Justus-Liebig-University Giessen

DISSERTATION

Adaptive user assistance in virtual reality shopping

Submitted in fulfillment of the requirements for the degree of

DOCTOR RERUM POLITICARUM (Dr. rer. pol.)

in the

Faculty of Economics and Business Studies

by

Tobias Weiß

Supervisors:

Prof. Dr. Jella Pfeiffer (first supervisor),

Chair for Information Systems,

Abt. VII: Wirtschaftsinformatik I, Stuttgart University

Prof. Dr. Nicolas Pröllochs (second supervisor),

Chair for Data Science and Digitization,

Faculty of Economics and Business Studies, Justus-Liebig-University Giessen

on

September 15, 2024

„God, grant me the serenity to accept the things I cannot change,

the courage to change the things I can,

and the wisdom to know the difference.“

Attributed to Reinhold Niebuhr, Lutheran theologian (1892–1971)

Acknowledgements

First, I would like to thank my supervisor, Prof. Dr. Jella Pfeiffer. Thank you for the freedom, the opportunities, the possibilities, your empathy, and your open-mindedness. I would also like to express my sincere gratitude to my second supervisor, Prof. Dr. Nicolas Pröllochs, who agreed to my request for supervision without hesitation.

Next, I would like to thank Dr. Christian Peukert and Prof. Dr. Thies Pfeiffer, who willingly shared their data and source code with me and thus got my work “kickstarted”. In addition, I am very grateful for the time of the amazing students, co-authors, and faculty members from various institutions who have accompanied me along the way.

Finally, I would like to thank Prof. Dr. Oliver Kirchkamp for his constant support and for serving as a role model for me, at least in some aspects of my life. Of course, I would also like to thank my mother for always supporting me and believing in me.

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – GRK2739/1 – Project Nr. 447089431 – Research Training Group: KD2School – Designing Adaptive Systems for Economic Decisions

Table of Contents

List of Abbreviations

AI	Artificial Intelligence
BG	Bayes Factor
AR	Augmented Reality
CAVE	CAVE Automatic Virtual Environment
CL	Cognitive Load
ECG	Electrocardiography
EEG	Electroencephalography
ELU	Exponential Linear Unit
ELPD	Expected Log Pointwise predictive Density
FB	Full Body
fNIRS	Functional Near-Infrared Spectroscopy
ET	Eye Tracking
FT	Face Tracking
HDI	Highest Density Interval
HUD	Heads-Up Display
HMD	Head-Mounted Display
LLM	Large Language Model
NUTS	No U-Turn Sampler
OC	Oculus Lip Sync
UAS	User Assistance System
GMM	Gaussian Mixture Model
HMM	Hidden Markov Model
ML	Machine Learning
SB	Static Body
SD	Standard Deviation
SF	Static Face
TP	Third Person
RB	Rocketbox
ReLU	Rectified Linear Unit
RM	Readyplayerme
VR	Virtual Reality
XR	Extended Reality

Publication Overview

This dissertation is comprised of the following publications and manuscripts:

Paper A Weiß, T., Pfeiffer, J. and Pfeiffer, T. (2024). "Early Bird – Predict healthy product choices in virtual commerce". ECIS 2024 Proceedings. 3.

Own contribution: 60% (Conceptualization, Methodology, Data analysis, Writing)

Paper B Weiß, T. and Pfeiffer, J. (2024). Consumer decisions in virtual commerce: Predict good help-timing based on cognitive load. Journal of Neuroscience, Psychology, and Economics, 17(2), 119.

Own contribution: 90% (Conceptualization, Experimental Design, Data collection, Data analysis, Methodology, Writing)

Paper C Weiß, T., Merkl, L., & Pfeiffer, J. (2023). Customer decision-making processes revisited: insights from an eye tracking and ECG study using a hidden Markov model. In NeuroIS Retreat (pp. 221-230). Cham: Springer Nature Switzerland.

Own contribution: 80% (Conceptualization, Methodology, Data analysis, Writing)

Paper D Weiß, T., Eilks, A., Pfeiffer, J., Putze, F., and Schultz, T. (2024). "Real agents in virtual commerce". Working paper.

Own contribution: 20% (All authors contributed equally)

Paper E Weiß, T., Pfeiffer, J., and Meißner, M. (2024). "Adaptive product comparison assistance in virtual reality". Working paper.

Own contribution: 60% (Data collection, Data analyses, Writing)

In addition, I have published the following article in the context of digital teaching and learning using VR. However, it is in German language and not included in this dissertation:

Weiß, T., Kirch, P., Büst, M., Schinder, S., and Pfeiffer, J. (2023). "Eine interdisziplinäre Kooperation in der Hochschullehre mit Hilfe der virtuellen Realität". Workshop-Proceedings der DELFI & HDI 2023.

1 General Introduction

I invite the reader to take a moment to think about what is possible and what determines it. As key concept in various disciplines, the possible describes the area of human experience that lies beyond the here and now. “The possible is not opposed either to the ‘actual’ or the ‘real’ and, in fact, our capacity to engage with what is possible grows out of concrete experiences and ends up transforming them” (Glăveanu, 2023, p. vii).

With the advent of Virtual Reality (VR), researchers have yet another tool at their disposal to push the boundaries of possibilities and make recently unimaginable experiences happen. Early VR pioneers, like Sutherland (1968) and Lanier (1989), were quick to embrace the technological opportunities, creating virtual environments already decades ago, with incredibly minimal hardware resources by today's standards. The miniaturization of transistors (Wu et al. 2007) and breakthroughs in organic light-emitting diodes (Kang et al. 2022) made it possible to produce today’s state-of-the-art VR devices, especially VR headsets. Soon, the idea of an interconnected world emerged, the Metaverse, that partially replaces the deteriorating real word, at least as a part-time habitat (Stevenson 1994). The author depicts a scenario in which the whole society adopts a parallel immersive virtual world into their everyday lives. With the rebranding of a large social media platform company to “Meta”, some enthusiasts were already heralding the dawn of this new era. However, current sales figures, customer sentiment, and technical developments point in a different direction. Inadequate network infrastructure, interoperability issues, and blockchain throughput limitations are just some of the major problems that let the Metaverse remain rather fiction than reality (Ball 2022). To sum up, the current state of VR technology is proving to be less disruptive than the smartphone, and barriers such as general technology aversion, discomfort in wearing, and a lack of VR applications are resulting in less adoption than some enthusiasts expected.

Nonetheless, in certain niche areas the current state of VR technology proves to be successful. For example, the learning and teaching domain shows promising use cases for VR (Renganayagalu et al. 2021). Consumer behavior and human computer interaction research also benefits from the latest VR technology (Stepanova et al. 2023). Eye tracking in VR and the recording of additional (bio-)sensors allows to adapt to the user and offers various research opportunities (Meißner et al. 2019).

Previous research has identified good timing as relevant factor for interactions between buyer and seller (Friemel et al. 2018; Lieven 2016). Our basic assumption is that most sellers will eventually offer virtual stores and showrooms that consumers can enter using a VR headset. Thus, we pose all research question of this dissertation in the realm of virtual commerce. We cover different aspects of consumer behavior and user interaction with the virtual environment, be it with a human-like agent or a modest interface element. A common denominator and thread that runs through the manuscripts is the pursuit of questions about adaptivity of these help providers, especially the timing of interference with the user.

Our articles document the boundaries of virtual commerce with the state of VR hardware and software limitations in the year 2025. At the same time, our artifacts and results shape the future virtual commerce landscape by providing guidance and applicable examples to practitioners and future generations of researchers.

The first presented paper, Paper A, is about a machine learning project that shows how InceptionTime, a deep learning time series classifier, can predict healthy product choices in a VR shopping environment. Our investigation is based on a large-scale VR data set of more than thousand product choices that was collected by Peukert et al. (2019). The goal is to predict healthy and unhealthy product choices using eye tracking data. Because the observations in the sample exhibit high class imbalance (mostly unhealthy product choices), we apply an evaluation metric that is geared towards the correct prediction of healthy choices (what introduces a flavor of nudging towards healthy products). We find superior performance of the deep time series classifier in comparison to a shallow gradient boosting baseline model. Overall, the results suggest that the presented method may be useful as feature generator for a gaze-based recommender system.

Paper B focuses on good interference timing of user assistance in VR and combines ideas from the educational and consumer behavior domain. We present an experimental design that covers two stages: in the first stage participants perform mentally demanding tasks; in the second stage they perform a purchase decision. We train a cognitive load classifier on the mentally demanding tasks of the first stage. The features consist of eye tracking and electrocardiography recordings that we synchronize and aggregate. Subsequently, the cognitive load classifier evaluates the purchase decisions based on the same features. Our results suggest that a good timing for algorithmic user assistance may be predicted based on cognitive load. However, the demand for help by an avatar seems to be affected by further influence factors, such as age and openness of the participant.

Paper C is a spin-off from the project presented in Paper B. We present an approach that utilizes a Hidden Markov Model with gaussian mixture distributions to discern decision-making sub phases. The Gaussian distributions represent different eye tracking and electrocardiography features, like fixations, saccades, and heart rate variability. The results suggest that sub phases of the decision-making processes and the transitions between the sub phases are detectable by means of the collected eye tracking and electrocardiography features.

Paper D is a lab linking project in cooperation with Bremen University. In a distributed setup, we mimic a customer interaction in the Metaverse and simulate sales conversations in a virtual commerce showroom. A human agent is steering an avatar either in third-person or with a full-body motion tracking suit, what entails different levels of fidelity. With a qualitative approach, we pursue the question how uncanny the agent is perceived and if we can improve the impression of the participants iteratively. Following the research questions of our previous study, we collect opinions about the right interference timing of the agent. We derive a simple appearance rule set to have actionable advice for the agent, based on the consumer gaze patterns.

As the final contribution of this dissertation, Paper E is a project that was also initialized by Christian Peukert, who created the initial experimental design. I took over his research by performing modifications to the questionnaire, experiment application, and by conducting the lab sessions with the help of my student assistants. The study investigates whether context-aware user assistance fosters trust, and if this relation is mediated by perceived intelligence of the system and perceived control over the system. Moreover, we investigate if explanations about the system's behavior alter these relations. In our case, context-awareness refers to whether the system is present from the very beginning or if it appears adaptively using eye tracking information. We report a Bayesian statistical analysis that provides evidence for the hypothesized mediation paths. In the analysis, we compare different variants of parallel mediations and an alternative moderated mediation approach using different prior distributions and control variables.

References

Ball, M. (2022). The metaverse: and how it will revolutionize everything. Liveright Publishing.

Friemel, C., Morana, S., Pfeiffer, J., & Maedche, A. (2018). On the role of users' cognitive-affective states for user assistance invocation. In *Information Systems and Neuroscience: Gmunden Retreat on NeuroIS 2017* (pp. 37-46). Springer International Publishing.

Glăveanu, V. P. (Ed.). (2023). The Palgrave encyclopedia of the possible. Springer Nature.

Kang, C. M., & Lee, H. (2022). Recent progress of organic light-emitting diode microdisplays for augmented reality/virtual reality applications. *Journal of Information Display*, *23*(1), 19-32.

Lanier, J., Conn, C., Minsky, M., Fisher, S., & Druin, A. (1989). Virtual environments and interactivity: Windows to the future. In ACM SIGGRAPH 89 Panel Proceedings (pp. 7-18).

Lieven, T. (2016). Customers' choice of a salesperson during the initial sales encounter. *Journal of Retailing and Consumer Services*, *32*, 109-116.

Meißner, M., Pfeiffer, J., Pfeiffer, T., & Oppewal, H. (2019). Combining virtual reality and mobile eye tracking to provide a naturalistic experimental environment for shopper research. Journal of Business Research, 100, 445-458.

Peukert, C., Pfeiffer, J., Meißner, M., Pfeiffer, T., & Weinhardt, C. (2019). Shopping in virtual reality stores: the influence of immersion on system adoption. *Journal of Management Information Systems*, *36*(3), 755-788.

Renganayagalu, S. K., Mallam, S. C., & Nazir, S. (2021). Effectiveness of VR head mounted displays in professional training: A systematic review. Technology, Knowledge and Learning, 1-43.

Stepanova, E. R., Desnoyers-Stewart, J., Kitson, A., Riecke, B. E., Antle, A. N., El Ali, A. & Howell, N. (2023). Designing with Biosignals: Challenges, Opportunities, and Future Directions for Integrating Physiological Signals in Human-Computer Interaction. In 2023 ACM Designing Interactive Systems Conference (pp. 101-103).

Stephenson, N. (1994). Snow crash. Penguin UK.

Sutherland, I. E. (1968, December). A head-mounted three dimensional display. In Proceedings of the December 9-11, 1968, fall joint computer conference, part I (pp. 757-764).

Wu, B., & Kumar, A. (2007). Extreme ultraviolet lithography: A review. *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena*, *25*(6), 1743-1761.

2 Paper A: Early Bird – Predict healthy product choices in virtual commerce

Tobias Weiß, Jella Pfeiffer, and Thies Pfeiffer

Abstract

Due to advances in extended reality technology, an increasing number of head-mounted displays are equipped with eye trackers. These sensors allow to predict customers' preferences on-the-fly. Such information can serve as features for recommender systems. We propose to treat eye tracking data as time series and utilize a deep time series classifier for inference. Our evaluation investigates possibly early predictions about customer preferences for healthy products in a virtual reality environment. The results, that are based on data from a large-scale laboratory experiment, demonstrate superior performance of the time series classifier, compared to a shallow gradient boosting baseline. They indicate a trade-off between prediction quality and how early this prediction is made. Overall, our study suggests that eye tracking and time series classification are valuable avenues for research and practice. Adaptive (shopping) assistants and recommendations based on artificial intelligence and bio sensors seem to be in close vicinity.

Keywords: Extended Reality, Eye Tracking, Healthy Consumption, Time Series Classification, Virtual Commerce

2.1 Introduction

Healthy food choices are a highly relevant topic for making predictions and recommendations in retail context (Cho et al. 2014; Naruetharadhol et al. 2023), as food choices are an important determinant of physical health and well-being (Wahl et al. 2017; Block et al. 2011; Bublitz et al. 2013). After the disruptive retail transformation from physical warehouses to e-commerce, a slower but continuous development towards virtual commerce is taking place (Evans and Wurster 1999; Bourlakis et al. 2009; Gadalla et al. 2013; Kovacova et al. 2022). Extended Reality (XR), an umbrella term for Augmented Reality (AR) and Virtual Reality (VR), found its way into Western society. Through interaction and high realism, this new technology offers unprecedented opportunities that may encourage consumers to make healthier choices. Research on the topic is needed that investigates new challenges and opportunities. Thus, we think that retailers should seize the opportunity and adjust their user assistance capabilities in order to meet the eminent needs of consumers who visit their future (at least partly virtual) commerce environments (Regt and Barnes 2019). Examples are adaptive head-up displays (HUDs that display customized product information and comparison options), personalized side-by-side recommendations, contextual advertising, and cross-platform nudges based on individual characteristics and preferences (Mariotti et al. 2023).

While acknowledging that research should advocate for rigid privacy measures within any XR environment, technological developments will most likely lead consumers to wear XR headsets, equipped with various bio sensors, for prolonged periods. Today, the first consumer-grade XR devices offer bio sensor based features, such as foveated rendering (Patney et al. 2016) and gaze-based interactions (Piumsomboon et al., 2017). One reason for our anticipated proliferation of biosensors is the privacy-personalization paradox, which describes the fact that people readily give personal information away if they expect utility while misjudging the real value of their personal information (Hoang et al. 2023).

Especially eye tracking (ET) based applications are a unique selling point in the current XR adoption phase. Eventually, ET could become a quality-of-life feature which consumers take for granted, like the camera in smartphones. ET can help to achieve a high degree of personalization and serve as an additional source of information for recommender systems (Meißner et al. 2019). In XR recommendation scenarios, ET may eventually replace click streams and historical data to a large extent. This is because ET allows close investigation of the user’s decision process and at the same time is available in the early phase of a purchase situation (Pfeiffer et al. 2020; Meißner et al. 2019).

Regarding the consumer preference (the dependent variable), we focus on healthy consumption because in different societies around the world, an increased attention on a healthy lifestyle is noticeable (Parashar et al. 2023). Policy makers are introducing healthiness indicators like the Nutri-score label and are actively fostering a healthy consumption (Hercberg et al. 2021), which is even included in the United Nations Sustainable Development Goals (Fernandez 2019). Therefore, a valuable customer insight is whether a person is open to suggestions that support healthy product choices or not (Tran et al. 2018). In the light of these developments, we pose following research question:

Can we identify customers who buy healthy products possibly early during their decision process in a virtual commerce scenario?

Shallow machine learning approaches have already been successfully applied in previous studies that predicted other aspects of the customer journey, for example the customers’ search motives (Pfeiffer et al. 2020) or the duration of intermediate decision stages (Weiß et al. 2023). A logic next step is to leverage deep learning to make predictions. An increasing amount of data and architectural improvements are likely to allow training of highly generalizing (or very precise, specialized) models. We treat the ET data as a discrete time series and, as further contrast to previously mentioned works, compare InceptionTime, one of the most promising deep learning approaches for time series classification, with the shallow gradient boosting method XGBoost which uses cross-sectional features.

With this paper, we contribute to the information systems literature in theoretical and practical manner. (i) As theoretical contribution, we show the superiority of using the complete time series of ET data in contrast to treating the ET data as cross-sectional data (by aggregating the number of fixations and other attributes). (ii) On the practical side, we show a promising way to personalize assistance systems in future metaverse applications based on the inobtrusive collection of ET data. Our paper describes a machine learning approach based on ET data which can be used to personalize XR experiences. The resulting features are of particular interest for new products or, more generally, in cases where user data is absent. (iii) Moreover, we investigate the trade-off between prediction quality and timing. Overall, our results inform the reader about interesting time windows during the decision process in our experimental purchase situation. From a broader research perspective, we show a promising way to personalize assistance systems in future metaverse applications.

2.2 Related Work

Already several Second Life studies pioneered connected 3D environments in virtual retail platforms (Bourlakis et al. 2009; Gadalla et al. 2013; Papagiannidis and Bourlakis 2010). The authors have depicted a transformation of traditional retail and outlined evolving marketing opportunities in the virtual space. Their conclusions emphasize the need for highly personalized and precisely timed customer service. Today, such connected virtual environments are thought of as the Metaverse, which are accessible via various XR devices. Recent comprehensive literature reviews about Metaverse shopping (Kliestik et al. 2022; Alcañiz et al. 2019; Shen et al. 2021) and AR shopping (Popescu et al. 2022) show how earlier claims, that were made for desktop environments, remain valid in XR. Virtual commerce research has diversified while recommendations and personalization remain highly relevant. A further recent review by Xi and Hamari (2021) categorizes 83 XR shopping studies along different axes (theories, in- and output devices, tracking technology, products, cognitive reactions, behavioral outcomes) and suggests a number of avenues for future research. Among these suggestions is an effective and efficient design of XR shopping, which is the area this work contributes to. The Metaverse is steadily taking shape (Peukert et al. 2022; Sriram 2022), head-mounted displays (HMDs) technology is advancing (Spagnolo et al. 2023), and HMD prices are deteriorating (Jensen and Konradsen 2018).

Various experiments have shown the significant impact of recommendations on the shopping behavior of customers, such as Li et al. (2022). Particularly in advertisement driven environments, recommender systems are very important business components. For instance, Google[1] accounts 40% of the Play Store app installations and 60% of the YouTube watch time to recommendations made by their recommender system. Collecting implicit information which reflects user preferences, like ET data, is an unobtrusive approach. This is important, as finding similarities between individuals should happen without any disruption of the consumer. Working with ET data in the context of recommender systems is nothing new (Castagnos et al. 2010; Xu et al. 2008; Zhao et al. 2016), but previous studies focused on

[1] https://developers.google.com/machine-learning/recommendation/overview

desktop based e-commerce websites. Moreover, these studies do not aim for an early prediction of user preferences.

Generally speaking, gaze patterns have potential to improve various aspects of digital and virtual commerce. Takahashi et al. (2022) presented a work in which they utilized ET to optimize a desktop-based 3D store layout. With the goal to support customers' decision-making processes, the experiment software used gaze information to rearrange the displayed products. Another step towards gaze-pattern utilization in shopping context was made by David-John et al. (2021). Their experimental design consisted of selection tasks of food items listed on recipes in a VR scene. The authors predicted the participants' intent to interact using logistic regression on gaze patterns. They treated the data as time series but only for a relatively short prediction horizon of 0.17 to 1 second. The results suggest that the used model can predict the users' interaction timing in real-time with above-chance accuracy.

Further ET studies have examined healthy food choices (Fenko et al. 2018; Kim et al. 2018) but the prediction horizon of these studies covered the whole decision-making process until the very end. Typical research using ET in the field of consumer behavior focuses on understanding and modelling the entire decision process up to the final purchase. For example, ET research has found the gaze cascade effect which describes a pre-decisional focus of attention on the chosen product (Shimojo et al. 2003; Krajbich and Rangel 2011). Regarding our research gap, none of these studies predicted customers' preferences early in the decision process.

In a hybrid field study, Pfeiffer et al. (2020) investigated grocery shopping behavior, especially the differences between a real and virtual supermarket. The authors did not predict consumers' preferences but two different shopping patterns, namely goal directed and exploratory search behavior. To predict shopping patterns, they analyzed the collected ET data of 29 participants in VR (a room-sized CAVE environment) and 20 in a real supermarket. Their evaluation covered increasing time windows on a per second basis. These windows were calculated using the intervals from the start of each trial to [5; 100] seconds into the decision-making process, increasing by one second. Due to the experimental setup, the classes were balanced, which is different compared to data presented in our study. They used shallow machine-learning approaches for point-in-time related features and not for time series. We call these features cross-sectional, as they are single values which are aggregated over the whole predefined period. This work identified the total number of fixated products and the variance of the average fixation duration among the most important predictor variables.

Millecamp et al. (2021) reported gaze pattern classification results for personality traits in the context of a browser-based music recommender system. The authors conducted a study with 30 participants in which eye movements were recorded using a desktop-based tracker. Their goal was to acquire predictions about the participants' openness, need for cognition, and musical sophistication. The authors considered 30%, 60%, and 90% of the data as time windows for their predictions. These time windows were less than the whole task duration but 60% and 90% of the decision-making process cannot be considered as particularly early stages. In general, their work showed the potential of using ET for adaptation of recommendations and explanations. However, in the conclusion they outlined improvement potential for the model's performance and called for further research on different tasks and interfaces.

Our search for related work indicates a research gap that previous authors did not particularly focus on early prediction of consumer preferences based on gaze patterns. So far, no proposal has been made to leverage ET data to generate features for recommender systems in VR which are generated possibly early in customer decision-making processes. Furthermore, to the best of our knowledge, no previous study used ET data with a state-of-the-art time series classification model to predict customer choices for healthy products. Using time series can improve performance because of leveraging information retrieved from behavior over time.

2.3 Method

2.3.1 Experimental Design

As dependent variable, we are interested in the healthiness of different muesli (cereal) purchase decisions. To categorize all products as healthy or unhealthy, the package label serves as a discriminative criterion. Representatives of the healthy and unhealthy classes are illustrated in Figure 1, where the left package is the healthy and the right package is the unhealthy alternative. The highlighted healthy label reads "without added sugar, wholegrain". We categorized a product as healthy if the packaging indicated at least reduced (or no) sugar or fat. According to this definition, seven out of the total 40 available products in the experiment were marked as healthy products. In total, out of 1040 product choices, 158 (15.2%) were for healthy products. The imbalanced class ratio leads to methodological challenges, which we discuss in the section on the treatment of class imbalances.

Figure 1. Criterion for healthy (left) and unhealthy (right) is the packaging label.

Our observations of retail purchase decisions in VR were collected in a controlled environment in a European University laboratory. The experimental design allowed our research group to answer several questions. Thus, the data is used in further studies which investigate the impact of low versus high immersion on system adoption (Peukert et al. 2019) and the impact of virtual reality in a conjoint-based choice analysis (Meißner et al. 2020). The VR scene was created using Unity 5.5.3f1 game engine. Participants were situated in a plain virtual room with a shelf of product packages and a shopping cart, as shown in Figure 2 (slightly distorted due to copyright reasons). We used an HTC Vive HMD with a dual display with 2160×1200 pixels resolution, an integrated SMI eye tracker, and HTC hand-held controllers.

Figure 2. Virtual environment with a muesli package shelf and a shopping cart.

After signing an informed consent form, all participants made multiple product choices in front of a product shelf. Participation compensation amounted to 14 Euro in total. To provide an economic incentive, participants received one of their product choices at random as part of their compensation. We instructed the participants to choose according to their natural preference and subtracted the cost of the chosen product from the monetary payout. Each experimental session was preceded by a training phase to familiarize the participants with the virtual environment. For this training, the shelf was filled with baking mixtures. In the subsequent experimental trials, the virtual shelf contained muesli products. In total, it held 24 different options which were selected from a product pool of 40 mueslis. Their arrangement followed a design which was suited for a conjoint-based choice analysis (Chrzan and Orme 2000). At any time, the product positioning ruled out centrality effects (Atalay et al. 2012). Furthermore, we positioned mueslis of the same brand close to each other. For each trial, one out of 171 product arrangements were displayed on the shelf. On average, the shelf contained 4.27 (SD 1.09) healthy products.

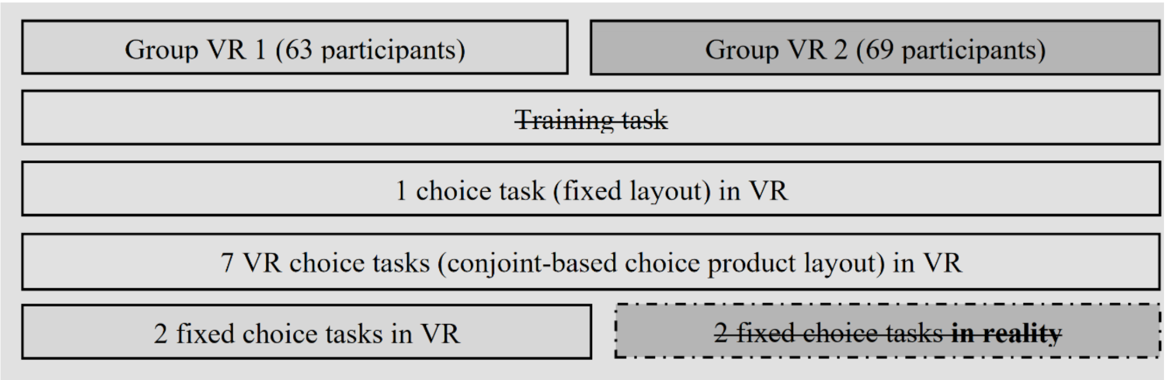

Figure 3. The experimental setup. We exclude the training task and real-world decisions.

Our sample consists of 132 student recordings, of which 45 were females and 87 males, with an average age of 22.13 (SD 1.98). The experiment followed a between-subjects design in which one treatment group was asked to make their last two purchase decisions in front of a real shelf (with real products). We had to exclude these real-world tasks because the ET equipment differed substantially between the VR and real-world setup. Thus, each participant made a total of either eight or ten product choices in VR, depending on the treatment group (see Figure 3). In other words, for the present study, we only used purchase decisions that were made in VR. After excluding the training task and erroneous recordings, the VR trials yield 1040 product choices, with an average decision duration of 54.91 seconds (SD

33.49). However, we further reduced the number of evaluated trials in the preprocessing because many of the respective decision-making processes were too short (less than 45 seconds) to separate them meaningfully into sub-phases (like orientation and evaluation). We chose 45 seconds as cutoff duration because of logic considerations about a decision process: a participant would need approximately 15 seconds to get an overview over the assortment and another 30 seconds to decide between the items in their consideration set (Hauser 2014).

