maryland.

Cao, A., Chintamani, K. K., Pandya, A. K., & Ellis, R. D. (2009). NASA TLX: Software for assessing subjective mental workload, *Behavior Research Methods, 41*, 113-117.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide, *Psychonomic Bulletin & Review, 12,* 769-786.

Diekmann, A. (2007). *Empirische Sozialforschung* [Empirical social research] (18th ed.), Reinbek: Rowohlt Verlag.

Ecker, U. K. H., Lewandowsky, S., Oberauer, K., & Chee, A. E. H. (2010). The components of working memory updating: An experimental decomposition and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 170-189.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General, 128*, 309-331.

Gillie, T., & Broadbent, D. E. (1989). What makes interruptions disruptive? A study of length, similarity, and complexity, *Psychological Research, 50*, 243-406.

Goldsmith, R. E., & Hofacker, C. F. (1991). Measuring consumer innovativeness. *Journal of the Academy of Marketing Science, 19*, 209-221.

Gopher, D., & Donchin, E. (1986). Workload – An examination of the concept. In K. R. Boff, L. Kaufmann, & J. P. Thomas (Eds.). *Handbook of Perception and Performance. Vol. II. Cognitive Processes and Performance* (pp. 41/1-41/49). New York: Wiley.

Gray, W. D., Young, R. M., & Kirschenbaum, S. S. (1997). Introduction to this special issue on cognitive architectures and human-computer interaction. *Human-Computer Interaction, 12*, 301-309.

Greene, K. K., & Tamborello, F. P. (2013). Initial ACT-R Extensions for User Modeling in the Mobile Touchscreen Domain, In Robert L. West & Terrence C. Stewart (Eds.), *Proceedings of ICCM 2013, 12th International Conference on Cognitive Modelling* (p. 348-353). Ottawa, Canada.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock, & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-183). Amsterdam, New York, Oxford: North Holland Press.

Hollnagel, E. (1998). *Cognitive reliability and error analysis method (CREAM)*. Elsevier.

Iqbal, S. T., & Horvitz, E. (2007). Disruption and recovery of computing tasks: field study, analysis, and directions. In *Proceedings of the Conference on Human Factors in Computing Systems CHI 2007* (pp. 677-686). San Jose, CA, USA: ACM Press.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189.

Karrer, K., Glaser, C., Clemens, C., & Bruder, C. (2009). Technikaffinität erfassen – der Fragebogen TA-EG [Measuring affinity for technology – the TA-EG questionnaire]. In A. Lichtenstein, C. Stößel, & C. Clemens (Eds.), *Der Mensch im Mittelpunkt technischer Systeme. 8. Berliner Werkstatt Mensch-Maschine-Systeme* (ZMMS Spektrum, Reihe 22, Nr. 29, pp. 196-201). Düsseldorf: VDI Verlag GmbH.

McFarlane, D. C., & Latorella, K. A. (2002). The scope and importance of human interruption in human-computer interaction design. *Human-Computer Interaction (17)*, 1-61.

Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Nelson, D., L., & Goodmon, L. B. (2003). Disrupting attention: The need for retrieval cues in working memory theories. *Memory & Cognition, 31*(1), 65-76.

Norman, D. A. (1988). *The psychology of everyday things*. New York: Basic Books.

O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufmann, & J. P. Thomas (Eds.). *Handbook of Perception and Performance. Vol. II. Cognitive Processes and Performance* (pp. 42/1-42/49). New York: Wiley.

Reid, G. B., & Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P.A. Hancock, & N. Meshkati (Eds.), *Human Mental Workload* (pp. 185–218). Amsterdam, New York, Oxford: North Holland Press.

Rey, G. D. (2009). E-Learning. Theorien, Gestaltungsempfehlungen und Forschung [E-learning. Theories, design recommendations and research]. Bern: Verlag Hans Huber, Hogrefe AG.

Rußwinkel, N., & Prezenski, S. (2014). ACT-R Meets Usability. In *Proceedings of the Sixth International Conference on Advanced Cognitive Technologies and Applications*. *COGNITIVE 2014,* Venice, Italy.

Salvucci, D. D., & Taatgen, N. A. (2010). *The multitasking mind.* New York: Oxford University Press.