2.3.2 Preprocessing

First, we determined fixations from the raw ET data and calculated the subject's gaze target for each fixation, which we tracked by means of ray casting (Pietroszek 2019). We did not consider blinks, pupil dilation and saccades. However, we emphasize that additional features could further improve predictive performance. In this paper we deliberately chose to focus on visual attention, which is best described by fixations (Holmqvist et al. 2011). In general, fixations last between 0.2 and 0.4 seconds. Fixations of less than 0.1 seconds were excluded, as they are too short for conscious information processing (Duchowski 2017). Fixations lasting longer than 10 seconds were also excluded, as they most likely indicate unnatural behavior or faulty sensor information. Predefined areas of interest comprised different parts of the individual product packages and their related price tags. This enabled us to discriminate fixations on different product groups (healthy and unhealthy products). Furthermore, fixations on each individual product and individual product's nutrition table were treated separately.

Transforming the gaze data into a discrete multivariate time series is the next preprocessing step. To aggregate the fixations into discrete bins, it was necessary to choose different step sizes for the cut-off points of the bins. We evaluated the step values (0.5, 0.6, 0.7, 0.8, 0.9, 1, 2, 3, 4, 5, 6, 9) seconds for the time series generation process. These values are based on reasoning about the average and maximum duration of a single fixation as described above. Shorter steps would often contain no fixation at all, and longer periods would cover too many fixations and be too coarse. We applied a sliding window technique (Hota et al. 2017) such that all bins overlapped with the previous one by 50%. The purpose of applying a sliding window is to capture interesting patterns that might be hidden by disjoint intervals. For each step size, we calculated the number of fixations, mean, variance, and skewness of the fixation duration (overall and for each of the areas of interest separately).

Our goal is to provide recommendations as early as possible during the evaluation phase of the respective decision. Therefore, we aimed to partially cut off the orientation and validation phase of the decision process as described in the on-the-fly-detection decision phase model by Peukert et al. (2020). In the orientation phase, consumers scan their environment, get an overview of the assortment, and do not compare different product choices in detail. For our data, the average transition from orientation to evaluation occurred in second 8 and the second transition from evaluation to verification occurred in second 47. Accordingly, we considered all integers in the interval [0;15] seconds as start values for our time series and all integers in the interval [20; 45] seconds as stop values. Using these intervals logically entailed to exclude decisions which lasted less than 45 seconds. Therefore, keeping shorter decisions would have confounded the input time series because trials shorter than 45 seconds would have to be filled with default values. After excluding all purchase processes shorter than 45 seconds, 516 relevant product choices remained for evaluation, with 78 (15.1%) healthy choices. To train and evaluate the classification models, a random split of training (60%), validation (20%), and test (20%) was used. We also allowed for recurring customers, i.e., we did not assign all trials of one participant to a single set. This means we assume that customers can return to the store, which is typical for grocery shopping.

2.3.3 Time Series Classifier

The deep learning approach InceptionTime (Ismail Fawaz et al. 2020) is a time series specific successor to the image classification model Inception, also referred to as GoogLeNet (Szegedy et al. 2015). InceptionTime is one of the current state-of-the-art deep learning approaches for time series classification (Middlehurst et al. 2021). The InceptionTime building blocks mainly consist of convolutional layers and pooling layers (Aggarwal 2018). The reference implementation proposes to stack six InceptionTime modules sequentially. As shown on the left in Figure 4, each module consists of several stages. A bottleneck layer (stage 1a) reduces the input dimensionality. The main components are three convolutional layers of different kernel sizes (stage 2a). Additionally, a parallel MaxPooling layer (stage 1b) makes the model invariant to small perturbations. This is followed by another bottleneck layer (stage 2b) to reduce dimensionality. At the end of each module (stage 3), the output of the convolutions and the max pooling operation are concatenated and serve as input to the next layer. As shown on the right in Figure 4, InceptionTime uses shortcut connections between every third InceptionTime module. These shortcuts help to overcome the vanishing gradient problem (Hochreiter 1998) and overfitting (Goodfellow et al. 2016). Finally, a dense classification head (a fully connected softmax layer) outputs the predicted probabilities for each class.

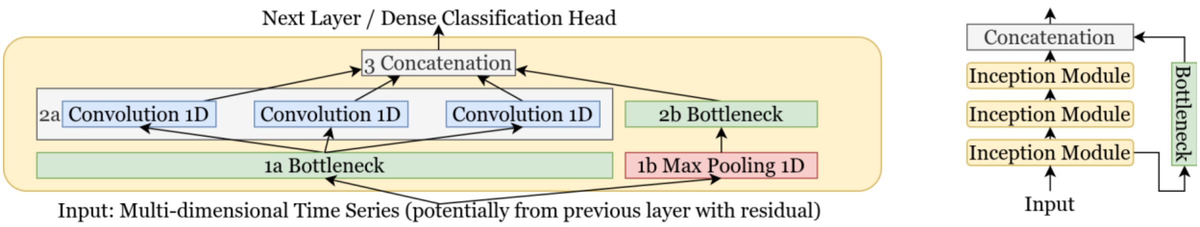

Figure 4. An InceptionTime module on the left and a shortcut connection on the right. Adapted from Ismail Fawaz et al. (2020).

The Inception architecture is based on two main ideas: First, reducing the dimensionality (via bottleneck layers) keeps the computational complexity low and mitigates overfitting for small datasets. Second, convolutional components with different receptive fields capture different aspects of the time series (Luo et al. 2016). For temporal data, the receptive field can be thought of as the maximum field of view of a neuron. The larger the receptive field is, the longer the patterns that can be detected by the neuron. The model uses multiple parallel, densely connected convolutional layers with different kernel sizes (see Figure 4, stage 2a) that allow to capture different aspects of the time series. During the trials, an asynchronous hyperband scheduler (Li et al. 2020) facilitated the exploration of 50 different combinations. Table 1 shows the complete hyperparameter space.

Table 1. The hyperparameter space which we used in the InceptionTime tuning process.

Name	**Values**	**Description**
Activation function	ReLU (Agarap 2019); eLU (Clevert et al. 2015)	
Alpha	[0.1; 0.3] Uniform	Focal loss
Bottleneck size	(32, 64, 128)	Inception Module 1a, 2b
Gamma	[0.1; 0.3] Uniform	Focal loss
Kernel Multiplier	(4, 6, 8, 18)	Inception Module 2a
Learning Rate	[1e-1, 1e-6] Log uniform	Optimizer
Num Filters	(8, 16, 32)	Inception Module 2a
Num Modules	(3, 6)	InceptionTime

In total, the different start, stop, and step size values resulted in 4990 possible combinations. A high-performance cluster was used to compute all respective trials. The Ray Tune framework (Liaw et al. 2018), combined with the slurm task scheduler (Yoo et al.), allowed us to partially parallelize the optimization of the InceptionTime instances, which all ran for a maximum of 100 epochs, using up to 75 compute nodes equipped with 24 CPU cores.

2.3.4 Class Imbalance Treatment

In our data, only 15.2% of choices were for healthy products. The applied methods need to take this class imbalance into account. Otherwise, classifiers tend to always predict the majority class. Different paradigms to treat imbalanced data exist, namely data-level, algorithm-level, and hybrid methods (Krawczyk 2016). We used an α balanced focal loss function (Lin et al. 2017) for the neural network optimizer to discount the majority classes. It is a hybrid approach that combines cost modifying and algorithmic adjustments. Focal loss is a modification of the widely used cross-entropy loss function (Goodfellow et al. 2016). The main idea is to discount correctly classified samples of the majority class, i.e., the contribution to the total loss value is large for wrong predictions of the minority class. Focal loss and can be denoted as

$$FocalLoss(p_t) = -\alpha_t\,(1-p_t)^{\gamma}\log(p_t), \text{ with } p_t = \begin{cases} p, & \text{if } y = 1 \\ 1-p, & \text{otherwise} \end{cases}. \qquad (1)$$

Parameter α_t specifies the minority class proportion in the test data set, $p_t \in [0, 1]$ is the predicted class probability for the sample, and $y \in \{0, 1\}$ is the target label. The focusing intensity $\gamma \geq 0$ determines the rate for discounting easy samples. Note that, when $\gamma = 0$, focal loss equals cross-entropy.

A further algorithmic measure is the evaluation with a suited scoring metric. For the prediction of imbalanced data the accuracy metric is unexpressive (Bekkar et al. 2013). Accuracy would put too much attention on unhealthy product choices (precision) and too little on healthy ones (recall). The F_β metric allows to adjust the trade-off between recall and precision (Maratea et al. 2014). A value for parameter β greater than one emphasizes the importance of recall while a value less than one emphasizes the importance of precision. For this study, $\beta = 1.5$ is used because we focus more on recall than on precision. Choosing $F_{\beta=1.5}$ means we deliberately expose some of the purchasers of unhealthy mueslis to recommendations for healthy products as trade-off for a higher classification rate of intended healthy product choices (which may be interpreted as a form of nudging). The F_β score (Maratea et al. 2014) can be denoted as

$$F_\beta = (1+\beta^2) \cdot \frac{Precision \cdot Recall}{(\beta^2 \cdot Precision) + Recall}. \qquad (2)$$

2.3.5 Gradient Boosting Trees Baseline

This gradient boosting baseline considers aggregated, one-dimensional, features, which is the current standard. Utilizing multi-dimensional features in form of time series is more promising because it allows considering the complete decision-making process in form of a vector, from the start of the purchase situation until the point-of-time when a recommendation should be made.

Gradient boosting served as a baseline for this study, as it has shown good results in similar setups (Millecamp et al. 2021; Pfeiffer et al. 2020). We implemented it using the XGBoost (Chen and Guestrin 2016) and scikit-learn (Pedregosa et al. 2011) packages. This model did not require a distinct validation dataset for training. Instead a 10-fold cross validation (Refaeilzadeh et al. 2009) ensured generalizability on the data set, permuting the combined training and validation subsets. The features for the gradient boosting model consist of the same underlying information (e.g., number of fixations) but aggregate it with respect to the total interval length. Analogous to the time series, we used $F_{\beta=1.5}$ as scoring metric and chose the intervals [0; 15] for start timestamps and [20; 45] for stop timestamps. To find a good set of hyperparameters (colsample_bytree, gamma, learning_rate, max_depth, min_child_weight, n_estimators, scale_pos_weight, subsample) a randomized search was performed for 100 trials on all possible start-stop combinations.

2.4 Results

Figure 5 shows two different prediction horizons (i) the first 25 seconds and (ii) the first 45 seconds of the decision process. The 25-second horizon is based on the idea of making recommendations early in the product evaluation process. A recommender system would have enough time to generate content after a feature extraction phase of 25 seconds at the beginning of the decision process. On average, the 45-second horizon covers the entire evaluation phase and can be seen as the upper limit for a recommender system to make suggestions.

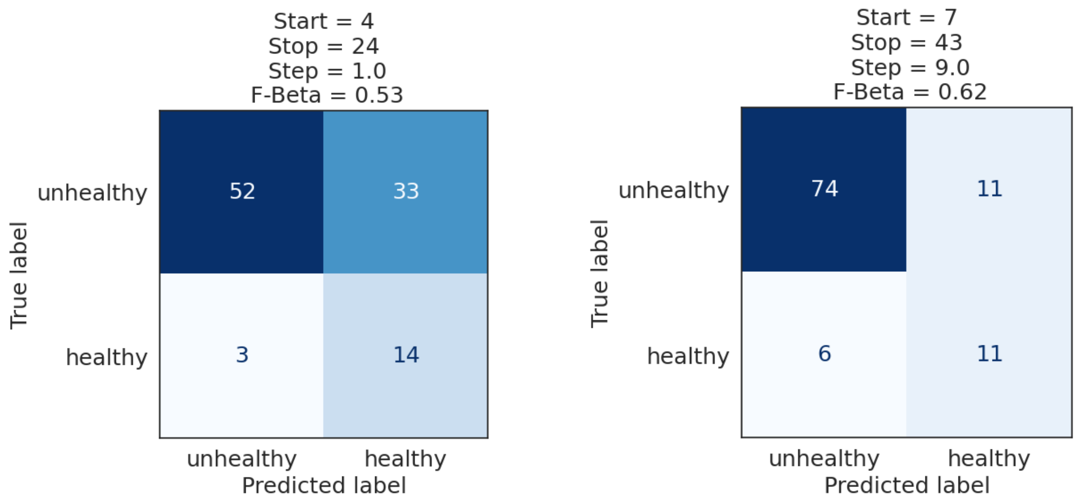

Figure 5. The confusion matrices represent the best InceptionTime models for healthiness preference predictions within the first 25 (left) and 45 (right) seconds.

The best model for the entire prediction horizon of 45 seconds ($F_{\beta=1.5} = 0.62$) did not use the full 45 seconds. It performed best when considering the time series from second 7 to 43, with a step size of 9.0 seconds. This model correctly classified 87.05% of the unhealthy choices and 64.70% of the healthy choices.

The model for the shorter prediction horizon of 25 seconds does not perform much worse overall (61.17% correct unhealthy classification and 82.35% correct healthy classification). It achieved an $F_{\beta=1.5}$ score of 0.53. We remind the reader that with a beta of 1.5, we value recall higher than precision, i.e., finding most of the healthy choices has priority. This model considered the period from second 4 to 24 as a time series, using a start-stop interval of [4; 24], and a step size of 1.0 second. It even correctly classified more healthy choices correctly compared to the best model for the 45-second prediction horizon.

In contrast, the best performing XGBoost model achieved an $F_{\beta=1.5}$ score of 0.48, using a start-stop interval of [0; 38]. It correctly classified 90.5% of the unhealthy choices but only 47.1% of the healthy choices. With respect to the prediction horizon of 25 seconds, the best XGBoost model performed slightly worse with an $F_{\beta=1.5}$ score of 0.42, using a start-stop interval of [5; 21].

In Figure 6, we provide information about the effect of different start and stop values on the maximum F_{β} classification performance. The left plot shows the average effect of different start values. For our data, starting in second 5 results in the best average F_{β} value. As expected, a decrease in performance occurs when a long onset duration is omitted before feeding the model. The right plot shows the average impact of different stop values with a peak at second 30. The positive trend for later stop values is also plausible, as more information becomes available over time.

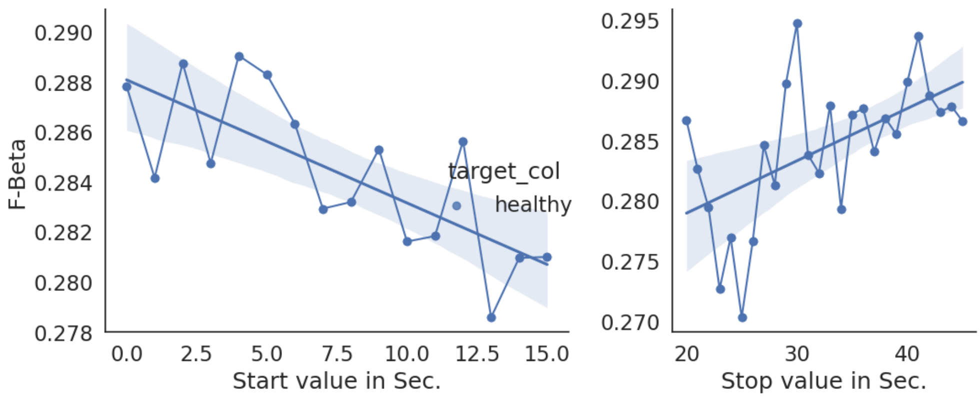

Figure 6. A timeline showing the average F-Beta value for healthiness preferences predictions regarding all evaluated start and stop values.

2.5 Discussion

Although the amount of training data we used was limited and the class distribution imbalanced, our work demonstrates a way to use gaze behavior, in our case the extracted fixations and gaze targets, as input for recommender systems. With the algorithmic adjustments regarding the misclassification of the majority class, we can clearly answer our research question. Yes, with reasonable performance in relation to the limited amount of data, we can identify customers who buy healthy products early during their decision-making process in a virtual commerce scenario using the InceptionTime deep learning approach. However, we acknowledge that current classification rates are not production ready and continuous model improvement and data collection are required to eventually allow for accurate predictions.

Our main aim was to correctly classify as many samples of the minority class (healthy choices) as possible during the evaluation phase of the decision processes. For the given data set, our results suggest that a time series-based approach like InceptionTime is a more appropriate classifier compared to the shallow XGBoost method. The InceptionTime model with a 1.0 second step size and a start-stop interval of [4; 24] is a promising predictor for healthy and unhealthy product choices early in the decision process. This model showed that focal loss and the F_β metric are effective measures to cope with the class imbalance inherent to the data set. It achieved the highest $F_{\beta=1.5}$ score of 0.53 in our evaluation and correctly classified most of the healthy choices (14 out of 17) while generating nudges candidates (a fraction of customers with unhealthy choices, 33 out of 87). The extent of candidate-generation could be adjusted by the β parameter for the evaluation metric (in our case we chose β=1.5 and argue that it was a good choice because the amount of nudge candidate seems to be appropriate).

The best XGBoost model achieved an $F_{\beta=1.5}$ score of 0.48 in our evaluation. It correctly classified less than half of the healthy choices correctly (8 out of 17) and generated only a small number of nudge candidates (8 out of 87). One reason for the lower performance could be the fact that the scikit-learn implementation of XGBoost does offer a focal loss function. However, recently Wang et al. (2020) implemented a focal loss function for the XGBoost algorithm that may serve as drop-in replacement in scikit-learn. A comparison between InceptionTime and an XGBoost model with implemented focal losses might offer a more equitable benchmark and could potentially alter the results' significance. The XGBoost model could also benefit from advanced sampling techniques, such as creating synthetic samples with small deviations from the real observations (Chawla et al. 2002), but this is beyond the scope of this research.

Regarding start values, the best InceptionTime models started early for both the 45 and the 25 second prediction horizon (second 4 and 7, respectively). Thus, it seems advisable to start the time series possibly early. Another viable option may be to decide on an individual case basis when the orientation phase ends, e.g., by detecting the gaze pattern which represents the first comparison of two products (Peukert et al. 2020). In terms of stop values, the results unsurprisingly exhibit a positive linear trend for the maximum F_β score, i.e., an increase of performance with the duration of the prediction horizon. However, the graph shows a lot of variances around second 25, 30, and 40 and it might be counterintuitive that for the stop values after second 40, the maximum F_β values mainly decrease. For earlier stop values, our results show that the prediction quality can remain relatively good, e.g., when stopping after 24 seconds. The corresponding InceptionTime model correctly classified only one healthy customer less (out of 17) than the best InceptionTime model which had access to additional 16 seconds of ET data of the decision processes. This further supports the importance of the early decision phase for the correct classification of healthy customers.

An open question remains the choice of a possibly ideal step size. We evaluated many different step values, which cost a lot of (computation) time and energy. Finding and validating a better theoretical foundation, to explain for what reasons a certain overlapping technique should be applied, could prove very helpful.

As theoretical contribution, our study confirms that leveraging a complete time series of ET data and feed it into a convolutional network can be superior to treating the ET data as cross-sectional data. However, the performance gain in comparison to a basic XGBoost model is only a first proof of concept and both the baseline and the classification model can further improve.

Before closing, we reflect on ethical considerations, particularly with regard the use of our classification model as input for recommender and other context-aware AI systems. We used gaze information of our participants to infer their willingness to buy healthy food and prioritized healthy purchases. In the design of our model, we accepted a bias towards healthy classification, what may lead to a nudge for a certain fraction of customers who would not necessarily appreciate suggestions for a healthy product. We argue that such a nudge would be ethically valid, as it fosters socially desired behavior. However, there is an ongoing debate in what situations nudging is desirable and when it should be avoided altogether (Hausman and Welch 2010). In any case, "[c]hoice architecture, both good and bad, is pervasive and unavoidable, and it greatly affects our decisions." (Thaler and Sunstein 2021, p. 252).

From a technical standpoint, our study suggests that time series classification enables real-time feature generation for recommender systems using gaze patterns. Our results indicate that the longitudinal point of view offers more relevant information than aggregations to statistical moments that span over the whole decision period. We acknowledge that further research and validation are needed to improve the reliability and generalizability of our findings. Nonetheless, we hope that the presented approach encourages practitioners to integrate recommender systems in virtual commerce environments. From our point of view, it is only a question of time until we experience various (most likely artificial intelligence assisted) tools which support and improve healthy food choices based on individual sensor data. Overall, the use of suitable deep learning models, such as InceptionTime, could potentially change the state-of-the-art for developing personalized interventions. In combination with large language models, time series classification and cutting-edge deep learning methods are likely to transform user assistance as we know it today. Researchers and practitioners might think about further contexts beyond classic collaborative filtering, such as personal trainers and instructors, medical advisors, psychotherapeutic treatments, and more. The presented approach could be applied everywhere where learning about users' preferences or their decision processes in general can be helpful. Therefore, it seems advisable to continue with data acquisition, model evaluation, and workflow integration.

2.6 Summary and Outlook

We proposed to use an InceptionTime classifier to infer customer preferences during the evaluation phase of customer decision processes using gaze patterns. Our focus was on classifying customers who buy healthy products in a VR setup. The results show that InceptionTime, in combination with class imbalance measures, can outperform a shallow gradient boosting model in classifying healthy purchase decisions while generating candidates for healthy food nudges.

The main limitation of this study is the fact that our sample consists of only 516 purchase decisions, of which only 15.1% were made for healthy products. Deep learning models are typically trained on much larger datasets (Szegedy et al. 2016), and we believe that the full potential of deep time series classification approaches will remain unexplored until such a large dataset becomes (publicly) available. However, in order to collect such a dataset, the legal consensus regarding privacy concerns for ET data needs to be solidified. Another limitation of this study is that we only considered product labels (the most visually salient information) to classify products as unhealthy or healthy when defining the ground truth. Future research could use more fine-grained information, such as ingredient lists and nutritional tables. With detailed information about a product's composition, recommendations could take additional aspects into account. A highly relevant example is the detection of allergies, e.g., many people are allergic to nuts. Consumers could decide whether to hide such products altogether or receive a multi-sensory warning when they focus on a critical product.

Overall, we see several avenues for future virtual commerce focused research. One prominent concern is the treatment of privacy issues. Deliberate actions, such as body movements or use of voice, can be controlled by the customers. In contrast, the gaze as such is less under consumer control and fundamental to decision-making. ET data can identify individuals and might reveal unwanted personal aspects (Cantoni et al. 2018). Thus, research should invent, evaluate, and reflect on different suitable (pseudo) anonymization techniques (Steil et al. 2019). Privacy research enables device vendors and digital commerce providers to avoid pitfalls and fosters trust among customers. The nudge aspect of this work is another route to follow. Healthiness is only one aspect of socially desirable behavior but there are further areas, such as sustainable consumption, which could be investigated by further research.

Regarding data collection, upcoming studies should include a broader variety of available information. Pupillometry and additional bio sensors seem to be a promising source for additional input features (Halbig and Latoschik 2021). Furthermore, time series classification evolves quickly and new classifiers emerge frequently, e.g., InceptionFCN (Usmankhujaev et al. 2021) or TapNet (Zhang et al. 2020). These models may have the potential to yield better classification rates and should be compared with the presented results.

Future research should predict further dependent variables and showcase a real recommendation pipeline. In addition to healthy products, we argue that brand and flavor preferences are particularly interesting. Such a follow-up study should rethink the large-scale hyperparameter searches. These searches do not necessarily enumerate all presented start and stop value combinations as presented in this study. Instead, it should benchmark different algorithmic design aspects, like predicting preferences for new customers only or limiting the feature set, which would provide further managerial insights. Next, a follow up should introduce a better baseline, e.g., by comparing InceptionTime with previously mentioned deep learning time series classification methods. Overall, we suggest iterative improvements by means of ongoing experiments with the latest sensor technology available, such as electroencephalography (event related potentials), facial features, body posture, pupil dilation, and maybe functional near-infrared spectroscopy (fNIRS). With all measures combined, we expect the predictive performance and validity to improve significantly (unfortunately, the same is true for complexity). From our perspective, a long-term goal should be to hone a publicly available machine learning pipeline, similar to the presented one, and ultimately showcase it as real-time feature generator for a recommender system in real virtual commerce setups.

References

Agarap, Abien Fred (2019): Deep Learning using Rectified Linear Units (ReLU): arXiv.

Aggarwal, Charu C. (2018): Neural networks and deep learning. In *Springer* 10, 978-3.

Alcañiz, Mariano; Bigné, Enrique; Guixeres, Jaime (2019): Virtual reality in marketing: a framework, review, and research agenda. In *Frontiers in psychology* 10, p. 1530.

Arya, Vikas; Sambyal, Rachita; Sharma, Anshuman; Dwivedi, Yogesh K. (2023): Brands are calling your AVATAR in Metaverse-A study to explore XR-based gamification marketing activities & consumer-based brand equity in virtual world. In *Journal of Consumer Behaviour*.

Atalay, A. Selin; Bodur, H. Onur; Rasolofoarison, Dina (2012): Shining in the center: Central gaze cascade effect on product choice. In *Journal of consumer research* 39 (4), pp. 848–866.

Bekkar, Mohamed; Djemaa, Hassiba Kheliouane; Alitouche, Taklit Akrouf (2013): Evaluation measures for models assessment over imbalanced data sets. In *Journal of Information Engineering and Applications* 3 (10).

Block, Lauren G.; Grier, Sonya A.; Childers, Terry L.; Davis, Brennan; Ebert, Jane E.J.; Kumanyika, Shiriki et al. (2011): From Nutrients to Nurturance: A Conceptual Introduction to Food Well-Being. In *Journal of Public Policy & Marketing* 30 (1), pp. 5–13.

Bourlakis, Michael; Papagiannidis, Savvas; Li, Feng (2009): Retail spatial evolution: paving the way from traditional to metaverse retailing. In *Electronic Commerce Research* 9 (1), pp. 135–148.

Bublitz, Melissa G.; Peracchio, Laura A.; Andreasen, Alan R.; Kees, Jeremy; Kidwell, Blair; Miller, Elizabeth Gelfand et al. (2013): Promoting positive change: Advancing the food well-being paradigm. In *Journal of Business Research* 66 (8), pp. 1211–1218.

Cantoni, Virginio; Musci, Mirto; Nugrahaningsih, Nahumi; Porta, Marco (2018): Gaze-based biometrics: An introduction to forensic applications. In *Pattern Recognition Letters* 113, pp. 54–57.

Castagnos, Sylvain; Jones, Nicolas; Pu, Pearl (2010): Eye-tracking product recommenders' usage. In: Proceedings of the fourth ACM conference on Recommender systems, pp. 29–36.

Chawla, Nitesh V.; Bowyer, Kevin W.; Hall, Lawrence O.; Kegelmeyer, W. Philip (2002): SMOTE: synthetic minority over-sampling technique. In *Journal of artificial intelligence research* 16, pp. 321–357.

Chen, Tianqi; Guestrin, Carlos (2016): Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785–794.

Cho, Jaehee; Park, Dongjin; Lee, H. Erin; others (2014): Cognitive factors of using health apps: systematic analysis of relationships among health consciousness, health information orientation, eHealth literacy, and health app use efficacy. In *Journal of medical Internet research* 16 (5), e3283.

Chrzan, Keith; Orme, Bryan (2000): An overview and comparison of design strategies for choice-based conjoint analysis. In *Sawtooth software research paper series* 98382, p. 360.

Clevert, Djork-Arné; Unterthiner, Thomas; Hochreiter, Sepp (2015): Fast and accurate deep network learning by exponential linear units (elus). In *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics* 20, pp. 924–932.

David-John, Brendan; Peacock, Candace; Zhang, Ting; Murdison, T. Scott; Benko, Hrvoje; Jonker, Tanya R. (2021): Towards gaze-based prediction of the intent to interact in virtual reality. In: ACM Symposium on Eye Tracking Research and Applications, pp. 1–7.

Duchowski, Andrew T. (2017): Eye tracking methodology: Theory and practice: Springer.

Evans, Philip; Wurster, Thomas S. (1999): Getting real about virtual commerce. In *Harvard Business Review* 77, pp. 84–98.

Fenko, Anna; Nicolaas, Iris; Galetzka, Mirjam (2018): Does attention to health labels predict a healthy food choice? An eye-tracking study. In *Food quality and preference* 69, pp. 57–65.

Fernandez, Rosa Maria (2019): SDG3 Good Health and Well-Being: Integration and Connection with Other SDGs. In: Encyclopedia of the UN Sustainable Development Goals. Cham: Springer International Publishing, pp. 629–636.

Gadalla, Eman; Keeling, Kathy; Abosag, Ibrahim (2013): Metaverse-retail service quality: A future framework for retail service quality in the 3D internet. In *Journal of Marketing Management* 29 (13-14), pp. 1493–1517.

Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron (2016): Deep learning: MIT press.

Halbig, Andreas; Latoschik, Marc Erich (2021): A Systematic Review of Physiological Measurements, Factors, Methods, and Applications in Virtual Reality. In *Front. Virtual Real.* 2, Article 694567, p. 89.

Hauser, John R. (2014): Consideration-set heuristics. In *Journal of Business Research* 67 (8), pp. 1688–1699.

Hausman, Daniel M.; Welch, Brynn (2010): Debate: To nudge or not to nudge. In *Journal of Political Philosophy* 18 (1), pp. 123–136.

Hercberg, Serge; Touvier, Mathilde; Salas-Salvado, Jordi (2021): The Nutri-score nutrition label: Hogrefe AG.

Hoang, Hanh Thi Hong; Nguyen, Lam Son; Nguyen, Chinh Hong; Le, Nga Viet; Nguyen, Nhung Thi; Dinh, Len Thi (2023): Personalisation-Privacy Paradox from Marketing Perspectives: Literature Review and Future Research Directions. In *Business Innovation for the Post-pandemic Era in Vietnam*, pp. 87–105.

Hochreiter, Sepp (1998): The vanishing gradient problem during learning recurrent neural nets. In *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 6 (02), pp. 107–116.

Holmqvist, Kenneth; Nyström, Marcus; Andersson, Richard; Dewhurst, Richard; Jarodzka, Halszka; van de Weijer, Joost (2011): Eye tracking: A comprehensive guide to methods and measures: OUP Oxford.

Hota, H. S.; Handa, Richa; Shrivas, Akhilesh Kumar (2017): Time series data prediction using sliding window based RBF neural network. In *International Journal of Computational Intelligence Research* 13 (5), pp. 1145–1156.

Ismail Fawaz, Hassan; Lucas, Benjamin; Forestier, Germain; Pelletier, Charlotte; Schmidt, Daniel F.; Weber, Jonathan et al. (2020): Inceptiontime: Finding alexnet for time series classification. In *Data Mining and Knowledge Discovery* 34 (6), pp. 1936–1962.

Jensen, Lasse; Konradsen, Flemming (2018): A review of the use of virtual reality head-mounted displays in education and training. In *Education and Information Technologies* 23 (4), pp. 1515–1529.

Kim, Eojina; Tang, Liang Rebecca; Meusel, Chase; Gupta, Manjul (2018): Optimization of menu-labeling formats to drive healthy dining: An eye tracking study. In *International Journal of Hospitality Management* 70, pp. 37–48.

Kliestik, Tomas; Novak, Andrej; Lazaroiu, George (2022): Live Shopping in the Metaverse: Visual and Spatial Analytics, Cognitive Artificial Intelligence Techniques and Algorithms, and Immersive Digital Simulations. In *Linguistic & Philosophical Investigations*.

Kovacova, Maria; Machova, Veronika; Bennett, Daniel (2022): Immersive extended reality technologies, data visualization tools, and customer behavior analytics in the metaverse commerce. In *Journal of Self-Governance and Management Economics* 10 (2), pp. 7–21.

Krajbich, Ian; Rangel, Antonio (2011): Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. In *Proceedings of the National Academy of Sciences of the United States of America* 108 (33), pp. 13852–13857.

Krawczyk, Bartosz (2016): Learning from imbalanced data: open challenges and future directions. In *Progress in Artificial Intelligence* 5 (4), pp. 221–232.

Li, Liam; Jamieson, Kevin; Rostamizadeh, Afshin; Gonina, Ekaterina; Hardt, Moritz; Recht, Benjamin; Talwalkar, Ameet (2020): A system for massively parallel hyperparameter tuning. In *Proceedings of Machine Learning and Systems* 2, pp. 230–246.

Li, Xitong; Grahl, Jörn; Hinz, Oliver (2022): How Do Recommender Systems Lead to Consumer Purchases? A Causal Mediation Analysis of a Field Experiment. In *Information Systems Research* 33 (2), pp. 620–637.

Liaw, Richard; Liang, Eric; Nishihara, Robert; Moritz, Philipp; Gonzalez, Joseph E.; Stoica, Ion (2018): Tune: A Research Platform for Distributed Model Selection and Training: arXiv.

Lin, Tsung-Yi; Goyal, Priya; Girshick, Ross; He, Kaiming; Dollár, Piotr (2017): Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision, pp. 2980–2988.

Luo, Wenjie; Li, Yujia; Urtasun, Raquel; Zemel, Richard (2016): Understanding the effective receptive field in deep convolutional neural networks. In *Advances in neural information processing systems* 29.

Maratea, Antonio; Petrosino, Alfredo; Manzo, Mario (2014): Adjusted F-measure and kernel scaling for imbalanced data learning. In *Information Sciences* 257, pp. 331–341.

Mariotti, Thomas; Schweizer, Nikolaus; Szech, Nora; Wangenheim, Jonas von (2023): Information nudges and self-control. In *Management Science* 69 (4), pp. 2182–2197.

Meißner, Martin; Pfeiffer, Jella; Peukert, Christian; Dietrich, Holger; Pfeiffer, Thies (2020): How virtual reality affects consumer choice. In *Journal of Business Research* 117, pp. 219–231.

Meißner, Martin; Pfeiffer, Jella; Pfeiffer, Thies; Oppewal, Harmen (2019): Combining virtual reality and mobile eye tracking to provide a naturalistic experimental environment for shopper research. In *Journal of Business Research* 100, pp. 445–458.

Middlehurst, Matthew; Large, James; Flynn, Michael; Lines, Jason; Bostrom, Aaron; Bagnall, Anthony (2021): HIVE-COTE 2.0: a new meta ensemble for time series classification. In *Machine Learning* 110 (11), pp. 3211–3243.

Millecamp, Martijn; Conati, Cristina; Verbert, Katrien (2021): Classifeye: classification of personal characteristics based on eye tracking data in a recommender system interface. In: Joint Proceedings of the ACM IUI 2021 Workshops co-located with 26th ACM Conference on Intelligent User Interfaces, vol. 2903. CEUR Workshop Proceedings.

Naruetharadhol, Phaninee; Wongsaichia, Sasichakorn; Pienwisetkaew, Teerapong; Schrank, Johannes; Chaiwongjarat, Kullanan; Thippawong, Peeranut et al. (2023): Consumer Intention to Utilize an E-Commerce Platform for Imperfect Vegetables Based on Health-Consciousness. In *Foods* 12 (6), p. 1166.

Papagiannidis, Savvas; Bourlakis, Michael A. (2010): Staging the New Retail Drama: at a Metaverse near you! In *Journal of Virtual Worlds Research* 2 (5), pp. 425–446.

Parashar, Sapna; Singh, Supriya; Sood, Gunjan (2023): Examining the role of health consciousness, environmental awareness and intention on purchase of organic food: A moderated model of attitude. In *Journal of Cleaner Production* 386, p. 135553.

Patney, Anjul; Salvi, Marco; Kim, Joohwan; Kaplanyan, Anton; Wyman, Chris; Benty, Nir et al. (2016): Towards foveated rendering for gaze-tracked virtual reality. In *ACM Transactions on Graphics (TOG)* 35 (6), pp. 1–12.

Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O. et al. (2011): Scikit-learn: Machine Learning in Python. In *Journal of Machine Learning Research* 12, pp. 2825–2830.

Peukert, Christian; Lechner, Jessica; Pfeiffer, Jella; Weinhardt, Christof (2020): Intelligent invocation: Towards designing context-aware user assistance systems based on real-time eye tracking data analysis. In: Information systems and neuroscience: Springer, pp. 73–82.

Peukert, Christian; Pfeiffer, Jella; Meißner, Martin; Pfeiffer, Thies; Weinhardt, Christof (2019): Shopping in virtual reality stores: the influence of immersion on system adoption. In *Journal of Management Information Systems* 36 (3), pp. 755–788.

Peukert, Christian; Weinhardt, Christof; Hinz, Oliver; van der Aalst, Wil M. P. (2022): Metaverse: How to Approach Its Challenges from a BISE Perspective: Springer. In *Business & Information Systems Engineering*.

Pfeiffer, Jella; Pfeiffer, Thies; Meißner, Martin; Weiß, Elisa (2020): Eye-tracking-based classification of information search behavior using machine learning: evidence from experiments in physical shops and virtual reality shopping environments. In *Information Systems Research* 31 (3), pp. 675–691.

Pietroszek, Krzysztof (2019): Raycasting in Virtual Reality: Springer International Publishing AG, part of Springer Nature 2018.

Popescu, Gheorghe H.; Valaskova, Katarina; Horak, Jakub (2022): Augmented Reality Shopping Experiences, Retail Business Analytics, and Machine Vision Algorithms in the Virtual Economy of the Metaverse. In *Journal of Self-Governance & Management Economics* 10 (2).

Refaeilzadeh, Payam; Tang, Lei; Liu, Huan (2009): Cross-validation. In *Encyclopedia of database systems* 5, pp. 532–538.

Regt, Anouk de; Barnes, Stuart J. (2019): V-Commerce in Retail: Nature and Potential Impact. In: Augmented Reality and Virtual Reality: Springer, Cham, pp. 17–25.

Shen, Bingqing; Tan, Weiming; Guo, Jingzhi; Zhao, Linshuang; Qin, Peng (2021): How to promote user purchase in metaverse? A systematic literature review on consumer behavior research and virtual commerce application design. In *Applied Sciences* 11 (23), p. 11087.

Shimojo, Shinsuke; Simion, Claudiu; Shimojo, Eiko; Scheier, Christian (2003): Gaze bias both reflects and influences preference. In *Nat Neurosci* 6 (12), pp. 1317–1322.

Spagnolo, Fanny; Corsonello, Pasquale; Frustaci, Fabio; Perri, Stefania (2023): Design of a low-power super-resolution architecture for virtual reality wearable devices. In *IEEE Sensors Journal* 23 (8), pp. 9009–9016.

Sriram, G. K. (2022): A Comprehensive Survey on Metaverse. In *International Research Journal of Modernization in Engineering Technology*.

Steil, Julian; Hagestedt, Inken; Huang, Michael Xuelin; Bulling, Andreas (2019): Privacy-aware eye tracking using differential privacy. In: Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications. New York, NY, USA. New York, NY, USA: ACM.

Szegedy, Christian; Liu, Wei; Jia, Yangqing; Sermanet, Pierre; Reed, Scott; Anguelov, Dragomir et al. (2015): Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9.

Szegedy, Christian; Vanhoucke, Vincent; Ioffe, Sergey; Shlens, Jon; Wojna, Zbigniew (2016): Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818–2826.

Takahashi, Shigeo; Uchita, Akane; Watanabe, Kazuho; Arikawa, Masatoshi (2022): Gaze-driven placement of items for proactive visual exploration. In *Journal of Visualization* 25 (3), pp. 613–633.

Thaler, Richard H.; Sunstein, Cass R. (2021): Nudge: The final edition: Yale University Press.

Tran, Thi Ngoc Trang; Atas, Müslüm; Felfernig, Alexander; Stettinger, Martin (2018): An overview of recommender systems in the healthy food domain. In *Journal of Intelligent Information Systems* 50 (3), pp. 501–526.

Usmankhujaev, Saidrasul; Ibrokhimov, Bunyodbek; Baydadaev, Shokhrukh; Kwon, Jangwoo (2021): Time Series Classification with InceptionFCN. In *Sensors* 22 (1), p. 157.

Wahl, Deborah R.; Villinger, Karoline; König, Laura M.; Ziesemer, Katrin; Schupp, Harald T.; Renner, Britta (2017): Healthy food choices are happy food choices: Evidence from a real life sample using smartphone based assessments. In *Scientific Reports* 7 (1), p. 17069.

Wang, Chen; Deng, Chengyuan; Wang, Suzhen (2020): Imbalance-XGBoost: leveraging weighted and focal losses for binary label-imbalanced classification with XGBoost. In *Pattern Recognition Letters* 136, pp. 190–197.

Weiß, Tobias; Merkl, Lukas; Pfeiffer, Jella (2023): Customer Decision-Making Processes Revisited: Insights from an Eye Tracking and ECG Study using a Hidden Markov Model. In: Davis, Riedl et al. (Hg.) 2023 – Information Systems and Neuroscience.

Xi, Nannan; Hamari, Juho (2021): Shopping in virtual reality: A literature review and future agenda. In *Journal of Business Research* 134, pp. 37–58.

Xu, Songhua; Jiang, Hao; Lau, Francis C. M. (2008): Personalized online document, image and video recommendation via commodity eye-tracking. In: Proceedings of the 2008 ACM conference on Recommender systems, pp. 83–90.

Yoo, Andy B.; Jette, Morris A.; Grondona, Mark: SLURM: Simple Linux Utility for Resource Management, pp. 44–60.

Zhang, Xuchao; Gao, Yifeng; Lin, Jessica; Lu, Chang-Tien (2020): TapNet: Multivariate Time Series Classification with Attentional Prototypical Network. In *AAAI* 34 (04), pp. 6845–6852.

Zhao, Qian; Chang, Shuo; Harper, F. Maxwell; Konstan, Joseph A. (2016): Gaze prediction for recommender systems. In: Proceedings of the 10th ACM Conference on Recommender Systems, pp. 131–138.

3 Paper B: Consumer Decisions in Virtual Commerce: Good Help-timing and its Prediction based on Cognitive Load

Tobias Weiß and Jella Pfeiffer

Abstract

The retail sector is steadily moving towards virtual commerce (v-commerce) and the process has recently gained momentum. With the latest developments in headset technology and the rise of artificial intelligence, virtual shopping becomes relevant for an increasing number of products. In this paper, we present a study that combines consumer behavior research, eye tracking, electrocardiography, machine learning, and the application of virtual reality. Fifty participants were invited to experience a virtual scenario, perform multiple mentally demanding tasks, and make a purchase decision for a product from one of two different product categories. In a post-hoc video analysis based on the first-person view, participants determined different points in time when they would have appreciated help from an algorithmic user assistance system or a digital human agent. Our statistical analysis suggests that the desired help timing depends on the product category. For fast-moving consumer goods, algorithmic help was requested particularly early. Furthermore, we collected eye tracking and electrocardiographic data to build and evaluate a predictive classification model that differentiates between three levels of cognitive load. The trained machine learning algorithm aims to classify cognitive load during decision-making, which may indicate the right time to offer help. Our findings provide evidence that particularly eye movements allow service providers to determine a good moment to approach consumers during their shopping experience.

Keywords: Consumer Behavior, Decision Making, Eye Tracking, Electrocardiography, Machine Learning, Metaverse, Virtual Commerce, Virtual Reality

3.1 Introduction

The popularity of online shopping has transformed traditional brick-and-mortar stores into highly competitive virtual marketplaces (Bourlakis et al. 2009). While technological advances provide new opportunities for consumers to visualize and experience their environment, new business rules pose challenges for retailers seeking to provide engaging and meaningful experiences (Reinartz et al. 2019). With the proliferation of immersive technologies such as virtual reality (VR), the idea of the Metaverse continues to fascinate many people. For immersive shopping scenarios, knowledge about cognitive processes can help to design highly personalized user assistance systems (UAS). Decision support systems are an elemental tool for retailers that can severely impact their business success (Shim et al. 2002). As a subclass, UAS can be seen as joint element which "bridge[s] the gap between the system's functionalities and the human's individual capabilities with the goal of positively influencing task outcomes" (Morana et al. 2020b, p. 189).

Due to the need for enhanced consumer experiences, several studies suggest that the provision of personalized user assistance will become highly relevant in v-commerce scenarios (Guo and Elgendi 2013; Zhang et al. 2013; Chen and Yang 2022). UAS in e- and v-commerce include conversational agents (Heßler et al. 2022), recommendation systems (Xiao and Benbasat 2007), and virtual assistants (Raut et al. 2023). In general, user assistance leverages analytics, data, and technology to help consumers make informed decisions about various aspects of their purchases. Examples of algorithmic help offerings include displaying the most relevant product reviews from other consumers (Pan and Zhang 2011) or assisting with interactive decision aids (Häubl and Trifts 2000; Pfeiffer 2011).

With the ability to collect data on neurophysiological responses in VR, new opportunities arise to create intelligent UAS that adapt to the individual's state. Machine learning (ML) plays a crucial role when building these new UAS as it provides the basis for an artificial intelligence (AI) steering the system. An intelligent, ML-based, adaptive system can learn about consumer search motives (Pfeiffer et al. 2020) using eye tracking (ET). Among latest VR headsets, the most common biosensors are ET cameras. For this reason, we utilize ET as the main neurophysiological sensor to detect visual attention and predict cognitive load. However, recent research-grade VR headsets offer further data sources, like electrocardiographic (ECG) sensors, and we forecast that a variety of different sensors will be available, as well as additional wearables. For example, electroencephalography (EEG) earbuds (Athavipach et al. 2019) for which a major tech company recently patented a design for.

One aspect that might help to create a good, highly personalized user experience (and therefore impact the success of these sales interactions) is the time when consumer assistance is invoked (Friemel et al. 2018). Adequate timing can influence consumers' attention (Bailey and Konstan 2006), perceived relevance, trust, urgency, and could be an enabler for UAS providers to beat the competition (Meurisch et al. 2020). Peukert et al. (2020) outlined how important it is to display a UAS with a good timing. They proposed a decision-phase-based detection algorithm and compared it with previously suggested decision phase models (Gidlöf et al. 2013; Russo and Leclerc 1994). However, they used simple gaze pattern rules to determine the phases, such as the first refixation on a product. A good timing to approach a consumer, however, depends on several factors, including their mental state (e.g., in the form of cognitive overload, personality, and habitual purchasing patterns). By carefully timing interactions, we claim that both consumers and providers can benefit due to the avoidance of dissonance between intended help offering and, in the worst case, perceived annoyance. While further previous work focused on assistance timing in generic software interface tasks, like finding appropriate software functionalities to alter an image (Ginon et al. 2016), this study is particularly geared towards the consumer decision-making context in v-commerce.

In this paper, we investigate cognitive load and its capability as an indicator to determine a good timing to engage with consumers in a shopping scenario. Previous work has identified cognitive load as a key mental state for decision-making (Deck and Jahedi 2015). In line with findings from the educational domain (Vaessen et al. 2014), we hypothesize that high levels of cognitive load can make it more difficult for consumers to understand and solve decision problems on their own, leading them to seek help (in form of an algorithmic support system or a digital human agent, i.e., a human sales representative controlling an avatar in the virtual shopping environment). Low levels of cognitive load might increase consumers' confidence and ability to solve problems independently, reducing the likelihood that they seek or appreciate help but rather want to browse the store independently. We argue that by estimating the cognitive load level during a consumer's purchase decision, it might be possible to determine a good timing to start an interaction. To account for varying levels of product knowledge, we employ two distinct products from two different categories: a fast-moving consumer good and a technology product. We expect differences between the product categories regarding desired help timing. Thus, the research questions read as follows:

1) When is the desired help timing for algorithmic user assistance compared to the desired help timing for a digital human agent in different shopping scenarios?
2) How does product knowledge influence the desired help timing?
3) Is desired help timing related to cognitive load and if yes, how can cognitive load be used to determine a good intervention timing?

We investigate these questions in an experimental VR environment, which gives our study particular relevance in the light of latest developments in the retail domain towards v-commerce. VR can improve consumer experiences (Moghaddasi et al. 2021) and offer high external validity while maintaining experimental control (Meißner et al. 2019). Furthermore, the used high-end VR headset allows us to collect gaze patterns and pupillometry in an unobtrusive and precise way. To answer our questions, we draw from two data sources. Both, ET and ECG, serve as an indicator of cognitive load (Haapalainen et al. 2010). This paper mainly builds upon two works. First, Peukert et al. (2020) have used ET to distinguish decision phases by using simple gaze patterns. These phases might indicate a good point in time when users seek help but a connection between decision phases and help timing was not investigated in their paper. Second, Pfeiffer et al. (2020) have estimated search motives based on fixations and their statistical moments. To complement the fixation data, we additionally include blinks, saccades, and pupillometry into the feature set. Additionally, we use ECG as secondary neuro-physiological sensor. We extend this existing stream of literature on consumer behavior in VR by focusing on the desired support type and good intervention timing.

Our contributions are twofold. First, we show that desired help timing depends on whether the help is provided by an algorithmic user assistance system or a digital human agent. The desired help timing also depends on the product category being purchased. As a result, when designing good shopping assistance, companies should be aware of this heterogeneity and strive for a high degree of personalization and context-awareness of the shopping situation. Second, we investigate cognitive load as an indicator to estimate the timing of assistance by using ET and ECG. The study demonstrates how ET and ECG can be used as features for shallow and interpretable ML models to predict optimal assistance offers. Overall, this article emphasizes the transformative nature of v-commerce and the high relevance of leveraging the recently available extended set of biosensors. We provide valuable practical guidance on how to approach the v-commerce transition and take advantage of the technological opportunities.

3.2 Related Work

3.2.1 Cognitive Load

The mental effort or capacity required to process and understand information is referred to as Cognitive Load (CL). Originating in psychology and education, Cognitive Load Theory (CLT) explains how the human brain processes information during learning and problem-solving (Plass et al. 2010; Sweller 2011). CLT suggests that humans have a limited amount of mental capacity (Miller 1956) and that the difficulty of a task can affect how much of this capacity it occupies. Furthermore, CLT can be applied to decision-making when choosing among several options (Deck and Jahedi 2015). For measuring CL, a variety of biosensors and ML techniques are available (Seitz and Maedche 2022). To minimize the negative impact of CL on decision-making, it is a viable option to simplify decision-making processes and reduce the amount of information that must be processed at a time (Todd and Benbasat 1994). Today's software solutions can reduce CL and improve decision making by providing help from a virtual agent (Brachten et al. 2020). Another option is breaking down overwhelming decision-making tasks into smaller, more manageable parts. Still, task optimization and atomization are no panacea. Even if the amount of options is limited, empirical results suggest that high CL levels can negatively impact the quality of decision-making (Allen et al. 2014; Dewitte et al. 2005). These studies consistently showed how a high CL level can lead to an increased likelihood of making errors in different task arrangements. Given this critical relation between CL and increased error rates, it is not surprising that marketing and shopping contexts are important domains to apply CLT (Schmutz et al. 2010; Grzyb et al. 2018; Wang et al. 2014). For example, a CLT-informed UAS can improve consumers' abilities to understand and process information about a product or service they consider buying. By reducing CL, product vendors can foster a positive shopping experience for their consumers. Building on the CLT principles, shop providers can actively design a UAS that increases their consumers' motivation and ability to seek help when needed. By making it easier for consumers to seek help when needed, or even offering the required help with perfect timing, companies can improve consumer satisfaction and reduce costs associated with providing assistance (Caruelle et al. 2023). Overall, CLT can provide a basis for understanding how different levels of CL influence consumers' motivation and ability to seek help. We hypothesize that after an initial exploration/orientation phase, consumers want to mitigate the imposed CL burden and value customer support. We further believe that CL can help to identify the moment when consumers engage with the product, viewing and comparing attributes or details. Such behavior indicates an increased likelihood of open questions. These questions could be answered by an algorithmic support system or a digital human agent.

3.2.2 Eye Tracking

Gaze patterns are suitable to track visual attention (Duchowski 2017), but their analysis relies on the eye-mind hypothesis by Just and Carpenter (1980), which assumes that human cognitive processes can be observed by their associated gaze patterns. However, it is evident that individuals can deliberately look at a certain position while thinking about something else (Anderson et al. 2004). Nonetheless, experimental findings indicate the validity of the eye-mind hypothesis in numerous scenarios (Holmqvist et al. 2011). Important movement-related gaze metrics are fixations and saccades. A fixation is a stationary state of the eyes and can last from milliseconds to seconds, while saccades are rapid eye movements between fixations.

Pupillometry investigates the changes in pupil dilation and frequently serves as estimator for CL (Kahneman 1973; Holmqvist et al. 2011; Hess and Polt 1964). In natural environments, pupillometry is not reliable for determining CL because small deviations in the lighting conditions have a strong impact on pupil dilation (Laeng et al. 2012). In a virtual environment, experienced by an individual using a VR headset, lighting confounds can be mitigated because the closed HMD cover offers fully controllable scene lighting.

3.2.3 Electrocardiography

ECG records the electrical activity of the heart, which emits a group of waves called PQRST (Goy 2013). Research has applied ECG to investigate various aspects of consumer behavior and is commonly used in combination with other biometric tools (Harris et al. 2018). Human-computer interaction research assesses additional factors, such as the usability of user interface design (Lee and Seo 2010) and emotional engagement with presented information (Ferdinando et al. 2016). ECG can serve as an indicator for CL (Haapalainen et al. 2010; Hughes et al. 2019). Data collection is typically performed with high frequency using electrodes that are attached to the skin.

3.2.4 Virtual Reality

In VR, the real-world environment is replaced as comprehensively as possible. A main goal of VR is to create realistic but completely virtual experiences with a high level of (tele-)presence for the users (Cummings and Bailenson 2016). An early head-mounted display (HMD), as it is common today, was already developed by Sutherland (1965). Another option to create virtual spaces is a CAVE automatic virtual environment (a recursive acronym), a cube-shaped room with projections on its walls (Cruz-Neira et al. 1992). Today, HMDs are common, and some models can even show mixed reality, which means everything on a spectrum from slightly augmented to fully immersive experiences. It is possible to combine an HMD with a variety of different sensors and cameras, particularly ET (Pfeiffer et al. 2020), which leads to many interesting research opportunities. Moreover, VR mitigates the trade-off between experimental control and ecological validity (Meißner et al. 2019).