Salvucci, D. D., Taatgen, N. A., & Borst, J. P. (2009). Toward a unified theory of the multitasking continuum: from concurrent performance to task switching, interruption, and resumption. In *Proceedings of the Conference on Human Factors in Computing Systems, CHI'04* (pp. 1819-1828), New York: ACM Press.

Schunn, C. D., & Wallach, D. (2005). Evaluating goodness-of-fit in comparison of models to data. In W. Tack (Ed.), *Psychologie der Kognition: Reden und Vorträge anlässlich der Emeritierung von Werner Tack* (pp. 115-154). Saarbrücken: University of Saarland Press.

Statista (2014a). *Anzahl der Smartphone-Nutzer in Deutschland in den Jahren 2009 bis 2014 (in Millionen)* [Amount of smartphone users in Germany from 2009 to 2014 in millions]. Retrieved from http://de.statista.com/statistik/daten/studie/198959/umfrage/anzahl-der-smartphonenutzer-in-deutschland-seit-2010/ at October 10th, 2014.

Statista (2014b). *Umsätze mit mobiler Display-Werbung in Deutschland in den Jahren 2012 bis 2013 und Prognose für 2014 (in Millionen)* [Sales in mobile display advertisement in Germany from 2012 to 2013 with forecast for 2014 in millions]. Retrieved from http://de.statista.com/statistik/daten/studie/296157/umfrage/entwicklung-der-umsaetze-fuer-mobile-display-werbung-in-deutschland/ at October 11th, 2014.

Städtler, T. (Ed.) (2003). *Lexikon der Psychologie: Wörterbuch, Handbuch, Studienbuch* (Sonderausgabe) [Lexicon of psychology: dictionary, compedium, transcript (Special edition)]. Stuttgart: Alfred Kröner.

Trafton, J. G., Altmann, E. M., & Brock, D. P. (2005). Huh, what was I doing? How people use environmental cues after an interruption. In *Proceedings of the Human Factors and Ergonomics Society 49th annual meeting* (pp. 468-472). Orlando, Florida.

Trafton, J. G., Altmann, E. M., Brock, D. P., & Mintz, F. E. (2003). Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies, 58,* 583-603.

Trafton, J. G., Altmann, E. M., & Ratwani, R. M. (2011). A memory for goals model of sequence errors. *Cognitive Systems Research, 12*(2), 134-143.

Trafton, J. G., Jacobs, A., & Harrison, A. M. (2012). Building and verifying a predictive model of interruption resumption. *Proceedings of the IEEE, 100*(3), 648-659.

Trafton J. G., & Monk C. M. (2008). Task interruptions. In D. A. Boehm Davis (Ed.), *Reviews of Human Factors and Ergonomics, Vol. 3* (pp. 111-126), Santa Monica: Human Factors and Ergonomics Society.

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2013). *Engineering psychology and human performance* (4th ed.). Upper Saddle River, New Jersey: Pearson Education.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*, 625-636.

Yi, Y. (1990). Cognitive and affective priming effects of the context for print advertisements. *Journal of Advertising, 19*(2), 40-48.

Appendix

List of figures within the appendix

Figure A1. Introduction to the test scenario. ………………………………………………………... 73

Figure A2. Description of the desired shopping behavior. ……………………………………………….. 73

Figure A3. Introduction of the neighbor Norbert. ……………………………………………………... 73

Figure A4. Introduction of the ill friend Fiona. ……………………………………………………….... 73

Figure A5. Description of the app use and outline of the task. ………………………………………... 73

Figure A6. Question on the comprehension of the task. ……………………………………………….. 73

Figure A7. Introduction to the first product selection run. ……………………………………………….. 74

Figure A8. Instruction to memorize the product list. ………………………………………………….... 74

Figure A9. Product list indicating the respective person. ………………………………………………… 74

Figure A10. Introduction to the second product selection run. ……………………………………….... 74

Figure A11. Introduction to the third product selection run. …………………………………………….... 74

Figure A12. Introduction to the fourth product selection run. …………………………………………. 74

Figure A13. Introduction to the final recall part. ………………………………………………………. 75

Figure A14. Instruction to recall all remembered products. ……………………………………………… 75

A Experiment material

As the experiment was conducted in German, all subsequent questionnaires and instruction materials are depicted in their original language. An English description of the respective content can be found in detail in *section 3.3.3*. For the instruction material, just those used for female participants within the high workload variation is shown to avoid redundancy. Instructions for male participants simply refer to "Dennis" instead of "Diana", but apart from that stay the same. In the case of low workload, passages introducing "Norbert" and "Fiona" and reporting their involvement in the main character's shopping day are left out. Slides which appear several times in the same or a similar way are just shown once. Moreover, only material specially developed within this thesis is included, whereas standardized tasks and questionnaires (CSPAN, TA-EG, NASA-TLX), as well as the source code of the used shopping list application are solely part of the digital appendix.