VR has changed the landscape of v-commerce, ushering in a new era of immersive and personalized shopping experiences (Evans and Wurster 1999). The technology might transform the way consumers interact with products and purchase them online by providing a more engaging and lifelike representation. VR showrooms allow customers to view products in three dimensions, enabling a more informed decision-making process. In addition, VR has enhanced the social aspect of v-commerce through shared virtual spaces where friends or family can shop together and share opinions in real time (Zhang et al. 2014). A recent review by Branca et al. (2023) provides a comprehensive overview of different literature streams that address v-commerce. The authors identify four key research streams: customers, products, product testing, and VR compared to other conditions. As our study mainly focuses on desired help timing, it fits into the customer category. However, we propose to introduce a fifth label called sales agents which covers related research. We argue that the interface between provider and consumer is a key success factor which needs increased attention. Table 2 provides a list of selected previous customer behavior experiments in VR. It briefly describes the experimental setups, contributions, and allows the reader to understand the contribution and positioning of our article.

Table 2. Related VR experiment categorization.

Study	Setup	Contributions
Bigné et al. (2016)	N=41 CAVE ET data Spatial data Questionnaire	This study investigates brand preferences for fast food products and suggests that high attention to a brand and slow eye movements between brands lead to additional brand purchases. The applied method consists of regressions with aggregated parameters related to the entire decision-making process.
Martinez-Navarro et al. (2019)	N=178 HMD Questionnaire	The authors compare the effectiveness of different VR formats and devices. They find that virtual stores are more effective in generating cognitive and conative responses. They apply a structural equation model that suggests a dual path via brand recall and presence which both influence consumers' purchase intention in virtual stores.
Meißner et al. (2020)	N=132 HMD Questionnaire	This article compares high immersive (using a HMD) and low immersive shopping environments (using a desktop computer) and examines consumers' variety-seeking, price sensitivity, and choice satisfaction. In an incentive-aligned choice experiment, participants make repeated purchase decisions for cereal products. The statistical analysis suggests that consumers are less price sensitive and seek more variety in highly immersive environments.
Pfeiffer et al. (2020)	N=50 CAVE ET data Questionnaire	The authors investigate two classic shopping motives: goal-directed search and exploratory browsing. They compare decisions in a real-world supermarket with decisions in a virtual reality supermarket. They collect ET data on which they train three shallow ML models. They found that an ensemble method can classify the two motives with about 90% accuracy.
Alzayat and Lee (2021)	N = {48, 35} HMD Questionnaire	Using two VR stages and an Amazon mturk questionnaire, the authors investigate the differences in hedonic purchase value between a VR retail environment and a website. Their analysis comprises three structural equation models. The results suggest that a VR retail environment is more appropriate for products that are perceived as an extension of the body (e.g., tools) rather than a representation of the body (e.g., clothing).
Huang et al. (2021)	N=80 HMD Brain activity	This article focuses on search behavior, which is involved in the evaluation phase of each decision-making process. The authors investigate the congruence or incongruence between text and

	Questionnaire	color of flavor labels on product packaging. They provide evidence for a color-flavor incongruence effect in visual search and correlate it to the violation of user expectations. The method involves subsequent VR and fMRI phases, which the authors analyze using multiple regressions and regional homogeneity analyses, respectively.
Park and Kim (2021)	N = 196 HMD Questionnaire	This research examines how offering a virtual try-on in Augmented Reality, a 3D store on a desktop computer, and a VR store affect consumers' purchase intentions. The study also analyzes how thinking more deeply about an item influences the decision-making process in different shopping scenarios (searching versus browsing). Results indicate that purchase intentions are highest when participants browse in the VR condition. A moderated mediation analysis supports the hypothesis that cognitive elaboration mediates purchase intentions for those consumers in the browsing mode, while such a mediating effect was absent in the searching mode.
Schnack et al. (2021)	N = 36 HMD EEG data Spatial data Purchase data Questionnaire	This study compares instant teleportation with motion-tracked walking in VR and investigates whether different locomotion techniques correlate with altered shopping behavior. Using a split-sample experimental design, the authors apply electroencephalography (EEG) to track emotional states such as stress. In the scenario, participants experience a VR grocery store. Overall, the results suggest that different locomotion techniques have no impact on the consumers' emotional state and engagement. However, different spatial movement patterns are noticeable when comparing the different conditions.
Harz et al. (2022)	N = 210 HMD Questionnaire	The authors report on a combination of a real-world field study which is followed by a laboratory experiment. They examine how durable goods companies can use VR for new product development, and how VR can improve pre-launch sales forecasting. One of the three experimental conditions takes place in VR, the other conditions take place online and in the real world. The analysis of variances suggests that sales forecasting in VR provides the most accurate predictions compared to the other conditions. Moreover, it confirms the first evidence of the field study that VR correlates with a more consistent consumer behavior, and that virtual reality might create superior behavioral consistency compared to the real world.

Our work	N = 50 HMD ET data ECG data Questionnaire	In contrast to previous work, our study focuses on the desired help timing in a VR scenario for an algorithmic UAS versus a digital human agent. As second dimension, we compare technical products (3D printers) with fast-moving consumer goods (washing powders). We present the statistical analysis of our questionnaire and apply a machine learning approach to identify a good intervention timing. During our experiment, participants solve CL inducing tasks before making a purchase decision. ET and ECG provide the features for an ML classifier. Algorithmic help was requested particularly early for the washing powder. The results further indicate that CL-based classification works for the desired help timing of an algorithmic UAS but not for a digital human agent. The approach could be refined to invoke an AI agent based on a fine-tuned LLM, who has in-depth product knowledge.

3.3 Method

3.3.1 Experimental Design

The experimental setup was based on a virtual showroom in VR. Participants performed generic CL tasks of three difficulty levels and a subsequent purchase decision. The experiment focused on the utilitarian aspect of consumer behavior, as we asked participants to make decisions based on a set of criteria, leaving little room for their own hedonic motives. A web-based questionnaire on a desktop computer complemented the VR recordings. For all experiment sessions, we collected ET and ECG data.

To answer the first research question, we examined consumers' desired help timing for an algorithmic UAS versus a digital human agent. To identify potential differences across product categories, we used two product sets of four items. One set represented technology products (3D printers) and the other set represented fast-moving consumer goods (washing powders). We asked participants to identify good intervention timings for the two different types of help providers, an algorithmic UAS and a digital human agent, because participants might perceive relevant differences for these help providers. We argue that an algorithmic UAS may appear earlier during a decision-making process compared to a digital human agent because it is comparatively inexpensive. For the intervention of a digital human agent, timing is critical because it translates into substantial costs for human resources on the seller’s side. Providers should therefore be confident that an engagement is desired and that it takes place at the appropriate time.

As an exploratory aspect related to the first research question, we also wanted to identify the specific desired help type for algorithmic user assistance. In other words, do users prefer interactive decision aids, recommendations, or other algorithmic help types? This insight may guide practitioners in deciding which system type to implement in a certain scenario.

To investigate the second research question, we compared participants' product knowledge for the different product categories and examined its relationship with desired help times. We expected low product knowledge for the 3D printers because they are niche products, whereas a broad range of participants should be familiar with different washing powders. However, it was not clear what effect this (un-)familiarity would have on desired help times.

To control for possible confounding, we collected the participants' demographic information, personality traits, and their general attitude toward sales representatives. We also asked the participants about their product involvement but expected little difference because the monetary incentive for solving the purchase task was the same in both the washing powder and 3D printer scenarios.

To answer our third research question which aims to increase the understanding of CL in relation to the point in time when consumers want help, we measured CL levels that participants experienced when solving three generic tasks of low, middle, and high complexity before transitioning to the actual purchase task. To verify the difficulty levels, we controlled for subjectively perceived complexity during the generic tasks. Using the recorded ET and ECG data, we trained an XGBoost model to predict the CL level during a short period prior to the desired help timing.

All virtual scenes were implemented using the Unity 2021.3 game engine. Participants experienced our virtual environments using a Varjo VR 3 HMD with Valve Index controllers. This headset offered high-frequency ET capability with a sampling rate of up to 200 Hz, and its display resolution of 2880×2720 pixels per eye led to high visual immersion. The ET sensor was calibrated at the beginning of each experiment stage using a five-dot calibration protocol. For ECG recording, a wireless bioPLUX device captured signals throughout the experiment with a sampling rate of 1000 Hz. To be able to clarify possible confounds post-hoc during data analysis, we additionally recorded all experimental sessions on video using a room camera. Overall, the experiment followed a between-subjects design (regarding the two product categories) and included several questionnaire parts which alternated with the VR stages. Mandatory VR breaks for the questionnaires had additionally reduced the risk of cybersickness (Davis et al. 2014) and exhaustion of the participants within the VR environment.

3.3.2 Participants

Our self-hosted online registration platform (Bock et al. 2014) helped to recruit participants and manage the experiment sessions. Additionally, we actively solicited participation from students on our campus. Participation requirements were an age between 18 and 65 years and good command of English and German. Furthermore, we only accepted participants with normal or corrected-to-normal vision. Participation compensation was 10 Euros fixed plus a performance-based component of up to 5.5 Euros. After arriving at the lab, participants signed a consent form. It ensured the participants' basic knowledge of the experimental procedure, informed them that the experiment complied with ethical standards, and required them to grant the permission to publish their pseudonymized data as an open-source dataset.

3.3.3 Behavioral Measurements

We measured all questionnaire items on a 7-point Likert scale. In terms of demographics, we tracked participants' age, gender, and occupation. To estimate personality traits, we used the BFI-10 short scale (Rammstedt et al. 2013) which allows for the evaluation of personality traits with acceptable validity in a compact manner. We measured the general desire to interact with a salesperson using eight items validated by Lee and Dubinsky (2017). To collect self-assessments about CL, for both the multitasking-stage and decision-stage, we asked participants to answer the six item NASA Task Load Index (TLX) questionnaire (Hart and Staveland 1988; Hart 2006). Overall, four TLX batteries were collected per participant, one for each of the three generic CL task difficulty levels and one for the purchase decision. The product knowledge scale, consisting of three items, was adapted from Park and Moon (2003) to fit the presented products (see appendix). Moreover, the questions regarding participants' product involvement comprised 20 bipolar items (Zaichkowsky 1985).

3.3.4 Neuro-Physiological Measurements

To generate features for the ML model from the collected sensor data, we aggregated the raw ET and ECG recordings. The extracted features are listed in the supplemental Table 7 for ET features, and in supplemental Table 8 for ECG features.

For the ET data, we utilized both gaze-based metrics and pupillometry. Gaze events, namely fixations, saccades, and blinks were created using a velocity-based algorithmic approach (I-VT) as described by Salvucci and Goldberg (2000). For saccades, we set 50°/second as the lower angular speed threshold (Holmqvist et al. 2011). We limited fixation durations to 0.1 seconds as the lower threshold and 10 seconds as the upper threshold (Duchowski 2017). After creating the gaze events, we aggregated statistical moments to determine if attention was directed to different areas of interest (AOI, for example a product) and how often attention shifted between different AOIs. For pupillometry, we used the pupil-iris ratio of the dominant eye and complemented the gaze events with this information.

Using the raw ECG data, we extracted time- and frequency-domain-related features that covered different aspects of the heart rate and its variability (HRV) in linear and nonlinear representations (Xiong et al. 2020; Chanel et al. 2019; Pham et al. 2021). Regarding ECG feature selection, we rely on a recent review that covers the “most up-to-date and commonly used HRV indices” by Pham et al. (2021). Due to our relatively short task periods, some of the common HRV measures could not be investigated, such as the standard deviations of average heartbeat intervals (SDANN) which compare longer segments (by default 1, 2, and 5 minutes).

Overall, a crucial step for feature engineering was setting the time window size because it determined how the features were aggregated. For the ET related features, we evaluated 6 different window sizes (3, 5, 7, 10, 15, 30 seconds, where 30 seconds is the full trial duration) which yielded equally long segments without overlapping or artificial padding.

Further assumptions are necessary for the ET post-processing. An average fixation lasts about 0.3 seconds (Holmqvist et al. 2011) and average blinks and saccades are even shorter. Thus, we argue that 3 seconds yield enough data to calculate meaningful statistical moments in many cases. Considering increasing window sizes makes sense because CL might not be present from the onset of the task. Comparing different parts of a trial could yield a good contrast, such as the first versus the second half of a trial.

For ECG measurements, we only considered the full trial length (30-second windows). For shorter periods, only a limited set of features is computable, such as heart rate variability (HRV), while several features from the frequency domain and nonlinear domain suffer from numeric instabilities.

3.3.5 Procedure

The experiment lasted approximately 80 minutes, and it consisted of five different stages, as shown in Figure 7. The stages were streamlined with a web-based questionnaire on a desktop computer which alternated with the VR scenes and guided participants through the different stages from start to end. During the pre-stage, our participants completed an onboarding procedure and answered general questions. A multitasking-stage followed in which participants performed nine generic CL tasks (with three levels of complexity: easy, medium, and hard). A decision stage followed, in which participants made a product purchase to meet a list of given criteria. A video-analysis-stage followed during which participants retrospectively analyzed their first-person view during the purchase decision. A final post-stage, in which participants went through our offboarding procedure, concluded the experiment.

3.3.5.1 Pre-Stage

We assigned arriving participants randomly to one of two groups by flipping a coin and started the corresponding questionnaire on the computer. In the subsequent decision-stage, Group A was assigned to decide upon 3D printer products and Group B was assigned to washing powder products. A welcome screen explained the general purpose and modalities of the experiment. Before continuing, we asked the participant to read and sign our consent form. Only after accepting the terms of the experiment, participants were asked to provide demographic data, information about their personality traits, and to answer questions about their general attitude towards salespersons. Next, we determined their dominant eye using the Miles test (Miles 1929). For electrocardiographic data acquisition, we asked participants to go to the restroom and to attach electrodes to their body according to a reference picture, and to connect them to the transmitter. We decided to triangulate the heart in a wide triangle, spanning from the shoulders to the hip, to receive a high-quality signal that is robust to noise caused by body movements.

Next, we explained the VR hardware, controller usage, ET calibration procedure, and the upcoming task. Then we familiarized participants with movement, teleportation, and interaction using a training environment very similar to the subsequent task environments. The training scene consisted of the same showroom which was later used for the CL and decision environments. Participants were asked to use two in-world buttons which invoked the appearance of example models, one low-quality model with low polygon count and single-colored texture and one high-quality model with high polygon count and high-fidelity texture. Additionally, participants were asked to interact with a menu that started a timer and transitioned to the next stage after successful activation.

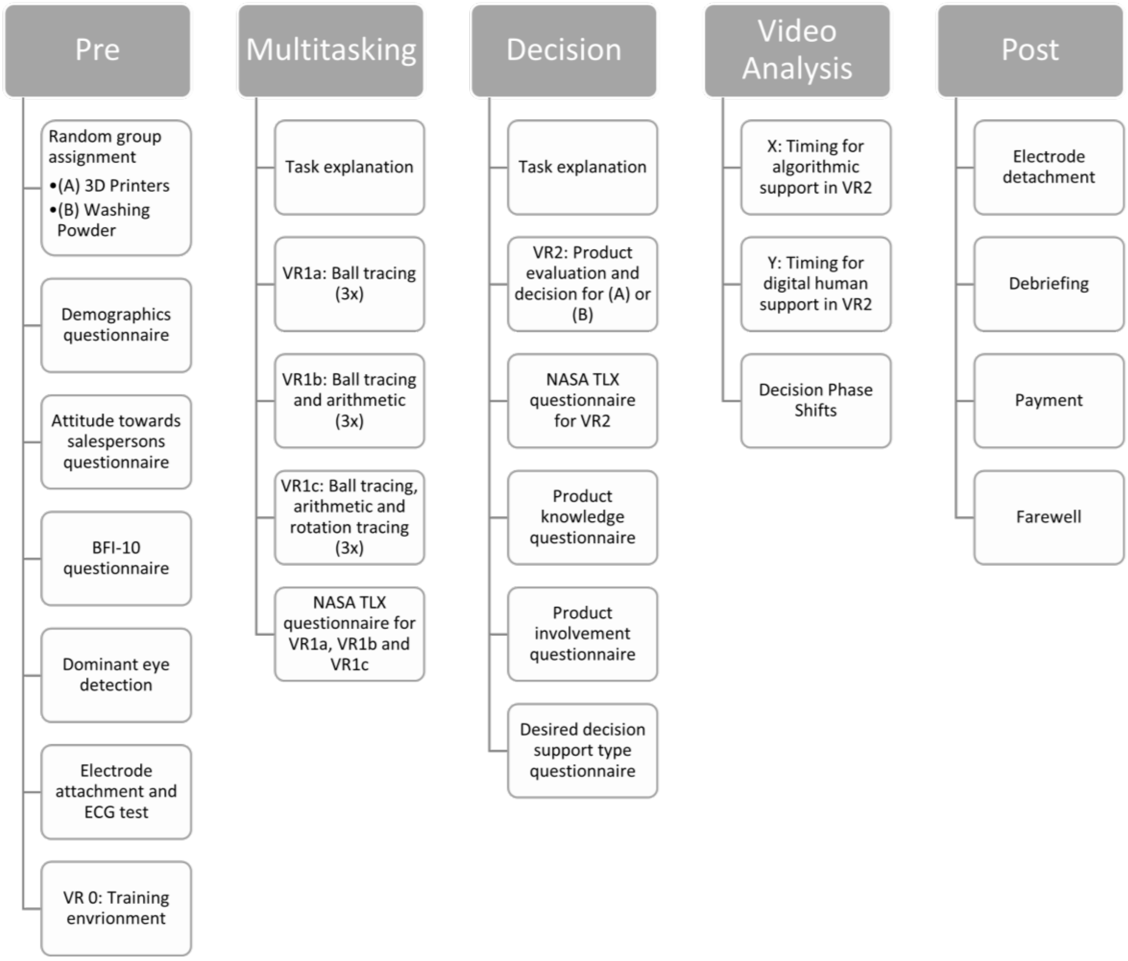

Figure 7. Experiment procedure.

3.3.5.2 Multitasking-Stage

To generate different generic CL levels, we designed a gamified CL task with three difficulty levels, as shown in Figure 8. This task was inspired by the work of Siegel et al. (2021). It consisted of three components – ball tracing, arithmetic, and rotation tracing. In the easy variant, participants had to trace one out of five moving balls. The target ball was colored red for 10 seconds. Afterwards, the trial began, and the target ball changed its' color to the same gray as the other four balls. All five balls moved around pseudo-randomly within a predefined area for 30 seconds. Finally, all balls stopped moving and displayed an identifying number. Participants then had to press a button labeled with the corresponding number to indicate which ball they considered as the target.

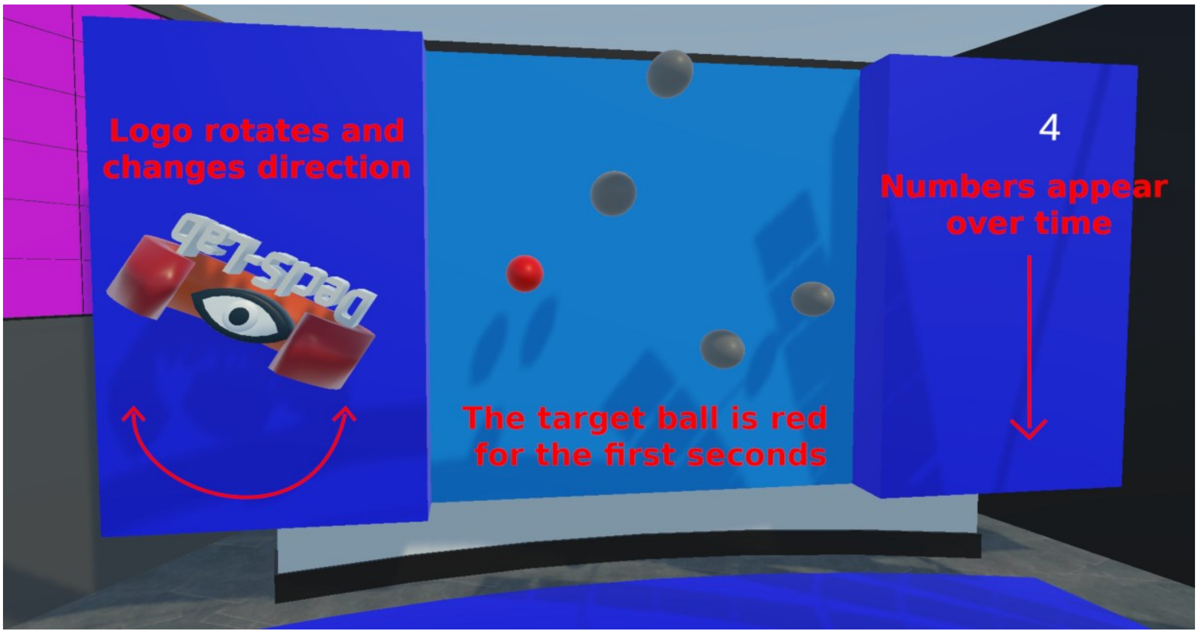

Figure 8. Multitasking VR environment.

A text message informed the participants whether the answer was correct or not, and the task was reset after a short waiting time. The medium variant was more difficult as it included the easy variant but additionally introduced an arithmetic component. To the right of the ball tracing area, small pseudo-random numbers (ranging from -10 to 10) appeared sequentially on the wall within a pseudo-random time interval and the participants had to aggregate them, while still tracing the ball in parallel. At the end of each trial, a slider was presented with which the calculated sum could be entered. An additional text message informed the participants whether the answer was correct or not. The hard variant was even more difficult as it included the medium variant but additionally introduced a rotation tracking component. To the left of the ball tracing area, a spinning logo appeared which changed its rotational direction between clockwise and counterclockwise within pseudo-random time intervals. Participants had to count the amount of rotational direction changes, in addition to the ball tracing and arithmetic components. After the 30 seconds of trial time, participants saw an additional slider to enter the counted number of rotational direction changes. All difficulty levels were repeated three times, and we incentivized the correct completion of each trial by increasing participants' performance-based extra payment by 0.5 Euro, if all components of a trial were answered correctly. Before these real trials started, all participants performed a training round in which they experienced the hard variant but without monetary incentive. During the training round, they could familiarize themselves with the task and ask questions. However, repetition was not possible. Then they began with the easy variant, followed by the medium and hard variant, until all nine trials were completed. Afterwards, participants took a VR break and continued the desktop-based questionnaire which sequentially asked for their perceived task difficulty for all three levels.

3.3.5.3 Decision-Stage

Both groups were presented with different task descriptions to create a realistic situation. Group A was asked to imagine being part of a board game designer team who needed a 3D printer to evaluate their game design. Group B was asked to imagine being a member of a residential community and being responsible for weekly grocery shopping which included buying washing powder (see appendix for the exact wordings of both task descriptions). To incentivize the decisions and increase external validity, participants had the chance to gain one additional Euro performance-based participation compensation if their product choice matched a previously determined team decision. This team decision was negotiated by a group of five individuals in advance of the experimental sessions.

In the virtual environment, participants first saw a blackboard containing the requirements specified by their imaginary peers. We designed these requirements so that the difficulty level matched among the groups (see supplemental Table 9). To this end, we chose three easy and three hard decision criteria. We considered attributes as easy that were obvious by looking at the product packaging or the product description from a distance. On the other hand, we considered attributes as hard for which participants either had to interact with the product (e.g., by starting or turning it) or needed further information to be able to judge the product. An example of a required interaction is that the print quality of a 3D printer could only be determined by pressing the print button and looking at the printed object. An example of a criterion which needed more information is whether a washing powder is environmentally friendly. This is because the roommates could have been looking for environmentally friendly packaging, environmentally friendly product ingredients, or both. We believe that external help could be strongly appreciated to clarify the requirements for both groups.

To begin the decision phase after memorizing the requirements, participants had to press a start button that concealed the requirements on the blackboard and displayed the products on a table behind them (refer to Figure 9 for Group A and Figure 10 for Group B). After this, participants could approach and engage with the products. To make their decision, participants of Group A had to choose the respective 3D printer name from a drop-down menu and click a purchase button while participants of Group B had to put the desired washing powder into a shopping cart next to the product table. After making a choice and detaching the HMD, participants continued to answer questions about their product knowledge, product involvement, task difficulty, and the preferred type of help for algorithmic user assistance (from a list of five common algorithmic user assistance types as shown in the appendix).

Figure 9. 3D printer decision VR environment.

Figure 10. Washing powder decision VR environment.

3.3.5.4 Video-Analysis-Stage

During this stage, participants answered time-related questions about their decision phase. Two questions regarding the desired timing for user assistance in the form of (X) an algorithmic UAS and (Y) a digital human agent (for exact wordings, see supplemental online material). These questions were displayed sequentially, and their order was randomized to avoid possible confounds induced by any static order. To find the corresponding timestamps, participants watched a video that showed their first-person view during the previous decision-stage and also displayed a gaze dot indicating their visual attention. Participants then selected the most appropriate moment for the assistance to appear and entered the corresponding timestamp in the questionnaire.

3.3.5.5 Post-Stage

We asked participants to go to the restroom and detach the ECG transmitter and electrodes. Then we continued with a debriefing (explanations about the experiment's purpose) and answered questions. Finally, we issued the participants' compensation and wished them farewell.

3.4 Results

The data analysis was performed in python 3.7 using neurokit2 0.2.3 (Makowski et al. 2021), scipy 1.7.3 (Virtanen et al. 2020), statsmodels 0.13.2 (Seabold and Perktold 2010), and pingouin 0.5.3 (Vallat 2018). ML was performed in python 3.10 using scikit-learn 1.0.2 (Pedregosa et al. 2011) and XGBoost 1.7.1 (Chen and Guestrin 2016).

3.4.1 Sample and Demographics

A total of 62 participants were observed resulting in 50 complete samples with 24 individuals in Group A (3D printers) and 26 individuals in Group B (washing powders). Regarding occupation, 49 of these 50 participants were students and one was a university staff member. Among the 12 discarded observations, one had to be excluded because of a recording interruption of the eye tracker during the decision phase. Another observation was excluded because the eye tracker was not able to calibrate, most likely due to a facial asymmetry of the participant. The remaining ten discarded observations had to be excluded due to ECG recording issues, particularly because of Bluetooth connection issues between the ECG transmitter and the host computer. The mean age of the remaining 50 participants (29 female and 21 male) was 24.5 years (SD = 4.9). Their average participation compensation amounted to 13.5 Euros (SD = 0.8).

3.4.2 Correlation of Neuro-physiological Features

We investigated correlations of ET and ECG metrics across different time windows for the different experimental periods. As expected, there were no significant correlations between the two sensors. The visualizations for the fixation duration over the different time windows are shown in the supplemental on top. Shorter intervals naturally show correlations with longer ones that comprise them. For example, the time windows from second 0 to 3 and from 3 to 6 overlap largely with the window from 0 to 5. This results in the red lines of high correlation in the fixation duration plot. While ET features were calculated for the interval lengths (3, 5, 7, 10, 15, 30), the ECG features only comprised the 30-second interval because shorter time windows would have been impractical for most HRV-based features. The bottom part of Figure 13 shows the correlations between the HRV features for this interval.

3.4.3 Attitude Towards Salespersons

To rule out possible confounds that could arise from different general attitudes toward salespersons, we asked the participants several questions before the actual purchase decision. The internal consistency of the general salesperson attitude scale was acceptable (Tavakol and Dennick 2011), measured by Cronbach's Alpha of $\alpha = .76$ and the mean rating was 4 ($SD = 1$) where a high rating corresponds with a high desire to interact with salespersons in general. A Shapiro-Wilk test indicated that the distribution of the mean rating did not depart significantly from normality ($W = 0.98$, $p = .73$), a Bartlett test indicated homoscedasticity ($T = 0.17$, $p = .68$), and a two-sample independent t-test did not indicate different means between the groups ($t = 0.7$, $p = .49$). Correlations with personality traits were determined via the BFI-10 scale (Rammstedt et al. 2013). It is plausible that agreeableness is significantly positively correlated ($r = 0.33$, $p = .02$) with a high desire to interact with salespersons.

3.4.4 Purchase Duration

The mean purchase duration (from pressing the start button to confirming the purchase) was 247.2 ($SD = 117.1$) seconds in total, and a normal distribution could not be assumed ($W = 0.94$, $p = .02$). The mean purchase time categorized by groups was 191.3 seconds ($SD = 85.1$) in Group A (3D printer) and 298.7 seconds ($SD = 120.2$) in Group B (washing powder). A Mann-Whitney U test indicated a significant difference between the groups ($U = 142.5$, $p < .01$). We see a reason for this difference in the fact that many participants interacted directly with the washing powder packages and regarded the product packages from all sides. For the printer decision, participants pressed the print button but rarely interacted with printed objects because they could visually judge the print quality without touching the objects.

3.4.5 RQ1: Desired Help Timing

We asked participants about (i) the desired help timing for an algorithmic UAS and (ii) the desired help timing for a digital human agent. As shown in Figure 11, an early appearance of the algorithmic UAS was particularly relevant for the fast-moving consumer good (FMCG). Reported mean values amounted to 125.5 seconds ($SD = 113.2$) for both help types (i and ii) combined, 103.1 seconds ($SD = 107.1$) for (i), and 148 seconds ($SD = 115.8$) for (ii). All values related to the duration after activating the start button which the participants pressed after memorizing the decision requirements on the blackboard. The mean difference (i) - (ii) for desired help timing between the two help providers (desired UAS timing - desired agent timing), was -44.9 seconds (SD = 123.3) for both groups, -12.3 seconds (SD = 81.4) for Group A and -75 seconds (SD = 147.4) for Group B. Multiple Wilcoxon signed rank tests (Wilcoxon 1992) for paired samples indicated that the difference (i) - (ii) for both product categories (W = 245, p = .02) and the difference (i) - (ii) for Group B (W = 39.5, p = .02) were significant, while the difference (i) - (ii) for Group A was not significant. Participants wanted help from an algorithmic UAS earlier than from a digital human agent, but this was mainly driven by the responses in Group B (washing powder). Overall, the differences in desired help timing showed the importance of investigating different product categories.

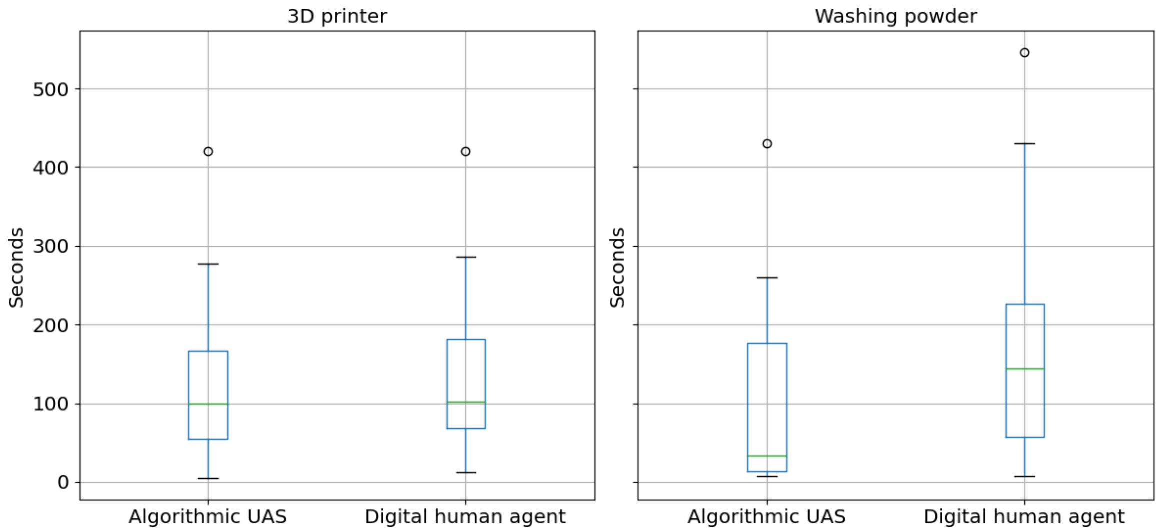

Figure 11. Desired help timing for algorithmic UAS and digital human consultant stratified by groups.

Regarding the most popular choices for algorithmic user assistance, 10 participants in Group A wished for reviews from other consumers and 14 participants in Group B wished for a product comparison matrix. Hiding irrelevant products and product feature highlighting were the least appreciated help types in both groups. Supplemental Figure 14 shows the complete distribution of the desired help types for an algorithmic user assistance.

3.4.6 RQ2: Influence of Knowledge on Help Timing

Internal consistency of the measured product knowledge items amounted to α = .76, which can be seen as acceptable (Tavakol and Dennick 2011). For the aggregated product knowledge measure, a normal distribution and homoscedasticity could be assumed. It amounted to 2.9 (SD = 1.4) for Group A, 4.2 (SD = 1.1) for Group B, and it significantly differed between the groups ($t = -3.37$, $p < .01$).

As a further control variable, we measured the participants' product involvement. For the respective items, a Cronbach's Alpha of $\alpha = .9$ indicated a very good consistency. Normal distribution and homoscedasticity could be assumed. The mean product involvement of 2.9 (SD = 1.4) for Group A and 4.2 (SD = 1.1) for Group B was not significantly different between the product categories ($t = 0.25$, $p = .8$). We expected such a similar product involvement for the different products, due to the equality in monetary incentivization for both groups.

Three linear regression (OLS) analyses provided further insight into whether product knowledge influenced desired help timings for different help providers. First, we considered only product knowledge and product category as independent variables and the absolute desired help timings as dependent variables (two separate OLS models for algorithmic UAS and digital human agent). For both help types, product knowledge had no significant linear association with desired help timings. Next, we investigated the same independent variables but used the difference between the desired help timings as dependent variable (algorithmic UAS help timing - digital human agent help timing). The respective OLS model showed that there was also no significant linear association between product knowledge and the difference in desired help timings. Finally, as a robustness check, we included our control variables and compared all three OLS models (desired help timing for the algorithmic UAS, digital human agent, and the timing difference between the two providers, see Table 2).

In all three constellations, there was no significant linear relationship between product knowledge and desired help timing. However, we did find a significant linear relationship between participants' openness and their desired help timing for an algorithmic UAS. Moreover, participants' age and extraversion showed significant linear associations with the desired help timing for a digital human agent. For the model that accounted for the timing difference between the help providers, the variables age, extraversion, and product involvement showed a significant linear association with the dependent variable.

Overall, we found no support for an influence of product knowledge on desired help timing. Instead, the OLS models suggested that age, personality traits, and product involvement influence desired help timing.

Table 3. OLS models of association between product knowledge and help timing.

	Model 1: *Algorithmic UAS timing*			Model 2: *Digital human agent timing*			Model 3: *Timing difference*		
	Coef.	**SE**	**p**	**Coef.**	**SE**	**p**	**Coef.**	**SE**	**p**
Knowledge	9.14	11.98	.45	10.79	11.6	.358	-1.65	13.26	.902
Involvement	5.47	9.61	.573	-16.64	9.32	.082	22.10	10.65	.045*
Sales Rep. Attitude	-6.73	16.72	.689	-7.71	16.2	.637	0.97	18.52	.958
Agreeableness	11.03	7.76	.164	12.65	7.5	.101	-1.63	8.60	.851
Conscientiousness	9.79	8.76	.271	-0.41	8.5	.962	10.20	9.71	0.3
Extraversion	-5.11	6.41	.43	14.82	6.2	.022*	-19.93	7.10	.008**
Openness	-22.12	8.78	.016*	-3.27	8.5	.703	-18.93	9.73	.059
Neuroticism	8.61	7.74	.273	-3.61	7.5	.633	12.23	8.57	.162
Age	3.88	3.34	.252	11.55	3.2	.001**	-7.66	3.70	.045*
Gender (Male)	-21.76	37.89	.569	-56.68	36.7	.131	34.92	41.97	.411
Group (B)	-39.00	35.33	.277	32.98	34.3	.342	-71.98	39.13	.074
Intercept	-13.51	167.85	.936	-199.01	162.7	.229	185.50	185.91	.325
R-squared	.30			.44			.354		

Note. *p < .05, **p < .01

3.4.7 RQ3: Cognitive Load Classification

3.4.7.1 Task Difficulties

We quantified the task difficulty of the generic CL tasks by counting the correct trials for each difficulty level (easy, medium, and hard). The correct completion rates were 146 out of 150 (97.3%) for the easy task, 131 out of 150 (87.3%) for the medium task, and 45 out of 150 (30%) for the hard task and a Kruskal-Wallis test indicated a significant difference between the medians ($H = 99.03$, $p < .01$). Using the NASA TLX questionnaire (Hart 2006), we measured how demanding our participants perceived the CL tasks and the purchase decision. Regarding the overall task load, a normal distribution could not be assumed for the easy task and the purchase decision (see supplemental Table 10). Therefore, we conducted a Kruskal-Wallis test that indicated significant differences between the three multitasking difficulty medians ($H = 65.14$, $p < .001$). Yet, due to the rather low internal consistency of the NASA TLX items (Cronbach's $\alpha < .7$, see supplemental Table 10), we considered only the single item concerning mental strain for further analyses (see supplemental Table 11). This single item also differed significantly between the tasks ($H = 83.73$, $p < .001$), suggesting that the three CL tasks evoked the desired low, medium, and high CL levels. Next, we tested which of the three CL task difficulty levels was most comparable to the purchase decision task. Pairwise Mann-Whitney U tests indicated significant differences for the tasks with easy and hard difficulty compared to the purchase decision, but this was not the case for the task with medium difficulty (see supplemental Table 12). The mean perceived task difficulty of the purchase decision was only 0.3 standard deviations less than the perceived medium task difficulty. Looking additionally at the box plots in the supplemental Figure 15, we interpret that, among the available options, the perceived difficulty of the purchase decision can best be matched to the perceived difficulty of the medium task. As a robustness check, we investigated the differences in perceived mental difficulty regarding the purchase decision between the groups. While the perceived mental difficulty in Group B exhibited less variance compared to Group A, we must assume equal mean difficulty between the groups, tested with a Mann-Whitney U test ($U = 368$, $p = .27$).

3.4.7.2 Machine Learning Model

To classify the CL tasks and desired help timings, we chose an 80% training and 20% testing split method. Instead of selecting a dedicated validation set, we applied a four-fold stratified cross-validation on the training set (Browne 2000). The optimization metric for classification was accuracy, while (multiclass) negative log-likelihood served as the loss function. Supplemental Table 6 shows the complete hyper parameter space. We used a randomized search approach on the hyper parameters to perform a lightweight tuning, limited to a maximum of 100 iterations. To interpret the feature importance, we used SHAP values (Lundberg and Lee 2017).

First, we solely investigated the generic multitasking difficulty levels. All participants performed three easy trials, three medium trials, and three hard trials for a duration of 30 seconds each. The best XGBoost model yielded a classification accuracy of .77 for the test set. This means that based on the ET and ECG measurements, we were able to predict with 77% accuracy whether a participant was performing the easy, medium, or hard task. Figure 12 shows the corresponding confusion matrix.

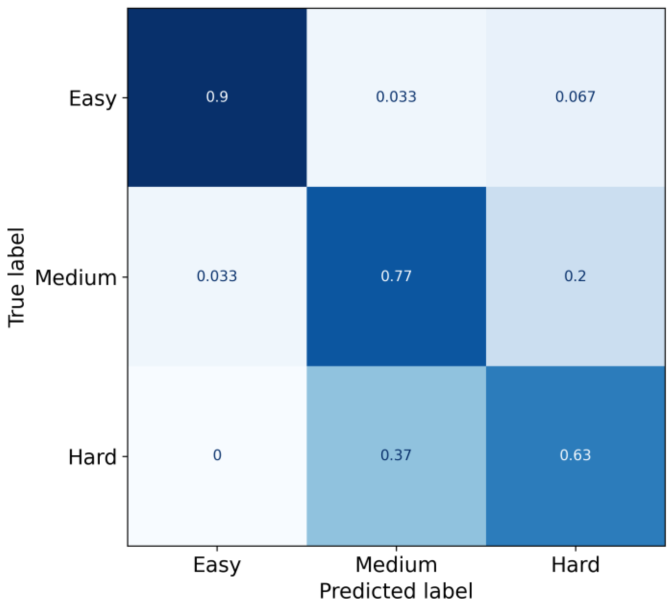

Figure 12. Confusion matrix for best multitasking classification model.

The easy task was classified with a high accuracy of .9 while the medium and hard tasks were not as clearly separable. Despite a correct classification rate of .77 for the medium and .63 for the hard CL levels, these tasks were frequently mutually misclassified. Nonetheless, the classification rates for these two classes were still clearly better than random guesses. A possible explanation for the misclassification between the medium and the hard tasks is the fact that 70% of the participants were unable to successfully complete the hard tasks. We observed that some participants only tracked two elements (the moving balls and appearing numbers) and ignored the additional spinning logo. Even though this strategy almost certainly resulted in an incorrect answer and no performance-based compensation for the respective round.

The mean absolute SHAP values, as shown in the supplemental Figure 16, represent the 20 most important features regarding the multitasking trials in the test set. Different saccade duration and angular speed related features were prominent (15 of the 20 most important features). This means the required time to jump between AOIs was most discriminative for the CL tasks. Overall, the most important feature was the saccadic mean duration for the whole 30-second periods. The number of uniquely fixated objects also played a role, as three features in this regard were among the 20 most important ones. Two blink-related features and one fixation-related feature were also present among them. Regarding the time window sizes, five features related to (3, 7, 15) second time spans, three features related to 30-

second time spans, and two features related to 5-second time spans. In our case, ECG and pupillometry features can be considered less important in discriminating between CL difficulty levels as they were not present among the 20 most important features. The best pupillometry feature was variance-related and ranked in 30th place. For ECG, the best feature was the HRV correlation dimension (HRV CD) for the whole trial duration (Bolea et al. 2014), a nonlinear measure for correlations within the signal which ranked in 51st place.

We applied the trained multitasking model to the purchase decisions and considered the intervals [t-30; t] prior to the indicated help timestamps t. Our intention was to identify the prevailing CL level shortly before help was requested. Choosing the same interval duration of 30 seconds allowed us to create the features analogously to the generic CL tasks. We classified each of the time spans as having either a low, medium, or high CL level. To compare one help interval with one respective non-help interval, we used the interval [t-60; t-30] as a non-help benchmark. For example, if a participant desired help two minutes after pressing the start button, we considered the data for the interval from timestamp 01:00 to 01:30 as the non-help benchmark and the data for the interval from timestamp 01:30 to 02:00 as the desired help timing period. For the desired timing periods of the algorithmic UAS, the model classified high (78%) and medium (12%) CL levels (see Table 4 for absolute counts and

Table 5 for classification probabilities). In comparison, most of the non-help benchmark intervals (96%) were classified as low CL level, and only 4% were classified as high CL level. For the desired timing of the digital human agent, the model classified 88% of the observations as high and 12% as medium. However, the benchmarks for these observations were also mostly classified as high (76%) and medium (18%) while only one observation (2%) was classified as low CL. This implies a difference in CL (an increase from low to high) during the 60 seconds before the algorithmic UAS was desired but no change in CL during the 60 seconds before a digital human agent should appear.

Table 4. Classification results (match help timespans to CL levels).

Help type	Low	Medium	High
Algorithmic UAS	0	11	39
Algorithmic UAS benchmark	48	0	2
Digital human agent	0	6	44
Digital human agent benchmark	1	11	38

Table 5. Average classification probabilities (match help timespans to CL levels).

Help type	Low	Medium	High
	P (SD)	P (SD)	P (SD)
Algorithmic UAS	.05 (.08)	.31 (.17)	.65 (.2)
Algorithmic UAS benchmark	.94 (.17)	.01 (.02)	.05 (.15)
Digital human agent	.03 (.04)	.28 (.16)	0.69 (.18)
Digital human agent benchmark	.05 (.1)	.31 (.18)	0.64 (.2)

3.5 Discussion

For our first research question, relevant insights emerged from the statistical analysis. We found that participants want help earlier from an algorithmic UAS than from a digital human agent. An early appearance of the algorithmic UAS was particularly relevant for the FMCG presented to Group B. The fact that a comparison matrix was the most desired algorithmic help type for the washing powders (see supplemental Figure 14) suggests that participants were primarily looking for ways to compare the product attributes efficiently. It is likely that they wanted to reduce extraneous CL induced by the rather unfamiliar VR environment. In contrast, when considering the 3D printer decisions, reviews from other consumers were the most desired algorithmic help type. Combined with the insignificant difference in desired help timing between the algorithmic UAS and the digital human agent when stratifying for Group A, it suggests that these participants were likely seeking help to cope with intrinsic CL.

Reviews were the second most desired help type. As a review by another consumer and an expressed opinion by a digital human agent are comparable, we claim that offering a digital human agent as help provider is more important for the technology product compared to the FMCG. This is further supported by the fact that 3D printers were the product for which our participants reported the least amount of product knowledge. For both groups, participants exhibited a certain reluctance to call for the digital human agent early in the process. A good idea could be to provide a digital human agent as optional help, in addition to algorithmic help types which are offered in the first place. Also, when considering non-binary choices for a certain help offering, our findings clearly highlight the need to customize timing and type of assistance offerings contingent on different scenarios and product categories.

Regarding the second research question, the experiment confirms significant differences in average product knowledge between the technical product and the FMCG. However, we did not find significant linear relationships between product knowledge and desired help timing for either of the two help providers (and not for the difference in desired help timing). When controlling for demographics, personality traits and product involvement, the respective OLS models indicate that participants' age, extraversion, openness, and product involvement have significant linear associations with desired help timings. The participants' age shows a strong positive linear association with the desired help timing for a digital human agent ($p = .001$, as shown in Table 3). The positive coefficient indicates that older participants wish to receive help from a digital human agent comparatively late (11.6 seconds per year). With increasing age, the difference (desired algorithmic UAS timing - desired digital human agent timing) between the desired help timing also decreases, but this effect is not as strong. Note that the product involvement is not significantly different between the product categories (likely due to the equal monetary incentivization) but displays a positive linear association with the difference between desired timings for the two help providers. More specifically, a one-unit increase on the 7-point Likert scale for product involvement corresponds to a 22.1-second increase in difference. Considering the product involvement coefficient for the timing of the digital human agent ($\beta = -16.64$, $p = .08$), we speculate that as product involvement increases, a digital human agent should appear earlier. To summarize, our OLS models suggest that product knowledge has a subordinate role with respect to desired help timings. Instead, demographic aspects and personality traits are likely to be more relevant. Product involvement could also play an important role, particularly in scenarios where the variance of product involvement is larger than in ours. In our experiment, we kept the variance in product involvement low by offering the same type of monetary incentive to solve both the 3D printer and the washing powder task.

The analysis of the ML classifications allows us to answer the third research question. Our results suggest that the 30-second periods before the desired help timings can be mapped with good accuracy to previously determined CL levels, even though the generic tasks were quite different compared to the purchase decisions. This is a promising result, as it suggests that further ML paradigms can potentially be trained with generic CL tasks that are quite different from the actual product decision. Regarding the input-features for the XGBoost model, saccade-based metrics were most relevant. Both saccadic angular velocity and saccade duration were highly discriminative. ECG measures were not among the 20 most important features, which suggests the superiority of the ET sensor over ECG for CL measurement, at least in our relatively brief scenario. As a supplemental data source, ECG can be useful to objectively measure CL, especially over an extended period.

For the 30-second intervals prior to the desired algorithmic UAS help timing, the ML model predicted medium and high CL levels but none of the observations were classified as low CL levels. In comparison, the model classified our benchmark interval (60 to 30 seconds prior to the desired help timing) mostly as low CL level. When considering the average class probabilities and their relatively low variances (see

Table 5), the benchmark and help intervals exhibit good separability. Overall, an adaptive intervention of an algorithmic UAS, which monitors changes in CL and automatically starts an interaction, seems possible.

Help timings for a digital human agent were also associated with a medium or high CL level. However, we did not find a significant change in CL levels compared to the respective baselines. The CL level is already medium or high during the baseline interval and does not change when help from a digital human agent is desired. Based on our findings, we argue that the CL level (at this likely later point in the decision-making process) should not be used as the sole indicator to inform a digital human agent about good intervention timing.

3.6 Conclusion

This study extends the consumer behavior literature in the emerging subfield of virtual commerce. Our statistical analysis investigates the desired help timings for two different product categories in detail and outlines the need for differentiated treatment. It also reveals behavioral and demographic factors which are linearly associated with desired help timing. Our study also provides information about the most desired algorithmic help types for different product categories.

Furthermore, we show how ET and ECG data can provide the features for a CL-based ML model which may benefit the consumer journey. The presented model indicates a good help timing for an algorithmic UAS while shopping for products or services in a v-commerce context. Even though, the ECG measures proved to be supplemental, our study still applies a larger number of ET features compared to previous studies. For instance, Peukert et al. (2020) used only one ET feature to detect decision phases and Pfeiffer et al. (2020) limited the number of predictors to four variables at a time.

In the v-commerce context, recognizing and reducing CL is applicable in many ways. Visual and other sensory aids can help to reduce CL and make it easier for consumers to understand information. Moreover, by personalizing a virtual environment, UAS can reduce CL and make it easier for consumers to perform their decision-making processes. CLT can provides twelve principles to break down complex information into smaller, more manageable parts and present it in a clear and concise manner. Our experiment suggests that an ML model can serve as indicator to invoke an algorithmic UAS which appears just-in-time and selectively provides the most relevant information to consumers. However, we believe that CL should not be used as the only criterion which determines the current consumer help seeking status. Instead, it should be included in multidimensional models to narrow down individualized help time spans for specific environments, products, and situations.

3.6.1 Theoretical Implications

With the proliferation of v-commerce, the emphasis in sales shifts towards providing consumers with a dynamic and interactive shopping experience. This increased attention to customer experience is driving providers to invest in innovative technology, such as AR and VR hardware, and the software to support it. The current rise of AI is likely to accelerate this trend even further, changing the rules for all kinds of retail activities. Our research gives answers to the question by Branca et al. (2023), who ask "What do we know and what do we not know about consumers' product evaluations in VR?". We complement previous research (i) by showing differences in desired help timings for different product categories, (ii) by identifying relevant impact factors on help timing, and (iii) by applying an extended set of sensors and features in an ML approach based on CLT. Our results show the feasibility of inferring CL from ET and ECG data, which then serves as a proxy for algorithmic UAS intervention. However, using CL as single predictor was not sufficient to determine a good point in time for a digital human agent.

Regarding the help type for an algorithmic UAS, our participants requested reviews and opinions of other customers most frequently. However, given the fake review problem (He et al. 2022) that currently prevails on several big e-commerce platforms, and combined with the rise of LLMs, we doubt that written messages or recorded videos will remain as compelling for consumers as they are today. On the second place were side-by-side product comparisons, which outline relevant and detailed information about products in tabular format. Taking CLT and Cognitive Fit Theory (Vessey 1991) into consideration, such a direct comparison might be feasible for a set of up to four products, which we deem a good maximum comparison capacity. Still, an optimal set size should be the object of further investigations.

The open research questions, such as good intervention timing for digital human agents, require combined efforts, methods, and theories from fields such as economics, neuroscience, and psychology. As sensors like EEG and functional near-infrared spectroscopy (fNIRS) become more precise while steadily shrinking in size and price, collaborative work can help to understand behavioral phenomena in the new context of immersive virtual domains. Applying new combinations of input features and incorporating further psychological effects such as flow (Berger et al. 2023) may also help to explain and model desired help timing and eventually allow for a better understanding of consumers.

3.6.2 Managerial Implications

We urge practitioners to embrace the challenges and opportunities which new virtual sales channels offer, sometimes even impose. Tech giants are racing for the next breakthrough device after the smartphone and consumers are wearing an increasing number of sensors that integrate into HMDs and additional wearables, such as wristwatches and earphones. Future shopping assistance will likely involve neurophysiological sensor data, apply ML, and be intelligent. Still, we believe that the human in the loop remains a crucial factor, for instance, as a digital human agent. Although delivered through an avatar, a genuine and actionable recommendation from a real person can still hold more trustworthiness than an automated suggestion from a recommender system (Castelo et al. 2019), particularly for contexts where the user wants to that to be seen by others, and to see themselves, as fully human (Heßler et al. 2022). However, LLMs are improving and a specialized model (in combination with further AI techniques) may soon allow for an intelligent, objective, and thus trustworthy AI sales agent that is perceived as very human-like (Seeger et al. 2021). Revolutionizing real-world call centers and drop-in stores, v-commerce industry pioneers should evaluate how a combination of basic UAS, LMM-based AI agents, and digital human agents may provide most value to the consumer experience.

With respect to ML, our described feature engineering process with different sensors and window sizes may inform how to create an appropriate inference pipeline for help timing. Our study provides a guideline on how to design a CL-based model that infers desired help timing for v-commerce customers. For practitioners with the capability to collect much larger samples than we had, we recommend evaluating time-series-based models. In our approach, we used a small dataset but with more data available, deep time series classifiers like InceptionTime (Ismail Fawaz et al. 2020) or TapNet (Zhang et al. 2020) might be suitable models to determine help-timing for a digital human agent. Providers could further combine it with an LLM-based AI agent, who has in-depth product knowledge. Overall, such a fine-tuned ML pipeline is likely to enhance customer experience, increase consumer engagement, and ultimately improve the likelihood of making a sale.

ET has proven to be an accurate sensor that provides both attentional and cognitive metrics. In contrast, we note that the ECG features only had a supplemental character for our study. In a brief period of 30 seconds, the heart rate is not as informative as the change in pupil dilation or the gaze duration for a certain product. While highlighting the key role of ET, we speculate about the impact of face tracking (FT) in our help timing prediction endeavors. Realistic synchronization of the cheeks, eyelids, and lips may help to improve the interaction between conversation partners. The next generation of wireless HMDs will integrate ET and FT because good animations and mapping of avatar movements is key in future virtual interactions, not only sales. Thus, incorporating FT seems like a logical next step.

V-commerce providers should consider ethical and privacy-related aspects, as the use of neurophysiological sensors raises many questions. To prevent privacy issues, inference could be done on the edge device itself, but this would be power-consuming and limited by the embedded processing unit. The European Union enforces special regulatory measures with the AI Act which could limit online data transfer for inference to a certain degree. However, these regulations are not yet established in detail and taking influence by means of close cooperation with the regulators seems advisable.