A1.1 Demographic questionnaire

Im Folgenden finden Sie nun einige Fragen zu Ihrer Person. Bitte kreuzen Sie jeweils die zutreffenden Antwortoptionen an oder antworten Sie durch eine numerische oder stichwortartige Angabe in den entsprechenden Feldern.

Angaben zur Person

1. Wie alt sind Sie?

2. Sind Sie männlich oder weiblich?

männlich ☐ weiblich ☐

3. Welchen Beruf üben Sie aus?

4. Nutzen Sie ein Smartphone?

Ja ☐ Nein ☐

5. Falls Sie Smartphonenutzer/in sind, wie intensiv nutzen Sie Ihr Smartphone?

Tägliche Nutzung in Stunden ☐

Wöchentliche Nutzung in Stunden ☐

Monatliche Nutzung in Stunden ☐

6. Wofür nutzen Sie ihr Smartphone insgesamt am häufigsten?

7. Benötigen Sie eine Sehhilfe?

Keine Sehhilfe ☐ Kontaktlinsen ☐ Brille ☐

A1.2 Instructions for the shopping task

Stellen Sie sich bitte folgendes Szenario vor:

Sie sind Diana Schiller, Architekturstudentin und 24 Jahre alt. Diana spielt neben dem Studium im Hochschulorchester Klarinette und kocht in ihrer Freizeit gerne mit ihren Freunden. Außerdem arbeitet sie ehrenamtlich für ein Nachbarschaftshilfeprojekt.

Figure A1. Introduction to the test scenario.

Zum Einkaufen nutzt Diana das Einkaufszentrum in der Nähe der Universität, das über eine eigene Einkaufs-App verfügt. Mit deren Hilfe kann sie die benötigten Produkte in den jeweiligen Läden auswählen und damit ihren Einkauf planen. Darüber hinaus informiert die App sie auch über aktuelle Sonderangebote. Von diesen kauft die sehr preisbewusste Diana durchschnittlich die Hälfte.

Versuchen Sie nun bitte, sich Diana möglichst genau vorzustellen!

Figure A2. Description of the desired shopping behavior.

Im Rahmen des Nachbarschaftshilfeprojekts kauft Diana regelmäßig für ihren Nachbarn Norbert ein. Der 70-jährige Rentner war früher als Lehrer tätig, und ist aufgrund einer Gehbehinderung nicht mehr so gut zu Fuß. Mit seinem Hund Mischka kann er zwar noch kleinere Spaziergänge machen, aber der Weg ins Einkaufszentrum ist für ihn inzwischen zu weit geworden.

Versuchen Sie nun bitte, sich auch Norbert so genau wie möglich vorzustellen!

Figure A3. Introduction of the neighbor Norbert.

Fiona, 26 Jahre, studiert ebenfalls Architektur und ist gut mit Diana befreundet. Die beiden haben schon so manche Klausur zusammen durchgestanden und kochen in ihrer Freizeit regelmäßig zusammen. Im Moment liegt Fiona allerdings mit einer schweren Grippe im Bett und kann ihre Einkäufe daher nicht selbst erledigen, so dass Diana das gerne für sie übernimmt.

Versuchen Sie nun bitte, sich auch Fiona so genau wie möglich vorzustellen!

Figure A4. Introduction of the ill friend Fiona.

Ihre Aufgabe besteht im Folgenden darin, für Diana, Norbert und Fiona einzukaufen und dafür die bereitgestellte Einkaufsapp zu nutzen.

Aus dem Startmenü heraus gelangen Sie durch die Auswahl von „Läden“ in die verfügbaren Geschäfte. In diesen sind die jeweiligen Produkte zu finden und lassen sich durch Setzen eines Hakens in das graue Kästchen auswählen. Nachdem Sie alle Produkte ausgewählt haben, lassen sie sich durch Drücken des Einkaufswagens oben links auf der Startseite zur Einkaufsliste hinzufügen.