Overall, our paper suggests that a virtual showroom is a feasible virtual shopping platform for both FMCG and technical products. Still, we believe that it is not enough to copy prevailing real-world patterns and paradigms into virtual environments. For instance, as space is no constraint in VR, we see a classic shelf arrangement with very low and high product positions as obsolete. Practitioners should put increased effort into identifying and adhering to these new v-commerce rules, such as the need for adjusted ergonomic considerations (Wilson 1997). Our showroom gives one idea of how a v-commerce sales platform might look, but it is still very close to what is possible in the real world. Engaging VR room designs could go beyond physical limitations and incorporate interesting architectural features. These environments could further incorporate fun games (Tayal et al. 2022) and social activities (Gallace and Girondini 2022), which might act as ice-breaker between the consumer and the vendor.

Finally, we advocate for iterative processes when transitioning to virtual sales and help offerings. Our study also describes one part of an iterative research process. Further iterations will introduce the much-spoken-of avatar, and we also plan to evaluate a product comparison matrix UAS for commodity products.

3.6.3 Limitations and Future Research

The limitations of this study can also provide directions and advice for future research. A first concern is the generalizability of the results as the sample mainly consisted of students. Future research should involve a broader cross-section of society including different education levels, occupations, and age groups.

Second, future studies should increase the sample size because we were rarely able to assume a normal distribution for statistical testing. For future experiments, it would also make sense to include further product categories (e.g., beverages, food, interior) to obtain a better understanding of product-specific needs. Our results regarding the desired help type also suggest taking a closer investigation of comparison matrices as algorithmic user assistance. A convenient algorithmic UAS for product detail comparison was particularly desired in the FMCG group.

Third, immersion, perceived telepresence, and perceived product involvement could have been increased by adding more sensory channels (particularly audio) to the virtual environment. Future research could mitigate these aspects, e.g., by adding sound effects to the products. The room size also had a limiting impact on immersion and telepresence. On several occasions the experimenter had to interrupt participants and ask them to remain within the defined VR space. Subsequently, they were not able to fully immerse themselves in the virtual space. Future studies with a similar showroom setup should ensure to have at least 25 square meters of dedicated VR space.

Fourth, the quantitative approach with questionnaires leads to methodical issues like centrality tendencies and questionable consistency, especially for the NASA-TLX items (Hart 2006). Future studies could mitigate this issue by applying a mixed methods approach and by implementing and validating a more consistent mental difficulty scale.

Fifth, our CL-based ML model predicted help timing for an algorithmic UAS well but not for a digital human agent. However, we believe that it is feasible to create a predictive model for both help providers. There seem to be other influencing factors for the right intervention timing of digital human agents that our ET and ECG features do not cover. Furthermore, other model families, such as Hidden Markov Models (Rabiner 1989) or a deep learning time series classifier might be able to mitigate the issue and predict timings for both help providers. For a review on different time series classifiers, we refer to Ruiz et al. (2021).

Sixth, future experiments could improve the generic CL tasks or introduce another CL inducing design, such as a n-back task variant (Jaeggi et al. 2010). We performed the single generic CL task trials sequentially from easy to hard with individually chosen rest periods. Future research could consider a randomized setup with fixed rest periods (which might result in better classification results but bears a risk of reporting confounds regarding the task order). A broader range of CL tasks could also be considered, for instance tasks with auditory or haptic components or a classic n-back task setup. Furthermore, the period of 30 seconds for the CL tasks is too short for ECG measurements and should be revised for future research. It is also advisable to consider further sensors for CL, such as measuring galvanic skin response (GSR) and electroencephalographic (EEG) activity, which might be available for future VR devices off-the-shelf.

Future studies may provide deeper insights, for the good of both customers and service providers. New generations of highly immersive VR hardware allow for integrated and appealing experiments. We see the use of neurophysiological sensors in VR as a valuable methodology in experimental consumer behavior research and advocate for further exploration. It remains future work to find indicators for precise help demand prediction regarding a digital human agent. Different age groups and personality traits (like extraversion) may serve as further predictors, as our data has indicated. Incorporating additional neurophysiological aspects, such as emotions (Martinez-Navarro et al. 2019) and stress (Riedl 2012; Ishaque et al. 2020), is another step to increase the accuracy and generalizability of the ML model. Future research should particularly focus on the prediction of the moment when a digital human (or AI) agent should appear. Most probably, this point in time is more heterogeneously distributed among participants compared to the algorithmic UAS timing.

References

Allen, P. M., Edwards, J. A., Snyder, F. J., Makinson, K. A., & Hamby, D. M. (2014). The effect of cognitive load on decision making with graphically displayed uncertainty information. *Risk Analysis*, *34*(8), 1495–1505.

Alzayat, A., & Lee, S. H. M. (2021). Virtual products as an extension of my body: Exploring hedonic and utilitarian shopping value in a virtual reality retail environment. *Journal of Business Research*, *130*, 348–363.

Anderson, J. R., Bothell, D., & Douglass, S. (2004). Eye movements do not reflect retrieval processes: Limits of the eye-mind hypothesis. *Psychological Science*, *15(4)*, 225–231.

Athavipach, C., Pan-Ngum, S., & Israsena, P. (2019). A wearable in-ear EEG device for emotion monitoring. *Sensors*, *19*(18), 4014.

Bailey, B. P., & Konstan, J. A. (2006). On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in Human Behavior*, *22*(4), 685–708.

Berger, C., Knierim, M. T., Benke, I., Bartholomeyczik, K., & Weinhardt, C. (2023). InterFlowCeption: Foundations for Technological Enhancement of Interoception to Foster Flow States during Mental Work: About the potential of technologically supported body awareness to promote flow experiences during mental work. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*.

Bigné, E., Llinares, C., & Torrecilla, C. (2016). Elapsed time on first buying triggers brand choices within a category: A virtual reality-based study. *Journal of Business Research*, *69*(4), 1423–1427.

Bock, O., Baetge, I., & Nicklisch, A. (2014). hroot: Hamburg Registration and Organization Online Tool. *European Economic Review*, *71*, 117–120.

Bolea, J., Laguna, P., Remartia, Rovira, E., Navarro, A., & Bailón, R. (2014). Methodological framework for estimating the correlation dimension in HRV signals. *Computational and Mathematical Methods in Medicine*, *2014*.

Bourlakis, M., Papagiannidis, S., & Li, F. (2009). Retail spatial evolution: Paving the way from traditional to metaverse retailing. *Electronic Commerce Research*, *9*(1-2), 135–148.

Brachten, F., Brünker, F., Frick, N. R. J., Ross, B., & Stieglitz, S. (2020). On the ability of virtual agents to decrease cognitive load: an experimental study. *Information Systems and E-Business Management*, *18*, 187–207.

Branca, G., Marino, V., & Resciniti, R. (2023). How do consumers evaluate products in virtual reality? A literature review for a research agenda. *Spanish Journal of Marketing-ESIC*.

Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology*, *44*(1), 108–132.

Caruelle, D., Lervik-Olsen, L., & Gustafsson, A. (2023). The clock is ticking—Or is it? Customer satisfaction response to waiting shorter vs. longer than expected during a service encounter. *Journal of Retailing*.

Chanel, C. P. C., Wilson, M. D., & Scannella, S. (2019). Online ECG-based Features for Cognitive Load Assessment. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC).* IEEE.

Chen, B., & Yang, D.-N. (2022). User Recommendation in Social Metaverse with VR. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management.* ACM.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*.

Cruz-Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. V., & Hart, J. C. (1992). The CAVE: Audio visual experience automatic virtual environment. *Communications of the ACM*, *35*(6), 64–72.

Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology*, *19*(2), 272–309.

Davis, S., Nesbitt, K., & Nalivaiko, E. (2014). A Systematic Review of Cybersickness. In *Proceedings of the 2014 Conference on Interactive Entertainment.* ACM.

Deck, C., & Jahedi, S. (2015). The effect of cognitive load on economic decision making: A survey and new experiments. *European Economic Review*, *78*, 97–119.

Dewitte, S., Pandelaere, M., Briers, B., & Warlop, L. (2005). *Cognitive Load has Negative After Effects on Consumer Decision Making.*

Duchowski, A. T. (2017). *Eye tracking methodology: Theory and practice*. Springer.

Evans, P., & Wurster, T. S. (1999). Getting real about virtual commerce. *Harvard Business Review*, *77*, 84–98.

Ferdinando, H., Seppanen, T., & Alasaarela, E. (2016). Comparing features from ECG pattern and HRV analysis for emotion recognition system. In *2016 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB).* IEEE.

Friemel, C., Morana, S., Pfeiffer, J., & Maedche, A. (2018). On the Role of Users' Cognitive-Affective States for User Assistance Invocation. In *Information Systems and Neuroscience* (pp. 37–46). Springer, Cham.

Gallace, A., & Girondini, M. (2022). Social touch in virtual reality. *Current Opinion in Behavioral Sciences*, *43*, 249–254.

Gidlöf, K., Wallin, A., Dewhurst, R., & Holmqvist, K. (2013). Using Eye Tracking to Trace a Cognitive Process: Gaze Behaviour During Decision Making in a Natural Environment. *Journal of Eye Movement Research*, *6*(1).

Ginon, B., Stumpf, S., & Jean-Daubias, S. (2016). Towards the Right Assistance at the Right Time for Using Complex Interfaces. In *Proceedings of the International Working Conference on Advanced Visual Interfaces.* ACM.

Goy, J.-J. (2013). *Electrocardiography (ECG)*. Bentham Science Publishers.

Grzyb, T., Dolinski, D., & Kozłowska, A. (2018). Is Product Placement Really Worse Than Traditional Commercials? Cognitive Load and Recalling of Advertised Brands. *Frontiers in Psychology*, *9*, 1519.

Guo, G., & Elgendi, M. (2013). A New Recommender System for 3D E-Commerce: An EEG Based Approach. *Journal of Advanced Management Science*, *1*(1), 61–65.

Haapalainen, E., Kim, S [SeungJun], Forlizzi, J. F., & Dey, A. K. (2010). Psycho-physiological measures for assessing cognitive load. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*.

Harris, J. M., Ciorciari, J., & Gountas, J. (2018). Consumer neuroscience for marketing researchers. *Journal of Consumer Behaviour*, *17*(3), 239–252.

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting.* Symposium conducted at the meeting of Sage publications Sage CA: Los Angeles, CA.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology* (pp. 139–183). Elsevier.

Harz, N., Hohenberg, S., & Homburg, C. (2022). Virtual reality in new product development: Insights from prelaunch sales forecasting for durables. *Journal of Marketing*, *86*(3), 157–179.

Häubl, G., & Trifts, V. (2000). Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids. *Marketing Science*, *19*(1), 4–21.

He, S., Hollenbeck, B., & Proserpio, D. (2022). The market for fake reviews. *Marketing Science*, *41*(5), 896–921.

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, *143*(3611), 1190–1192.

Heßler, P. O., Pfeiffer, J., & Hafenbrädl, S. (2022). When Self-Humanization Leads to Algorithm Aversion. *Business & Information Systems Engineering*, *64*(3), 275–292.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

Huang, J., Zhao, P., & Wan, X. (2021). From brain variations to individual differences in the color-flavor incongruency effect: A combined virtual reality and resting-state fMRI study. *Journal of Business Research*, *123*, 604–612.

Hughes, A. M., Hancock, G. M., Marlow, S. L., Stowers, K., & Salas, E. (2019). Cardiac Measures of Cognitive Workload: A Meta-Analysis. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *61*(3), 393–414.

Ishaque, S., Rueda, A., Nguyen, B., Khan, N., & Krishnan, S. (2020). Physiological signal analysis and classification of stress from virtual reality video game. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC).* Symposium conducted at the meeting of IEEE.

Ismail Fawaz, H., Lucas, B., Forestier, G., Pelletier, C., Schmidt, D. F., Weber, J., Webb, G. I., Idoumghar, L., Muller, P.-A., & Petitjean, F. (2020). Inceptiontime: Finding alexnet for time series classification. *Data Mining and Knowledge Discovery*, *34*(6), 1936–1962.

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, *18*(4), 394–412.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354.

Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Citeseer.

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on Psychological Science*, *7*(1), 18–27.

Lee, H., & Seo, S. (2010). *A comparison and analysis of usability methods for web evaluation: the relationship between typical usability test and bio-signals characteristics (EEG, ECG).*

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*.

Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., Schölzel, C., & Chen, S. H. A. (2021). Neurokit2: A Python toolbox for neurophysiological signal processing. *Behavior Research Methods*, *53*(4), 1689–1696.

Martinez-Navarro, J., Bigné, E., Guixeres, J., Alcañiz, M., & Torrecilla, C. (2019). The influence of virtual reality in e-commerce. *Journal of Business Research*, *100*, 475–482.

Meißner, M., Pfeiffer, J., Peukert, C., Dietrich, H., & Pfeiffer, T. (2020). How virtual reality affects consumer choice. *Journal of Business Research*, *117*, 219–231.

Meißner, M., Pfeiffer, J., Pfeiffer, T., & Oppewal, H. (2019). Combining virtual reality and mobile eye tracking to provide a naturalistic experimental environment for shopper research. *Journal of Business Research*, *100*, 445–458.

Meurisch, C., Mihale-Wilson, C. A., Hawlitschek, A., Giger, F., Müller, F., Hinz, O., & Mühlhäuser, M. (2020). Exploring user expectations of proactive AI systems. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *4*(4), 1–22.

Miles, W. R. (1929). Ocular dominance demonstrated by unconscious sighting. *Journal of Experimental Psychology*, *12*(2), 113–126.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97.

Moghaddasi, M., Marín-Morales, J., Khatri, J., Guixeres, J., Chicchi Giglioli, I. A., & Alcañiz, M. (2021). Recognition of Customers' Impulsivity from Behavioral Patterns in Virtual Reality. *Applied Sciences*, *11*(10), 4399.

Morana, S., Pfeiffer, J., & Adam, M. T. P. (2020). User Assistance for Intelligent Systems. *Business & Information Systems Engineering*, *62*(3), 189–192.

Pan, Y., & Zhang, J. Q. (2011). Born Unequal: A Study of the Helpfulness of User-Generated Product Reviews. *Journal of Retailing*, *87*(4), 598–612.

Park, C.-W., & Moon, B.-J. (2003). The relationship between product involvement and product knowledge: Moderating roles of product type and product knowledge type. *Psychology & Marketing*, *20*(11), 977–997.

Park, H., & Kim, S [Seeun] (2021). Do augmented and virtual reality technologies increase consumers' purchase intentions? The role of cognitive elaboration and shopping goals. *Clothing and Textiles Research Journal*, *41*(2), 91–106.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., & others (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825–2830.

Peukert, C., Lechner, J., Pfeiffer, J., & Weinhardt, C. (2020). Intelligent Invocation: Towards Designing Context-Aware User Assistance Systems Based on Real-Time Eye Tracking Data Analysis. In *Information Systems and Neuroscience* (pp. 73–82). Springer, Cham.

Pfeiffer, J. (2011). *Interactive decision aids in e-commerce*. Springer Science & Business Media.

Pfeiffer, J., Pfeiffer, T., Meißner, M., & Weiß, E. (2020). Eye-Tracking-Based Classification of Information Search Behavior Using Machine Learning: Evidence from Experiments in Physical Shops and Virtual Reality Shopping Environments. *Information Systems Research*, *31*(3), 675–691.

Pham, T., Lau, Z., Chen, S. H., & Makowski, D. (2021). Heart rate variability in psychology: a review of HRV indices and an analysis tutorial. *Sensors*, *21*(12), 3998.

Plass, J. L., Moreno, R., & Brünken, R. (2010). *Cognitive load theory*. Cambridge University Press.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257–286.

Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C., & Kovaleva, A. (2013). A Short Scale for Assessing the Big Five Dimensions of Personality: 10 Item Big Five Inventory (BFI-10). *Methods, Data, Analyses*, *7*(2), 17.

Raut, A., Tiwari, A., Das, S., Saha, S., Maitra, A., Ramnani, R., & Sengupta, S. (2023). Reinforcing personalized persuasion in task-oriented virtual sales assistant. *PLOS ONE*, *18*(1).

Reinartz, W., Wiegand, N., & Imschloss, M. (2019). The impact of digital transformation on the retailing value chain. *International Journal of Research in Marketing*, *36*(3), 350–366.

Riedl, R. (2012). On the biology of technostress: literature review and research agenda. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, *44*(1), 18–55.

Ruiz, A. P., Flynn, M., Large, J., Middlehurst, M., & Bagnall, A. (2021). The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery*, *35*(2), 401–449.

Russo, J. E., & Leclerc, F. (1994). An Eye-Fixation Analysis of Choice Processes for Consumer Nondurables. *Journal of Consumer Research*, *21*(2), 274.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*.

Schmutz, P., Roth, S. P., Seckler, M., & Opwis, K. (2010). Designing product listing pages—Effects on sales and users' cognitive workload. *International Journal of Human-Computer Studies*, *68*(7), 423–431.

Schnack, A., Wright, M. J., & Holdershaw, J. L. (2021). Does the locomotion technique matter in an immersive virtual store environment?-Comparing motion-tracked walking and instant teleportation. *Journal of Retailing and Consumer Services*, *58*, 102266.

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with Python. In *Proceedings of the Python in Science Conference.* SciPy.

Seitz, J., & Maedche, A. (2022). Biosignal-Based Recognition of Cognitive Load: A Systematic Review of Public Datasets and Classifiers. *NeuroIS Retreat*, 35–52.

Shim, J. P., Warkentin, M., Courtney, J. F., Power, D. J., Sharda, R., & Carlsson, C. (2002). Past, present, and future of decision support technology. *Decision Support Systems*, *33*(2), 111–126.

Siegel, E. H., Wei, J., Gomes, A., Oliviera, M., Sundaramoorthy, P., Smathers, K., Vankipuram, M., Ghosh, S., Horii, H., Bailenson, J., & others. (2021). *HP Omnicept Cognitive Load Database (HPO-CLD)-Developing a Multimodal Inference Engine for Detecting Real-time Mental Workload in VR.* Technical report, HP Labs, Palo Alto.

Sutherland, I. E. (1965). The ultimate display. *Proceedings of the IFIP Congress*, *2*, 506–508.

Sweller, J. (2011). Cognitive load theory. In *Psychology of learning and motivation* (Vol. 55, pp. 37–76). Elsevier.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53.

Tayal, S., Rajagopal, K., & Mahajan, V. (2022). Virtual reality based metaverse of gamification. In *2022 6th International Conference on Computing Methodologies and Communication (ICCMC).* Symposium conducted at the meeting of IEEE.

Todd, P. A., & Benbasat, I. (1994). The influence of decision aids on choice strategies under conditions of high cognitive load. *IEEE Transactions on Systems, Man, and Cybernetics*, *24*(4), 537–547.

Vaessen, B. E., Prins, F. J., & Jeuring, J. (2014). University students' achievement goals and help-seeking strategies in an intelligent tutoring system. *Computers & Education*, *72*, 196–208.

Vallat, R. (2018). Pingouin: statistics in Python. *J. Open Source Softw.*, *3*(31), 1026.

Vessey, I. (1991). Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision Sciences*, *22*(2), 219–240.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., & others (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272.

Wang, Q., Yang, S., Liu, M., Cao, Z., & Ma, Q. (2014). An eye-tracking study of website complexity from cognitive load perspective. *Decision Support Systems*, *62*, 1–10.

Wilcoxon, F. (1992). *Individual comparisons by ranking methods*. Springer.

Wilson, J. R. (1997). Virtual environments and ergonomics: needs and opportunities. *Ergonomics*, *40*(10), 1057–1077.

Xiao, & Benbasat (2007). E-Commerce Product Recommendation Agents: Use, Characteristics, and Impact. *MIS Quarterly*, *31*(1), 137.

Xiong, R., Kong, F., Yang, X., Liu, G., & Wen, W. (2020). Pattern Recognition of Cognitive Load Using EEG and ECG Signals. *Sensors*, *20*(18), 5122.

Zaichkowsky, J. L. (1985). Measuring the Involvement Construct. *Journal of Consumer Research*, *12*(3), 341.

Zhang, H., Lu, Y., Gupta, S., & Zhao, L. (2014). What motivates customers to participate in social commerce? The impact of technological environments and virtual customer experiences. *Information & Management*, *51*(8), 1017–1030.

Zhang, X., Gao, Y., Lin, J., & Lu, C.-T. (2020). Tapnet: Multivariate time series classification with attentional prototypical network. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Zhang, Z., Shang, S., Kulkarni, S. R., & Hui, P. (2013). Improving augmented reality using recommender systems. In *Proceedings of the 7th ACM conference on Recommender systems.* ACM.

3.7 Supplemental Material

3.7.1 Cover Story 3D printer

You and a team of fellow students develop a board game idea. The team decides to put the idea into practice and builds a prototype. In a collaborative effort, you design the game pieces in a 3D software. Now you want to evaluate these models.

For the production, your team decides to purchase a 3D printer. However, an abundance of different printer variants exists. A company called 3D Print Workshop Inc. offers you the opportunity to evaluate their 3D printers in a virtual environment. Now, you put on VR glasses and enter the showroom of 3D Print Workshop Inc. In the virtual environment, you will see decision criteria which your team considers important. Furthermore, you will see several 3D printers with their properties.

Your task is to choose the right product.

3.7.2 Cover Story Washing Powder

You share your apartment with several roommates. It is your turn to do the grocery shopping and realize that the washing powder is empty. Therefore, you ask all your roommates to note down what kind of washing powder they prefer. Of course, you are not going to the supermarket in real world. Instead, you use your VR headset and order the product in Virtual Reality. In the virtual environment, you will see decision criteria which your roommates consider important.

Your task is to choose the right product.

3.7.3 Choices for Desired Help Types

Product comparison matrix

Product recommendations

Reviews of other consumers

Product feature highlighting

Hiding irrelevant products

3.7.4 Help Timing Questions

The experimenter will show you a video of your purchase decision. While watching the video, please determine when you would have appreciated help during the task ((X) by a digital human consultant in the VR environment / (Y) (by an algorithm in the VR environment)). After watching the video, please answer the questions below.

(X) What time would be the best moment for an **algorithmic** decision support to appear?

(Y) What time would be the best moment for a **digital human consultant** to appear?

3.7.5 Product Knowledge Items

1) Compared to others, how familiar do you think you are with the product?

2) Do you know precisely what attributes of the product decide the function of the product?

3) Do you think you can make a satisfactory purchase of the product based on only your own knowledge, without another person's help?

7-point Likert Scale: 1 – absolutely not; 7 – absolutely yes

3.7.6 Supplemental Figures

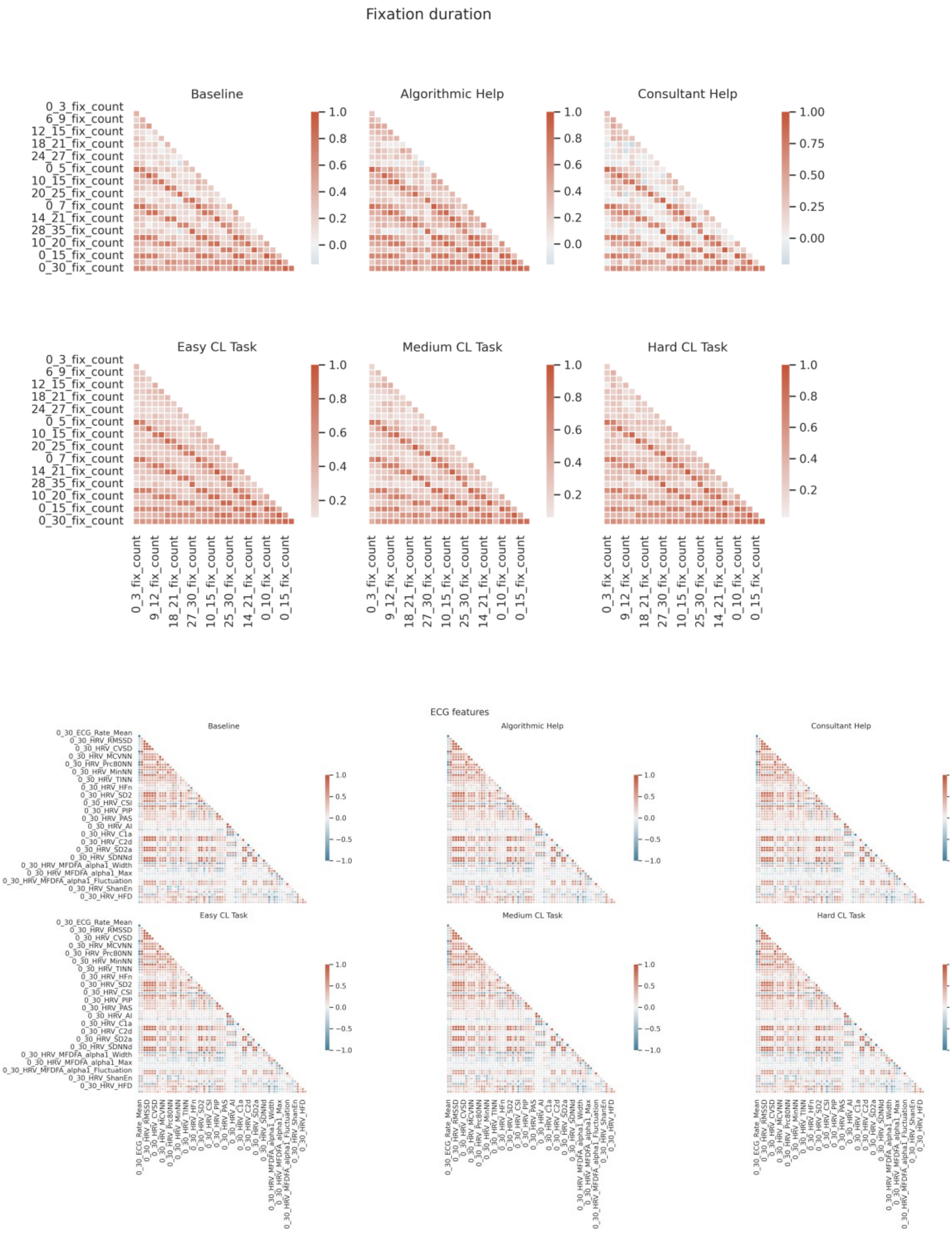

Figure 13. Correlation analysis for ET and ECG features.

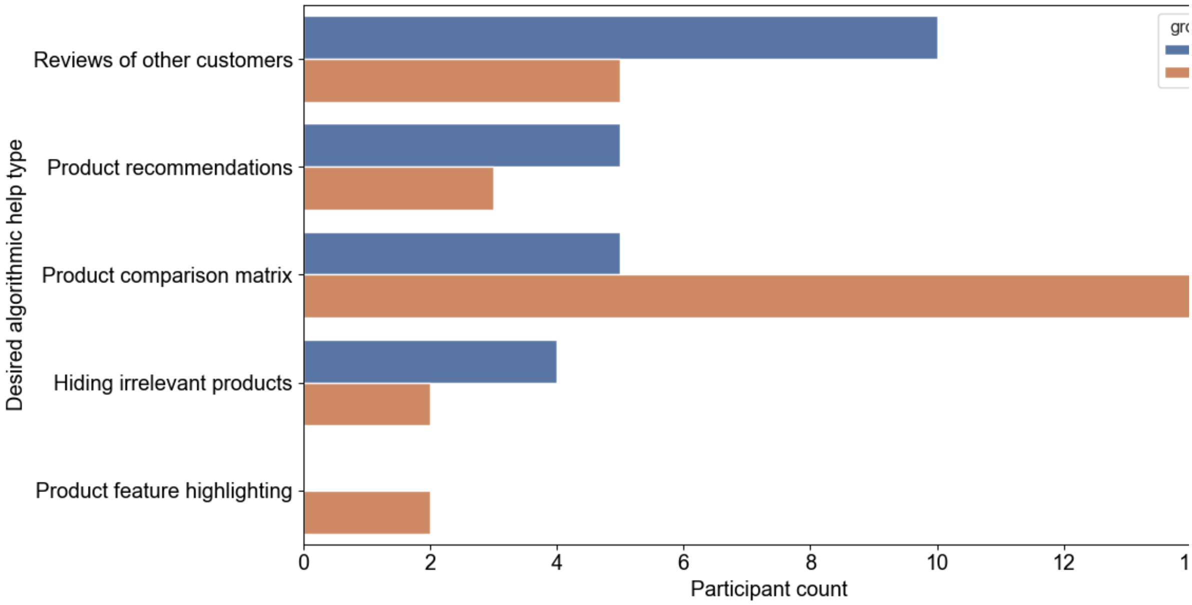

Figure 14. Desired algorithmic UAS help type stratified by groups.

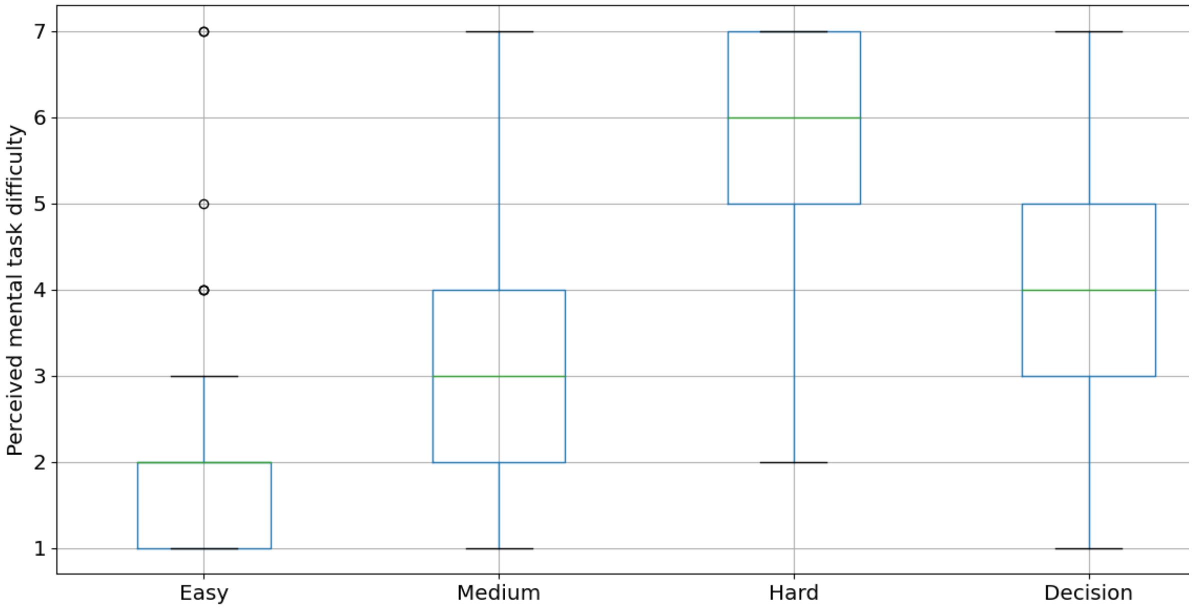

Figure 15. Boxplot showing the mental strain among the different tasks.

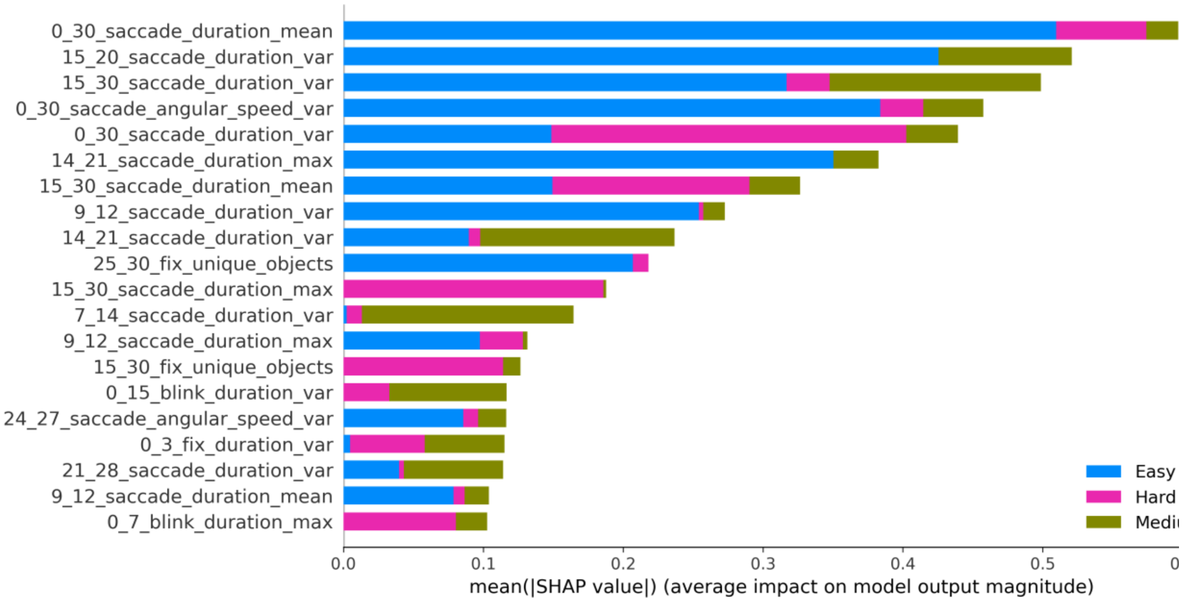

Figure 16. Mean absolute SHAP values explaining the multitasking classification model for the test set.

3.7.7 Supplemental Tables

Table 6. Hyper parameter space for the XGBoost model.

Parameter	Values
colsample_bytree	[0.6, 0.7, 0.8, 0.9]
gamma	[0, 0.1, 0.2, 0.3, 0.4]
learning_rate	[0.0005, 0.001, 0.005, 0.01, 0.05, .1, 0.5]
max_depth	[3, 5, 7, 9, 11, 13, 15, 17, 19]
min_child_weight	[1, 3, 4, 7, 9]
n_estimators	[25, 50, 75, 100, 125, 150, 175, 200]
reg_alpha	[0, 0.001, 0.005, 0.01, 0.05]
subsample	[0.6, 0.7, 0.8, 0.9]

Table 7. Eye tracking feature overview.

Feature*	Comment
Blink count	
Blink duration	Max, Mean, Var
Fixation count	
Fixation duration	Max, Mean, Var
Saccade count	
Saccade duration	Max, Mean, Var
Saccadic angular speed	Min, Max, Mean, Var
Dominant pupil iris ratio	Min, Max, Mean, Var
Unique object count	

** ET features were calculated for the different time windows 3, 5, 7, 10, 15, and 30 sec.*

Table 8. ECG feature overview.

Feature*	Comment
Time domain	
NN	Min, Max, Mean, Median, MAD
SDNN, SDSD, RMSSD, Prc20NN, Prc80NN, pNN20, pNN50, HTI, TTIN	
Frequency domain	
HF, VHF, LnHF, HFn, LFn	
Time-frequency domain	
STFT, WT, WVD, SWVD	
Nonlinear domain	
SD1, SD2, SD1/SD2, S, C, C1, C2, CSI, CVI, CSI_Modified, PIP, CD, HFD, KFD, LZC, CVNN, CVSD, MCVNN, IQRNN, IALS, PSS, PAS, GI, SI, AI, PI, DFA_alpha1, ApEn, ShanEn, FuzzyEn	
MFDFA_alpha1	Width, Peak, Mean, Max, Delta, Asymmetry, Fluctuation, Increment

** ECG features were only calculated for the full 30 second time windows.*

Table 9. Product criteria classification.

Category	(A) 3D printer	(B) Washing powder
Rather easy	- Easy device setup - Large model print size - PETG material printable	- Allows for 100 washing cycles - Conserve colors - Efficient dirt removal
Rather hard	- Fast print speed - High print quality - The device should not catch fire	- Environmentally friendly - Allow high washing temperature - Little powder amount per wash cycle

Table 10. Overall task difficulty and test results for normality and for consistency among NASA TLX items.

Task	Mean	SD	Shapiro-Wilk W	Shapiro-Wilk p	Cronbach's α
Easy	2.89	0.727	.897	<.01**	.629
Medium	3.60	0.767	.974	.345	.574
Hard	4.57	0.910	.967	.169	.582
Decision	2.98	0.972	.946	.024*	.699

Note. *p < .05, **p < .01

Table 11. Mental task difficulty descriptive analysis and test results for normality.

Task	Mean	SD	Shapiro-Wilk W	Shapiro-Wilk p
Easy	2.14	1.385	.740	<.01**
Medium	3.46	1.343	.913	<.01**
Hard	5.74	1.226	.849	<.01**
Decision	3.86	1.539	.939	.012*

Note. *p < .05, **p < .01

Table 12. Pairwise task comparison for differences in mental difficulty.

Task 1	Task 2	Mann-W. U	Mann-Whitney p (Bonf. corrected)	SD
Hard	Medium	2201.5	<.01**	1.773
Hard	Easy	2364.5	<.01**	2.752
Hard	Decision	2073	<.01**	1.351
Medium	Easy	1976	<.01**	0.967
Medium	Decision	1032.5	.76	-0.277
Easy	Decision	483.5	<.01**	-1.175

Note. *p < .05, **p < .01

4 Paper C: Customer Decision-Making Processes Revisited: Insights from an Eye Tracking and ECG Study using a Hidden Markov Model

Tobias Weiß, Lukas Merkl, and Jella Pfeiffer

Abstract

Good timing is key for many activities in business and society. In the context of adaptive user assistance, it can work as door opener to further engage with the user. This paper presents a virtual commerce study which combines eye tracking, electrocardiography, and virtual reality with the goal to detect decision phases in two different purchase scenarios. We therefore collect objective sensor data in combination with subjective decision phase annotations. Shifts between decision phases are determined subjectively by the participants via retrospective video analysis. For decision phase recognition, we demonstrate how to use the neurophysiological sensor data to train a Hidden Markov Model with multivariate mixed Gaussian emission distributions and how to use it for inference. A main benefit of our approach is its lightweight character regarding both training and inference.

Keywords: Customer Behavior, Decision-making, Eye Tracking, Electrocardiography, Hidden Markov Model, Gaussian Mixture Model, Machine Learning, Virtual Commerce, Virtual Reality.

4.1 Introduction

Approaching customers at the right time is crucial because it can significantly impact the interaction success (Sykes 2015). Specifically, good timing can help to maximize engagement, build trust, and increase conversion rates (Friemel et al. 2018). However, to determine the right point in time to approach a customer requires profound understanding of the target audience's behavior and preferences (Horvitz et al. 2013). Advances in conversational agents and user assistance systems (UAS) often focus on the right information, introduce context-awareness and improve interactivity (Maedche et al. 2016; Pfeiffer 2011; Sykes 2015). Rather scarcely, previous research has investigated invocation timing based on neurophysiological indicators (Peukert et al. 2020). Within the ongoing transformation of the retail sector towards virtual commerce (Bourlakis et al. 2009) and the rise of the metaverse idea (Ball

2022), good invocation timing is one of the key components for a variety of information system (IS) artifacts. Decades ago, metaverse and virtual reality (VR) advocates already envisioned that a large fraction of daily life and therewith a large fraction of shopping activities transfers to virtual spaces (Lanier and Biocca 1992; Stephenson 2003). Today, this process gains momentum, as big tech companies introduce new hardware and applications with rigorous commitment. Latest VR headsets ship with eye and face tracking technology which fosters the potential and feasibility of neurophysiological IS and therefore turns them into a game changer. With a VR headset on their head, future customers wear a variety of sensors in proximity to the most reliable information source about their attitudes and moods. In this paper, we present our approach to integrate neuroscientific methods into virtual commerce IS. Our research question states as follows:

RQ: Can we determine a good timing to approach customers in a virtual commerce scenario using eye tracking and electrocardiography?

We report our insights gained from a study in which 50 participants had to make purchase decisions for either washing powder or 3D printers while wearing a head-mounted VR headset. We collected participants' eye tracking (ET) data, electrocardiography (ECG) data, and created a prediction model that can distinguish between different decision phases. Our insight can be used to inform a UAS or digital human agent when help is wanted. As model for decision phase recognition, we chose a combination of multivariate Gaussian Mixed Model and Hidden Markov Model (GMM-HMM). The benefit of our approach is its lightweight character in both training and inference. Thus, the presented GMM-HMM approach offers itself as good candidate to make it into soon-to-be released virtual commerce (and other) neurophysiological sensor-based IS artifacts (vom Brocke et al. 2013). To the best of our knowledge, no study exists which applies machine learning approaches to differentiate between different decision phases using neurophysiological sensor data. Our research builds up on previous models but tries to apply a more generic inference method not solely dependent on product comparisons and re-dwells.

4.2 Method

4.2.1 Consumer Decision-Making

Several scholars in consumer behavior research suggested models to subdivide customer decision-making processes into different phases. Most studies support a phase theory which consists at least of an orientation and an evaluation phase. One prominent phase model is the five-stage Engel Kollat Blackwell (EKB) model (Engel et al. 1968), as shown in Figure 17. The EKB model is still widely accepted (Sihi 2018) and frequently serves as basis for further adjustment to integrate specific aspects and research field dependent needs, such as modifications for an eye tracking study in VR.

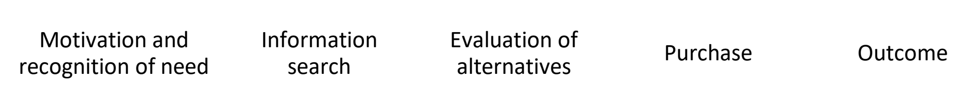

Figure 17. The EKB model, dividing customer decision processes into five phases (Engel et al. 1968).

In an eye tracking context, several other decision phase models were developed, e.g., by Russo and Leclerc (1994), Gidlöf et al. (2013), and Peukert et al. (2020). These models subdivide decision processes into three phases – orientation, evaluation, and validation. The transition between different phases is based on simple rules, like re-fixations on products. The VR study in Peukert et al. (2020) pursued an on-the-fly attempt to determine the phases. Its authors used eye tracking data and identified the first comparison between two products as shift between orientation and evaluation. Furthermore, the shift between evaluation and verification was considered as the moment when the first product entered the shopping cart (We believe this is a questionable criterion because putting a product into the shopping cart signals a certain level of confidence).

For the right timing of user assistance, we consider the shift between orientation and evaluation as particularly interesting. We conjecture that help is most appreciated by customers after being within the evaluation phase for a certain offset duration. To verify this assumption empirically, self-reported desired help timings can be used. Knowing the phase of a decision process and the offset duration at least approximately, a UAS or sales representative can determine a good starting point to approach the customer.

4.2.2 Neurophysiological Data Collection in VR

The development of visual VR has a longer history than one might expect. For example, an early head mounted display (HMD) was already developed by Sutherland (1965). Commercial endeavors of big tech companies still focus on HMD development. For research, the latest HMD generation is particularly interesting because many models ship with integrated neurophysiological sensors, particularly ET (Pfeiffer et al. 2020). ET is integrated because it can be used to optimize graphic card utilization via foveated rendering, a method which only renders the focused area in high detail while neglecting peripheral areas (Patney et al. 2016). Recent research-grade HMDs include further sensors as ECG and electroencephalography (EEG). The integration of EEG into consumer-grade hardware seems rather unrealistic in the near and intermediate-term future as the sensor itself is expensive and the electrodes are relatively uncomfortable to wear. ECG measures a person's heart rate and is more likely to find its way into consumer devices. Another sensor, which is very likely to be included into future customer-grade HMDs, is photoplethysmography (PPG). PPG is a light-based sensor which can also be used to measure heart rate and corresponding metrics. Compared to ECG, PPG is cheaper, easier to attach (e.g., a forehead-sensor integrated in the HMD-cover), but less accurate. It is also imaginable to couple wearables with an HMD, particularly fitness watches, which already include ECG or PPG sensors. Overall, ET and ECG/PPG are the most likely sensors for future off-the-shelf HMDs. Thus, it makes sense to use gaze patterns, pupillometry, and heart rate as data sources for inference.

4.2.3 Hidden Markov Model

An HMM is a statistical model which describes a Markov process with a set of states between which it can transition (Rabiner and Juang 1986; Eddy 2004). At each state, an HMM generates an observation or output symbol, which is associated with that state. Such observations generated by a state of the model are referred to as emissions. HMMs find application in a variety of disciplines (Liu et al. 2023; Krogh et al. 1994; Schultz and Waibel 2001). To match the characteristics of our purchase decision scenario in the experimental VR setup, we use elements of both the classic EKB phase model (Engel et al. 1968) and the eye tracking model proposed by Russo and Leclerc (1994). We begin with a memorization phase which corresponds to the motivation phase of the EKB model. During this phase, participants see purchase criteria on a blackboard and memorize them. The transition between memorization and the next phase is identified by a button press. For the subsequent phases, we use the phase labels orientation, evaluation and verification as proposed by Russo and Leclerc (1994). However, we outline that the state transitions in our model have nothing in common with the originally proposed transitions which were based on specific gaze patterns. Instead, shifts to evaluation and verification were determined via self-reported timestamps given by the participants. Next, we adopt the purchase phase from the EKB model, as participants remained inside the VR scenario after confirming the purchase. Furthermore, an initial and terminal state are added as they are needed for computation. The corresponding HMM with flat prior transition probabilities is shown in Figure 18. GMM-HMM with flat prior transition probabilities..

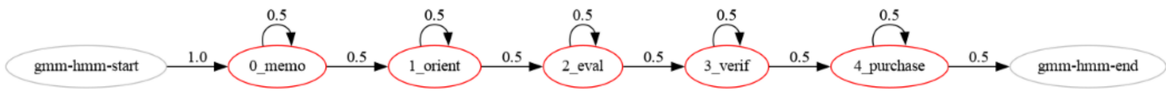

Figure 18. GMM-HMM with flat prior transition probabilities.

When the model transitions from one state to another, it refers to a (hidden) multivariate probability distribution which corresponds to the current input features. Internally, each state holds a multivariate Gaussian mixture distribution (what turns the model into a GMM-HMM), which is trained with ET and ECG features based on consecutive five second time windows. For each of these windows, our feature engineering pipeline creates 44 features which comprise 26 ET and 18 ECG features. ET features consist of statistical moments (mean, min, max, var) for blinks, fixations, fixation duration, pupil size, saccadic duration, and saccadic speed. ECG features are limited to the time domain, particularly the heart rate and its variability. Frequency domain related and non-linear ECG features are not considered because they would require longer window durations (Pham et al. 2021). If participants indicate a state transition during such a window, the label for the subsequent and all following windows changes to the next state.

For real-time inference, the GMM-HMM can even be simplified to a GMM classifier which decides if the evaluation phase is reached or not. Features of a current observation can be shown to the model which maps them to the probability distribution and stochastically decides whether the evaluation phase is reached or not. If the evaluation phase is indicated several times in a row, the offset of approximately fifty seconds could be added and finally the UAS or digital human agent could approach the customer with a help offering.

4.3 Experiment

4.3.1 Participants

Our sample was collected from 50 participants (29 female, mostly students) with a mean age of 24.5 years (SD=4.89). Only individuals with normal or corrected-to-normal vision via contact lenses were accepted since glasses would not fit into the HMD and not wearing them might confound the ET data. The participation compensation consisted of a fixed 10 Euro baseline plus a performance-based component. After arrival at the lab, participants signed a consent form. It ensured the participants' basic knowledge of the experiment procedure and informed them that the experiment complied with ethical standards. Further, it required them to grant permission to publish their data in anonymized form. For recruitment, we used the participant pool in our self-hosted online registration platform (Bock et al. 2014) and actively approached students on campus.

4.3.2 Procedure

We simulated customer purchase decisions in VR, collecting ET and ECG data. All virtual scenes were created using the Unity 2021.3 game engine. Participants entered our showroom using a Varjo VR 3 HMD with high-frequency ET capability (sampling rate of up to 200 Hz) and a display resolution of 2880 × 2720 pixels per eye. A bioPLUX device was used for ECG recording and captured signals with a sampling rate of 1000 Hz. Overall, the experiment followed a between-subjects design and included two different decision scenarios, one for 3D printers and one for washing powders (see Figure 19). To create realistic shopping scenarios, we presented dedicated cover stories to both groups. Participants were then shown a list of purchase decision criteria they had to memorize. The end of memorization phase was triggered by the participants using a button press which hid the criteria and spawned the products. Then, they had the chance to gain one Euro in addition to their participation compensation if they matched a previously determined team decision. This monetary incentive helped to motivate the participants and increased the external validity of the experiment. Participants confirmed their purchase decision either by putting the product into a shopping cart or by clicking a purchase button. After making the purchase, participants left the VR environment and answered questions about their decision phases by means of a first-person video. This video showed a gaze dot which indicated their visual attention. Participants determined the moments when they shifted (1) from orientation to evaluation and (2) from evaluation to verification. For each of these phase shifts, they entered the timestamp in a web-based questionnaire form. Furthermore, participants reported their desired help timing for a digital human agent in the same manner as for the phase shifts.

Figure 19. Experimental VR setup

(3D printer decision top, washing powder decision bottom).

4.4 Results

For our analysis, we used python 3.10 and the neurokit2 0.2.3 (Makowski et al. 2021), pomegranate 0.4.0 (Schreiber 2017), and scikit-learn 1.0.2 (Pedregosa et al. 2011) packages.

We verified our conjecture regarding the desired help timing. As expected, help was most frequently desired after the shift from orientation to evaluation but before entering the verification phase. On average, the phase shift from orientation to evaluation was indicated after 100.2 seconds (SD=79.8) and the shift from evaluation to verification was after 210 seconds (SD=97.2). Participants reported the average desired help timing for a digital human agent after 148 seconds (SD=115.8), i.e., with an average offset of 48 seconds after starting the evaluation phase and 62 seconds before entering the verification phase (see Figure 20).

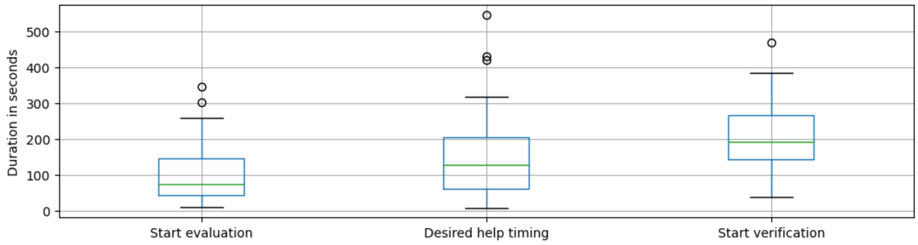

Figure 20. Boxplots of the self-reported phase shifts and the desired help timing.

Our trained model with posterior transition probabilities is shown in Figure 21. Each state is holding a multivariate GMM which consists of multiple Gaussian mixture distributions (see Figure 22 left for a univariate example). We showcase the inference of one full exemplary purchase process in Figure 22 right. Such phase predictions can be further refined and leveraged by a UAS or sales agent to find the best time to approach customers with an assistance offering. It is noteworthy that training duration only lasted 3.21 seconds and with very brief inference times a single observation can be predicted on the fly. The mean difference between classified and reported shifts from orientation to evaluation is -0.14 (SD=4.49) five second time windows. Overall, the model fits 84.89% of the five second windows correctly.

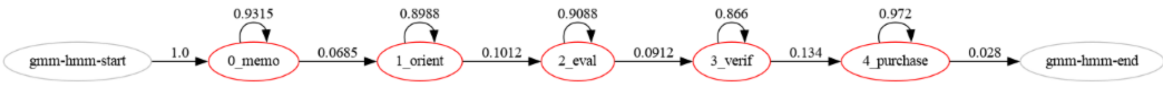

Figure 21. GMM-HMM with posterior transition probabilities.

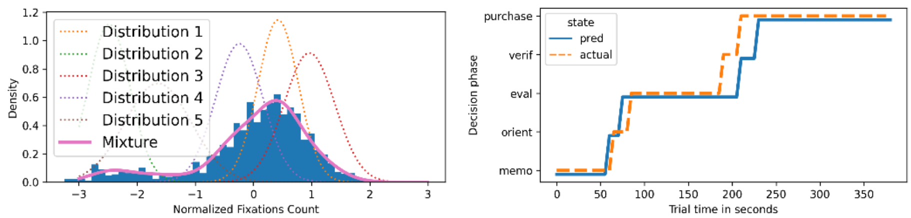

Figure 22. Exemplified univariate GMM for a single feature (left), comparison between reported state transitions and model prediction for one purchase decision (right).

4.5 Discussion

Our results show the feasibility of identifying a good timing to approach customers in a virtual commerce scenario using GMM-HMMs and thus yield an answer to our research question. The presented approach uses multiple neurophysiological sensors as input and meets our goal to overcome pure comparison and fixation-based phase determination. The presented methodology can be adopted by other researchers and practitioners to build a maybe soon to be realized overarching virtual platform, offering a multitude of interconnected virtual worlds and services.

This work has limitations which may serve as a guideline for future research. First, our sample almost exclusively consists of students, which limits generalizability. Future research should involve a broader cross-section of society. Second, the sample size should be increased. Our 50 observations yield little variety to equip the model with performant predictive power. Third, immersion, perceived telepresence, and perceived product involvement could have been increased by adding more sensory channels (particularly audio) to the virtual environment. Room size also played a limiting role, as participants had to remain relatively immobile and could not fully immerse themselves in the virtual space. Regarding the applied machine learning techniques, we plan to rigidly quantify the model performance and give detailed information about the most relevant features. We also want to consider further measurements as features, such as electrodermal activity and electroencephalography, which eventually might also be integrated into future HMDs off-the-shelf. Finally, we plan to evaluate the simplified GMM classifier version of the model in an experimental virtual commerce shopping scenario in which a digital human agent approaches a customer according to the timing suggested by the model.

References

Abowd, Gregory D.; Dey, Anind K.; Brown, Peter J.; Davies, Nigel; Smith, Mark; Steggles, Pete (1999): Towards a better understanding of context and context-awareness. In: Handheld and Ubiquitous Computing: First International Symposium, HUC'99 Karlsruhe, Germany, September 27-29, 1999 Proceedings 1. Springer, pp. 304–307.

Acharya, Nirmal; Sassenberg, Anne-Marie; Soar, Jeffrey (2022): Consumers' Behavioural Intentions to Reuse Recommender Systems: Assessing the Effects of Trust Propensity, Trusting Beliefs and Perceived Usefulness. In *Journal of Theoretical and Applied Electronic Commerce Research* 18 (1), pp. 55–78.