Überlegen Sie bei Sonderangeboten bitte, ob Diana diese für sich oder eine der anderen beiden Personen kaufen würde.

Figure A5. Description of the app use and outline of the task.

Haben Sie verstanden, worin Ihre Aufgabe besteht? Falls ja, erhalten Sie von der Versuchsleiterin nun weitere Anweisungen.

Figure A6. Question on the comprehension of the task.

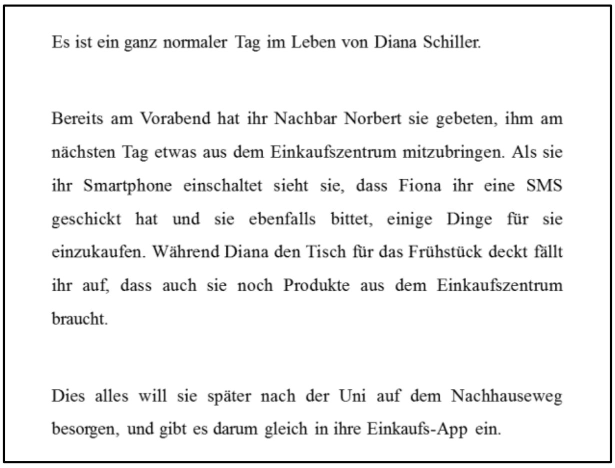

Figure A7. Introduction to the first product selection run. It serves to generate previous knowledge.

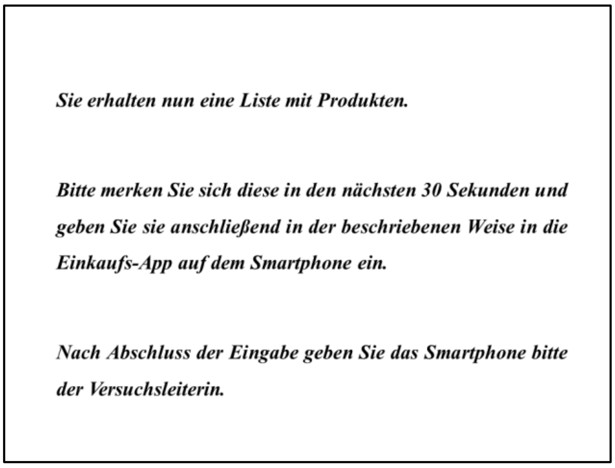

Figure A8. Instruction to memorize the product list. It appears four times in this way.

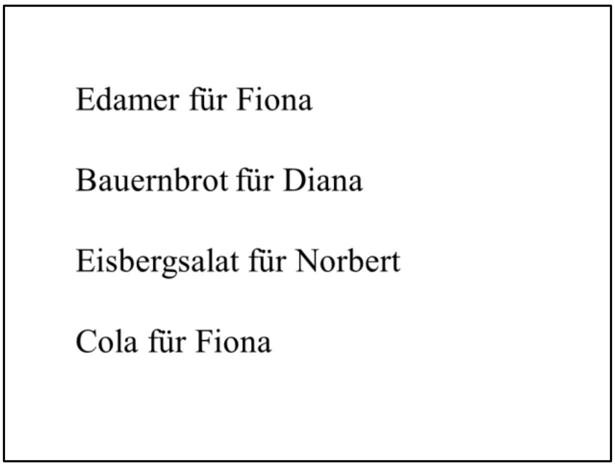

Figure A9. Product list indicating the respective person. It appears four times in this way but with various products.

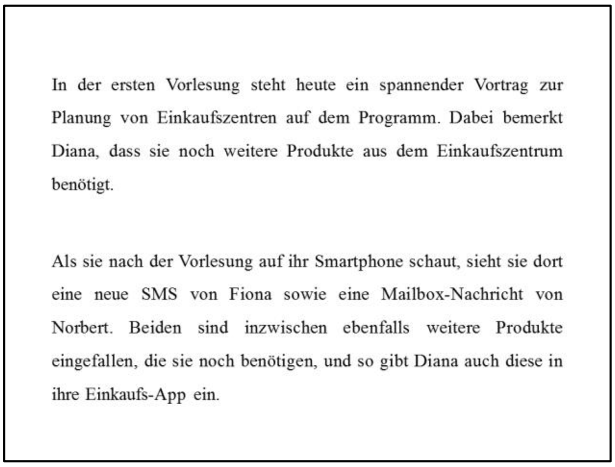

Figure A10. Introduction to the second product selection run. It depicts the first test run.

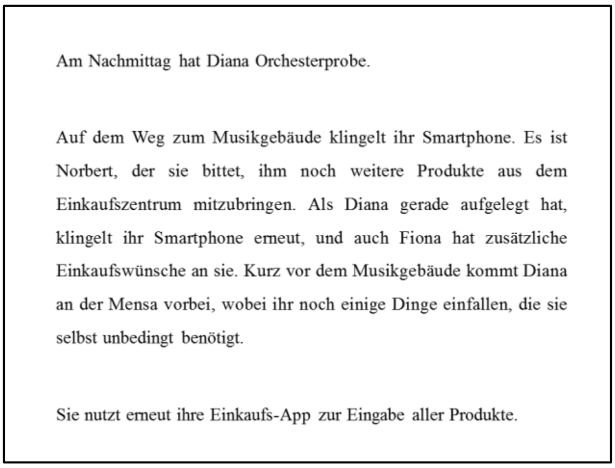

Figure A11. Introduction to the third product selection run. It depicts the second test run.

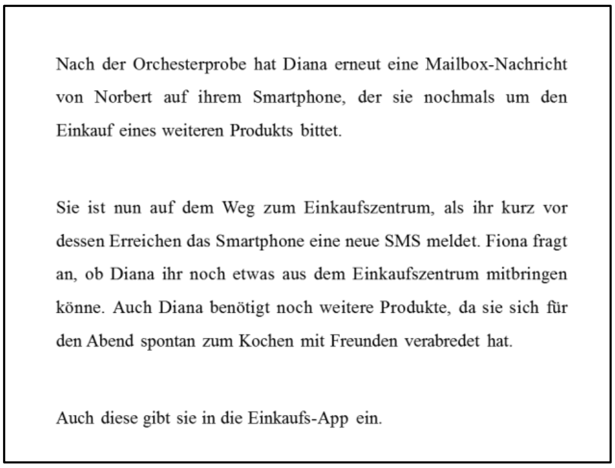

Figure A12. Introduction to the fourth product selection run. It depicts the third test run.

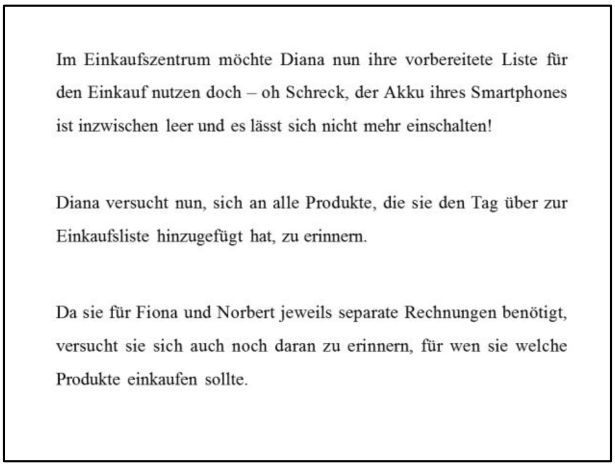

Figure A13. Introduction to the final recall part.

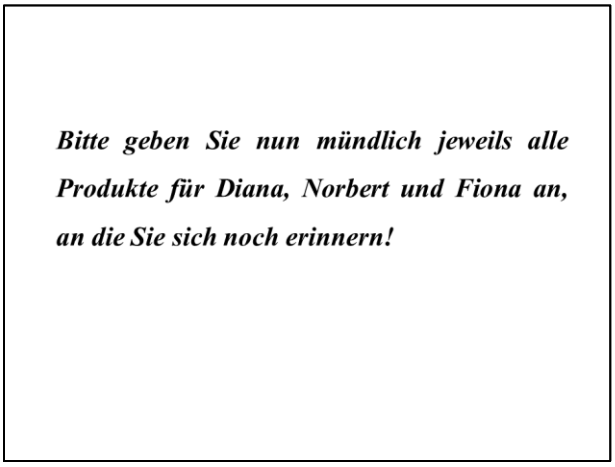

Figure A14. Instruction to recall all remembered products.

A1.2 Questions within the structured interview

Strukturiertes Interview zur Unterbrechung und den genutzten Wiederaufnahme-Strategien

1. Wie haben Sie sich Diana/Dennis vorgestellt?
2. In der High Workload Bedingung: Wie haben Sie sich Norbert vorgestellt?
3. In der High Workload Bedingung: Wie haben Sie sich Fiona vorgestellt?
4. Wie schwer ist Ihnen die Bearbeitung der Aufgabe am Smartphone gefallen?
5. Wie störend empfanden Sie die Werbeunterbrechungen?
6. Wie plausibel empfanden Sie die Werbeunterbrechung?
7. Wie gehen Sie für gewöhnlich mit Werbeunterbrechungen während der Smartphone-Nutzung um?
8. Sind Sie anders mit den Werbeunterbrechungen umgegangen, weil Sie nun in der Rolle einer anderen Person gehandelt haben?
9. Wie haben Sie es geschafft, nach der Werbeunterbrechung wieder an die Suche und Auswahl der Produkte anzuknüpfen?
10. Haben Sie versucht, sich daran zu erinnern, welches Produkt Sie unmittelbar vor der Werbeunterbrechung ausgewählt haben?
11. Haben Sie versucht, sich an alle bereits ausgewählten Produkte zu erinnern?
12. Haben Sie nach der Werbeunterbrechung auf dem Display nach dem Produkt gesucht, dass Sie unmittelbar davor ausgewählt haben?
13. In welcher Reihenfolge haben Sie sich die Produkte jeweils gemerkt?

B Digital appendix

B1 Model code

B1.1 High workload (folder containing LISP files for running the high workload model)

B1.2 Low workload (folder containing LISP files for running the low workload model)

B1.3 ACT-R 6 (folder containing the ACT-R 6 Standalone Version)

B2 Experiment material

B2.1 Consent form (.docx file containing a printable version)

B2.2 Demographic questionnaire (.docx file containing a printable version)

B2.3 CSPAN Task (folder containing .ebs and .bmp files as well as an answer sheet)

B2.4 TA-EG questionnaire (.pdf file containing a printable version)

B2.5 Instructions shopping task (folder containing .pptx files for female and male participants in both workload variations)

B2.6 Documentation final recall (.docx file containing printable versions for both workload variations)

B2.7 NASA-TLX questionnaire (.pdf file containing a printable version)

B2.8 Structured interview (.docx file containing a printable version for taking notes)

B2.9 Shopping list application (folder containing the Java source code to run the application)

B3 Data analysis

B3.1 Log files model (folder containing the log files generated during the model runs)

B3.2 Log files experiment (folder containing the log files generated by smartphone use)

B3.3 Data file model (.sav file containing the whole model dataset)

B3.4 Data file experiment (.sav file containing the whole experimental dataset)

B3.5 Analysis model data (.sps file containing the syntax for analyzing model data)

B3.6 Analysis experiment data (.sps file containing the syntax for analyzing experiment data)

B3.7 Data file model fit (.sav file containing the dataset for calculating the model fit)

B3.8 Analysis model fit (.sps and .xlsx files containing syntax and results for the model fit calculations)

B3.9 Figures (.xlsx file containing the set of result related figures)

Technische Universität Berlin

Prüfungsausschuss Human Factors
Institut für Psychologie und Arbeitswissenschaft
Franklinstr. 28/29 • Sekr. FR 3-8 • D-10587 Berlin

Fakultät V

IPA TU Berlin

Nachweis über 10 abgeleistete Versuchspersonen-Stunden als Bestandteil der Masterarbeit im Fach Human Factors

Name: Maria Würzberger Matr.-Nr.: 346577

Versuchstitel	Zeitdauer	Datum	Name des/der Versuchsleiters/ Versuchsleiterin (in Druckschrift)	Unterschrift des verantwortlichen Prof. oder WM
Blicksteuerung	1h	01.03. 2013	Juliane Bock	[signature]
Operateurstätigkeiten in Leitwarten	2h	10.06. 13	Zirk	[signature]
Reliabilität TUBSS (Müdigkeitsmessung)	6h	10.5. & 17.5.13	Michel Peters	[signature]
GESTEN	1h	05.07. 2013	RUFF	[signature]