Afzal, Hasan; Khan, Muhammad Aslam; ur Rehman, Kashif; Ali, Imran; Wajahat, Sobia (2010): Consumer's trust in the brand: Can it be built through brand reputation, brand competence and brand predictability. In *International business research* 3 (1), p. 43.

Agarap, Abien Fred (2019): Deep Learning using Rectified Linear Units (ReLU): arXiv.

Aggarwal, Charu C. (2018): Neural networks and deep learning. In *Springer* 10, 978-3.

Alcañiz, Mariano; Bigné, Enrique; Guixeres, Jaime (2019): Virtual reality in marketing: a framework, review, and research agenda. In *Frontiers in psychology* 10, p. 1530.

Allen, Pamela M.; Edwards, John A.; Snyder, Frank J.; Makinson, Kevin A.; Hamby, David M. (2014): The effect of cognitive load on decision making with graphically displayed uncertainty information. In *Risk Analysis* 34 (8), pp. 1495–1505.

Alzayat, Ayman; Lee, Seung Hwan Mark (2021): Virtual products as an extension of my body: Exploring hedonic and utilitarian shopping value in a virtual reality retail environment. In *Journal of Business Research* 130, pp. 348–363.

Anderson, John R.; Bothell, Dan; Douglass, Scott (2004): Eye movements do not reflect retrieval processes: Limits of the eye-mind hypothesis. In *Psychological Science* 15(4), pp. 225–231.

Arcand, Manon; Nantel, Jacques; Arles-Dufour, Mathieu; Vincent, Anne (2007): The impact of reading a web site's privacy statement on perceived control over privacy and perceived trust. In *Online Information Review* 31 (5), pp. 661–681.

Atalay, A. Selin; Bodur, H. Onur; Rasolofoarison, Dina (2012): Shining in the center: Central gaze cascade effect on product choice. In *Journal of consumer research* 39 (4), pp. 848–866.

Athavipach, Chanavit; Pan-Ngum, Setha; Israsena, Pasin (2019): A wearable in-ear EEG device for emotion monitoring. In *Sensors* 19 (18), p. 4014.

Baber, C. (1996): Humans, servants and agents: human factors of intelligent domestic products. In: IEE Colloquium on Artificial Intelligence in Consumer and Domestic Products: IEE.

Bailey, Brian P.; Konstan, Joseph A. (2006): On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. In *Computers in Human Behavior* 22 (4), pp. 685–708.

Ball, Matthew (2022): The Metaverse. [S.I.]: Liveright.

Bansal, Sakshi; Gupta, Chetna; Sinha, Adwitiya (2017): Clickstream & behavioral analysis with context awareness for e-commercial applications. In: 2017 Tenth International Conference on Contemporary Computing (IC3), pp. 1–6.

Bartneck, Christoph; Kulić, Dana; Croft, Elizabeth; Zoghbi, Susana (2009): Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. In *International journal of social robotics* 1, pp. 71–81.

Bekkar, Mohamed; Djemaa, Hassiba Kheliouane; Alitouche, Taklit Akrouf (2013): Evaluation measures for models assessment over imbalanced data sets. In *Journal of Information Engineering and Applications* 3 (10).

Berg, Joyce; Dickhaut, John; McCabe, Kevin (1995): Trust, reciprocity, and social history. In *Games and economic behavior* 10 (1), pp. 122–142.

Berger, Christoph; Knierim, Michael Thomas; Benke, Ivo; Bartholomeyczik, Karen; Weinhardt, Christof (2023): InterFlowCeption: Foundations for Technological Enhancement of Interoception to Foster Flow States during Mental Work: About the potential of technologically supported body awareness to promote flow experiences during mental work. In: Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems, pp. 1–8.

Bigné, Enrique; Llinares, Carmen; Torrecilla, Carmen (2016): Elapsed time on first buying triggers brand choices within a category: A virtual reality-based study. In *Journal of Business Research* 69 (4), pp. 1423–1427.

Bijlsma-Frankema, Katinka; Costa, Ana Cristina (2005): Understanding the trust-control nexus. In *International sociology* 20 (3), pp. 259–282.

Block, Lauren G.; Grier, Sonya A.; Childers, Terry L.; Davis, Brennan; Ebert, Jane E.J.; Kumanyika, Shiriki et al. (2011): From Nutrients to Nurturance: A Conceptual Introduction to Food Well-Being. In *Journal of Public Policy & Marketing* 30 (1), pp. 5–13.

Bock, Olaf; Baetge, Ingmar; Nicklisch, Andreas (2014): hroot: Hamburg Registration and Organization Online Tool. In *European Economic Review* 71, pp. 117–120.

Bolea, Juan; Laguna, Pablo; Remartia; Rovira, Eva; Navarro, Augusto; Bailón, Raquel (2014): Methodological framework for estimating the correlation dimension in HRV signals. In *Computational and mathematical methods in medicine* 2014.

Bouneffouf, Djallel; Bouzeghoub, Amel; Gançarski, Alda Lopes (2012): A contextual-bandit algorithm for mobile context-aware recommender system. In: Neural Information Processing: 19th International Conference, ICONIP 2012, Doha, Qatar, November 12-15, 2012, Proceedings, Part III 19. Springer, pp. 324–331.

Bourlakis, Michael; Papagiannidis, Savvas; Li, Feng (2009): Retail spatial evolution: paving the way from traditional to metaverse retailing. In *Electronic Commerce Research* 9 (1), pp. 135–148.

Brachten, Florian; Brünker, Felix; Frick, Nicholas R. J.; Ross, Björn; Stieglitz, Stefan (2020): On the ability of virtual agents to decrease cognitive load: an experimental study. In *Information Systems and e-Business Management* 18, pp. 187–207.

Bradshaw, Jeffrey M. (1997): An introduction to software agents. In *Software agents* 4, pp. 3–46.

Branca, Generoso; Marino, Vittoria; Resciniti, Riccardo (2023): How do consumers evaluate products in virtual reality? A literature review for a research agenda. In *Spanish Journal of Marketing-ESIC*.

Browne, Michael W. (2000): Cross-validation methods. In *Journal of mathematical psychology* 44 (1), pp. 108–132.

Bublitz, Melissa G.; Peracchio, Laura A.; Andreasen, Alan R.; Kees, Jeremy; Kidwell, Blair; Miller, Elizabeth Gelfand et al. (2013): Promoting positive change: Advancing the food well-being paradigm. In *Journal of Business Research* 66 (8), pp. 1211–1218.

Cantoni, Virginio; Musci, Mirto; Nugrahaningsih, Nahumi; Porta, Marco (2018): Gaze-based biometrics: An introduction to forensic applications. In *Pattern Recognition Letters* 113, pp. 54–57.

Capurso, Nicholas; Mei, Bo; Song, Tianyi; Cheng, Xiuzhen; Yu, Jiguo (2018): A survey on key fields of context awareness for mobile devices. In *Journal of Network and Computer Applications* 118, pp. 44–60.

Carrera-Rivera, Angela; Larrinaga, Felix; Lasa, Ganix (2022): Context-awareness for the design of Smart-product service systems: Literature review. In *Computers in Industry* 142, p. 103730.

Caruelle, Delphine; Lervik-Olsen, Line; Gustafsson, Anders (2023): The clock is ticking—Or is it? Customer satisfaction response to waiting shorter vs. longer than expected during a service encounter. In *Journal of Retailing*.

Castagnos, Sylvain; Jones, Nicolas; Pu, Pearl (2010): Eye-tracking product recommenders' usage. In: Proceedings of the fourth ACM conference on Recommender systems, pp. 29–36.

Castelfranchi, Cristiano; Falcone, Rino (2000): Trust and control: A dialectic link. In *Applied Artificial Intelligence* 14 (8), pp. 799–823.

Castelo, Noah; Bos, Maarten W.; Lehmann, Donald R. (2019): Task-dependent algorithm aversion. In *Journal of Marketing Research* 56 (5), pp. 809–825.

Chanel, Caroline P. C.; Wilson, Matthew D.; Scannella, Sebastien (2019): Online ECG-based Features for Cognitive Load Assessment. In: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC): IEEE.

Chawla, Nitesh V.; Bowyer, Kevin W.; Hall, Lawrence O.; Kegelmeyer, W. Philip (2002): SMOTE: synthetic minority over-sampling technique. In *Journal of artificial intelligence research* 16, pp. 321–357.

Chen, Bing; Yang, De-Nian (2022): User Recommendation in Social Metaverse with VR. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management. New York, NY, USA. New York, NY, USA: ACM.

Chen, Tianqi; Guestrin, Carlos (2016): Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785–794.

Chen, Yi-Cheng; Thaipisutikul, Tipajin; Shih, Timothy K. (2020): A learning-based POI recommendation with spatiotemporal context awareness. In *IEEE Transactions on Cybernetics* 52 (4), pp. 2453–2466.

Chittaro, Luca; Ration, Roberto (2000): Adding adaptive features to virtual reality interfaces for e-commerce. In: International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems. Springer, pp. 86–97.

Cho, Jaehee; Park, Dongjin; Lee, H. Erin; others (2014): Cognitive factors of using health apps: systematic analysis of relationships among health consciousness, health information orientation, eHealth literacy, and health app use efficacy. In *Journal of medical Internet research* 16 (5), e3283.

Chrzan, Keith; Orme, Bryan (2000): An overview and comparison of design strategies for choice-based conjoint analysis. In *Sawtooth software research paper series* 98382, p. 360.

Clevert, Djork-Arné; Unterthiner, Thomas; Hochreiter, Sepp (2015): Fast and accurate deep network learning by exponential linear units (elus). In *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics* 20, pp. 924–932.

Corbitt, Brian J.; Thanasankit, Theerasak; Yi, Han (2003): Trust and e-commerce: a study of consumer perceptions. In *Electronic commerce research and applications* 2 (3), pp. 203–215.

Cruz-Neira, Carolina; Sandin, Daniel J.; DeFanti, Thomas A.; Kenyon, Robert V.; Hart, John C. (1992): The CAVE: audio visual experience automatic virtual environment. In *Commun. ACM* 35 (6), pp. 64–72.

Cummings, James J.; Bailenson, Jeremy N. (2016): How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. In *Media psychology* 19 (2), pp. 272–309.

Dasgupta, Archi; Manuel, Mark; Mansur, Rifat Sabbir; Nowak, Nabil; Gračanin, Denis (2020): Towards real time object recognition for context awareness in mixed reality: a machine learning approach. In: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). IEEE, pp. 262–268.

David-John, Brendan; Peacock, Candace; Zhang, Ting; Murdison, T. Scott; Benko, Hrvoje; Jonker, Tanya R. (2021): Towards gaze-based prediction of the intent to interact in virtual reality. In: ACM Symposium on Eye Tracking Research and Applications, pp. 1–7.

Davis, Simon; Nesbitt, Keith; Nalivaiko, Eugene (2014): A Systematic Review of Cybersickness. In: Proceedings of the 2014 Conference on Interactive Entertainment. New York, NY, USA. New York, NY, USA: ACM.

Deck, Cary; Jahedi, Salar (2015): The effect of cognitive load on economic decision making: A survey and new experiments. In *European Economic Review* 78, pp. 97–119.

Deljoo, Ameneh; van Engers, Tom; Gommans, Leon; Laat, Cees de (2018): The impact of competence and benevolence in a computational model of trust. In: Trust Management XII: 12th IFIP WG 11.11 International Conference, IFIPTM 2018, Toronto, ON, Canada, July 10-13, 2018, Proceedings 12. Springer, pp. 45–57.

Dewitte, Siegfried; Pandelaere, Mario; Briers, Barbara; Warlop, Luk (2005): Cognitive Load has Negative After Effects on Consumer Decision Making.

Dey, Anind K. (2009): Explanations in Context-Aware Systems. In: ExaCt, pp. 84–93.

Drif, Ahlem; Zerrad, Houssem Eddine; Cherifi, Hocine (2021): Context-awareness in ensemble recommender system framework. In: 2021 International Conference on Electrical, Communication, and Computer Engineering (ICECCE). IEEE, pp. 1–6.

Duchowski, Andrew T. (2017): Eye tracking methodology: Theory and practice: Springer.

Eddy, Sean R. (2004): What is a hidden Markov model? In *Nature biotechnology* 22 (10), pp. 1315–1316.

Engel, James F.; Kollat, David T.; Blackwell, Roger D. (1968): A model of consumer motivation and behavior. In *Research in Consumer Behavior, Holt, Rinehart and Winston, Inc., New York, NY*, pp. 3–20.

Evans, Philip; Wurster, Thomas S. (1999): Getting real about virtual commerce. In *Harvard Business Review* 77, pp. 84–98.

Fenko, Anna; Nicolaas, Iris; Galetzka, Mirjam (2018): Does attention to health labels predict a healthy food choice? An eye-tracking study. In *Food quality and preference* 69, pp. 57–65.

Ferdinando, Hany; Seppanen, Tapio; Alasaarela, Esko (2016): Comparing features from ECG pattern and HRV analysis for emotion recognition system. In: 2016 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB): IEEE.

Fernandez, Rosa Maria (2019): SDG3 Good Health and Well-Being: Integration and Connection with Other SDGs. In: Encyclopedia of the UN Sustainable Development Goals. Cham: Springer International Publishing, pp. 629–636.

Friemel, Celina; Morana, Stefan; Pfeiffer, Jella; Maedche, Alexander (2018): On the Role of Users' Cognitive-Affective States for User Assistance Invocation. In: Information Systems and Neuroscience: Springer, Cham, pp. 37–46.

Gadalla, Eman; Keeling, Kathy; Abosag, Ibrahim (2013): Metaverse-retail service quality: A future framework for retail service quality in the 3D internet. In *Journal of Marketing Management* 29 (13-14), pp. 1493–1517.

Gallace, Alberto; Girondini, Matteo (2022): Social touch in virtual reality. In *Current Opinion in Behavioral Sciences* 43, pp. 249–254.

Gellersen, Hans W.; Schmidt, Albrecht; Beigl, Michael (2002): Multi-sensor context-awareness in mobile devices and smart artifacts. In *Mobile Networks and Applications* 7, pp. 341–351.

Gidlöf, Kerstin; Wallin, Annika; Dewhurst, Richard; Holmqvist, Kenneth (2013): Using Eye Tracking to Trace a Cognitive Process: Gaze Behaviour During Decision Making in a Natural Environment. In *JEMR* 6 (1).

Ginon, Blandine; Stumpf, Simone; Jean-Daubias, Stéphanie (2016): Towards the Right Assistance at the Right Time for Using Complex Interfaces. In: Proceedings of the International Working Conference on Advanced Visual Interfaces. New York, NY, USA. New York, NY, USA: ACM.

Ginting, Y.; Chandra, Teddy; Miran, Ikas; Yusriadi, Yusriadi (2023): Repurchase intention of e-commerce customers in Indonesia: An overview of the effect of e-service quality, e-word of mouth, customer trust, and customer satisfaction mediation. In *International Journal of Data and Network Science* 7 (1), pp. 329–340.

Gonzalez-Franco, Mar; Ofek, Eyal; Pan, Ye; Antley, Angus; Steed, Anthony; Spanlang, Bernhard et al. (2020): The rocketbox library and the utility of freely available rigged avatars. In *Frontiers in virtual reality* 1, p. 20.

Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron (2016): Deep learning: MIT press.

Goy, Jean-Jacques (2013): Electrocardiography (ECG): Bentham Science Publishers.

Grubert, Jens; Langlotz, Tobias; Zollmann, Stefanie; Regenbrecht, Holger (2016): Towards pervasive augmented reality: Context-awareness in augmented reality. In *IEEE transactions on visualization and computer graphics* 23 (6), pp. 1706–1724.

Grzyb, Tomasz; Dolinski, Dariusz; Kozłowska, Agnieszka (2018): Is Product Placement Really Worse Than Traditional Commercials? Cognitive Load and Recalling of Advertised Brands. In *Front. Psychol.* 9, p. 1519.

Guo, Guibing; Elgendi, Mohamed (2013): A New Recommender System for 3D E-Commerce: An EEG Based Approach. In *JOAMS* 1 (1), pp. 61–65.

Guo, Ruoxi; Izzouzi, Lisa; Steed, Anthony (2023): User Motion Accentuation in Social Pointing Scenario. In: 2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). IEEE, pp. 575–576.

Gupta, Kunal; Hajika, Ryo; Pai, Yun Suen; Duenser, Andreas; Lochner, Martin; Billinghurst, Mark (2020): Measuring human trust in a virtual assistant using physiological sensing in virtual reality. In: 2020 IEEE Conference on virtual reality and 3D user interfaces (VR). IEEE, pp. 756–765.

Haapalainen, Eija; Kim, SeungJun; Forlizzi, Jodi F.; Dey, Anind K. (2010): Psycho-physiological measures for assessing cognitive load. In: Proceedings of the 12th ACM international conference on Ubiquitous computing, pp. 301–310.

Halbig, Andreas; Latoschik, Marc Erich (2021): A Systematic Review of Physiological Measurements, Factors, Methods, and Applications in Virtual Reality. In *Front. Virtual Real.* 2, Article 694567, p. 89. Available online at https://www.frontiersin.org/articles/10.3389/frvir.2021.694567/full.

Haring, Kerstin Sophie; Matsumoto, Yoshio; Watanabe, Katsumi (2013): How do people perceive and trust a lifelike robot. In: Proceedings of the world congress on engineering and computer science, vol. 1, pp. 425–430.

Harris, Joanne M.; Ciorciari, Joseph; Gountas, John (2018): Consumer neuroscience for marketing researchers. In *Journal of consumer behaviour* 17 (3), pp. 239–252.

Hart, Sandra G. (2006): NASA-task load index (NASA-TLX); 20 years later. In: Proceedings of the human factors and ergonomics society annual meeting, vol. 50. Sage publications Sage CA: Los Angeles, CA, pp. 904–908.

Hart, Sandra G.; Staveland, Lowell E. (1988): Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Advances in Psychology: Elsevier, pp. 139–183.

Harz, Nathalie; Hohenberg, Sebastian; Homburg, Christian (2022): Virtual reality in new product development: Insights from prelaunch sales forecasting for durables. In *Journal of Marketing* 86 (3), pp. 157–179.

Häubl, Gerald; Trifts, Valerie (2000): Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids. In *Marketing Science* 19 (1), pp. 4–21.

Hauser, John R. (2014): Consideration-set heuristics. In *Journal of Business Research* 67 (8), pp. 1688–1699.

Hausman, Daniel M.; Welch, Brynn (2010): Debate: To nudge or not to nudge. In *Journal of Political Philosophy* 18 (1), pp. 123–136.

Hayes, Andrew F. (2017): Introduction to mediation, moderation, and conditional process analysis: A regression-based approach: Guilford Publications.

He, Sherry; Hollenbeck, Brett; Proserpio, Davide (2022): The market for fake reviews. In *Marketing Science* 41 (5), pp. 896–921.

Hercberg, Serge; Touvier, Mathilde; Salas-Salvado, Jordi (2021): The Nutri-score nutrition label: Hogrefe AG.

Hess, Eckhard H.; Polt, James M. (1964): Pupil size in relation to mental activity during simple problem-solving. In *Science* 143 (3611), pp. 1190–1192.

Heßler, Pascal Oliver; Pfeiffer, Jella; Hafenbrädl, Sebastian (2022): When Self-Humanization Leads to Algorithm Aversion. In *Bus Inf Syst Eng* 64 (3), pp. 275–292.

Hoang, Hanh Thi Hong; Nguyen, Lam Son; Nguyen, Chinh Hong; Le, Nga Viet; Nguyen, Nhung Thi; Dinh, Len Thi (2023): Personalisation-Privacy Paradox from Marketing Perspectives: Literature Review and Future Research Directions. In *Business Innovation for the Post-pandemic Era in Vietnam*, pp. 87–105.

Hochreiter, Sepp (1998): The vanishing gradient problem during learning recurrent neural nets. In *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 6 (02), pp. 107–116.

Holmqvist, Kenneth; Nyström, Marcus; Andersson, Richard; Dewhurst, Richard; Jarodzka, Halszka; van de Weijer, Joost (2011): Eye tracking: A comprehensive guide to methods and measures: OUP Oxford.

Hornsey, Matthew J.; Greenaway, Katharine H.; Harris, Emily A.; Bain, Paul G. (2019): Exploring cultural differences in the extent to which people perceive and desire control. In *Personality and Social Psychology Bulletin* 45 (1), pp. 81–92.

Horvitz, Eric J.; Breese, John S.; Heckerman, David; Hovel, David; Rommelse, Koos (2013): The Lumiere project: Bayesian user modeling for inferring the goals and needs of software users. In *arXiv preprint arXiv:1301.7385*.

Hota, H. S.; Handa, Richa; Shrivas, Akhilesh Kumar (2017): Time series data prediction using sliding window based RBF neural network. In *International Journal of Computational Intelligence Research* 13 (5), pp. 1145–1156.

Hu, Xiaohan (2023): Empowering consumers in interactive marketing: examining the role of perceived control. In: The Palgrave Handbook of Interactive Marketing: Springer, pp. 117–147.

Huang, Jianping; Zhao, Ping; Wan, Xiaoang (2021): From brain variations to individual differences in the color-flavor incongruency effect: A combined virtual reality and resting-state fMRI study. In *Journal of Business Research* 123, pp. 604–612.

Huang, Lihua; Ba, Sulin; Lu, Xianghua (2014): Building online trust in a culture of Confucianism: The impact of process flexibility and perceived control. In *ACM Transactions on Management Information Systems (TMIS)* 5 (1), pp. 1–23.

Hughes, Ashley M.; Hancock, Gabriella M.; Marlow, Shannon L.; Stowers, Kimberly; Salas, Eduardo (2019): Cardiac Measures of Cognitive Workload: A Meta-Analysis. In *Hum Factors* 61 (3), pp. 393–414.

Ishaque, Syem; Rueda, Alice; Nguyen, Binh; Khan, Naimul; Krishnan, Sridhar (2020): Physiological signal analysis and classification of stress from virtual reality video game. In: 2020

42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, pp. 867–870.

Ismail Fawaz, Hassan; Lucas, Benjamin; Forestier, Germain; Pelletier, Charlotte; Schmidt, Daniel F.; Weber, Jonathan et al. (2020): Inceptiontime: Finding alexnet for time series classification. In *Data Mining and Knowledge Discovery* 34 (6), pp. 1936–1962.

Izzouzi, Lisa; Steed, Anthony (2022): Integrating rocketbox avatars with the ubiq social vr platform. In: 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). IEEE, pp. 69–70.

Jaeggi, Susanne M.; Buschkuehl, Martin; Perrig, Walter J.; Meier, Beat (2010): The concurrent validity of the N-back task as a working memory measure. In *Memory* 18 (4), pp. 394–412.

Jensen, Lasse; Konradsen, Flemming (2018): A review of the use of virtual reality head-mounted displays in education and training. In *Education and Information Technologies* 23 (4), pp. 1515–1529.

Johnson, Wendy; Carothers, Andrew; Deary, Ian J. (2008): Sex differences in variability in general intelligence: A new look at the old question. In *Perspectives on psychological science* 3 (6), pp. 518–531.

Jones, Kiku; Leonard, Lori N. K. (2008): Trust in consumer-to-consumer electronic commerce. In *Information & Management* 45 (2), pp. 88–95.

Just, Marcel A.; Carpenter, Patricia A. (1980): A theory of reading: From eye fixations to comprehension. In *Psychological Review* 87 (4), pp. 329–354.

Kahneman, Daniel (1973): Attention and effort: Citeseer (1063).

Kim, Eojina; Tang, Liang Rebecca; Meusel, Chase; Gupta, Manjul (2018): Optimization of menu-labeling formats to drive healthy dining: An eye tracking study. In *International Journal of Hospitality Management* 70, pp. 37–48.

Kim, Kyung Kyu; Prabhakar, Bipin (2004): Initial trust and the adoption of B2C e-commerce: The case of internet banking. In *ACM SIGMIS database: the DATABASE for advances in information systems* 35 (2), pp. 50–64.

Kliestik, Tomas; Novak, Andrej; Lazaroiu, George (2022): Live Shopping in the Metaverse: Visual and Spatial Analytics, Cognitive Artificial Intelligence Techniques and Algorithms, and Immersive Digital Simulations. In *Linguistic & Philosophical Investigations*.

Kovacova, Maria; Machova, Veronika; Bennett, Daniel (2022): Immersive extended reality technologies, data visualization tools, and customer behavior analytics in the metaverse commerce. In *Journal of Self-Governance and Management Economics* 10 (2), pp. 7–21.

Krajbich, Ian; Rangel, Antonio (2011): Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. In *Proceedings of the National Academy of Sciences of the United States of America* 108 (33), pp. 13852–13857. DOI: 10.1073/pnas.1101328108.

Krawczyk, Bartosz (2016): Learning from imbalanced data: open challenges and future directions. In *Progress in Artificial Intelligence* 5 (4), pp. 221–232.

Krogh, Anders; Brown, Michael; Mian, I. Saira; Sjölander, Kimmen; Haussler, David (1994): Hidden Markov models in computational biology: Applications to protein modeling. In *Journal of molecular biology* 235 (5), pp. 1501–1531.

Kruschke, John (2014): Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan.

Kulkarni, Saurabh; Rodd, Sunil F. (2020): Context Aware Recommendation Systems: A review of the state of the art techniques. In *Computer Science Review* 37, p. 100255.

Laeng, Bruno; Sirois, Sylvain; Gredebäck, Gustaf (2012): Pupillometry: A window to the preconscious? In *Perspectives on psychological science* 7 (1), pp. 18–27.

Lallemand, Carine; Koenig, Vincent (2020): Measuring the contextual dimension of user experience: development of the user experience context scale (UXCS). In: Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society, pp. 1–13.

Lanier, Jaron; Biocca, Frank (1992): An Insider's View of the Future of Virtual Reality. In *Journal of Communication* 42 (4), pp. 150–172.

Lao, J.; Louf, R. (2020): Blackjax: A sampling library for JAX. In *URL http://github. com/blackjax-devs/blackjax*.

Lee, Haeinn; Seo, Ssanghee (2010): A comparison and analysis of usability methods for web evaluation: the relationship between typical usability test and bio-signals characteristics (EEG, ECG).

Lee, John; Moray, Neville (1992): Trust, control strategies and allocation of function in human-machine systems. In *Ergonomics* 35 (10), pp. 1243–1270.

Lee, Yun Jung; Dubinsky, Alan J. (2017): Consumers' desire to interact with a salesperson during e-shopping: development of a scale. In *International Journal of Retail & Distribution Management* 45 (1), pp. 20–39.

Li, Liam; Jamieson, Kevin; Rostamizadeh, Afshin; Gonina, Ekaterina; Hardt, Moritz; Recht, Benjamin; Talwalkar, Ameet (2020): A system for massively parallel hyperparameter tuning. In *Proceedings of Machine Learning and Systems* 2, pp. 230–246.

Li, Xitong; Grahl, Jörn; Hinz, Oliver (2022): How Do Recommender Systems Lead to Consumer Purchases? A Causal Mediation Analysis of a Field Experiment. In *Information Systems Research* 33 (2), pp. 620–637.

Liaw, Richard; Liang, Eric; Nishihara, Robert; Moritz, Philipp; Gonzalez, Joseph E.; Stoica, Ion (2018): Tune: A Research Platform for Distributed Model Selection and Training: arXiv.

Lim, Brian Y. (2010): Improving trust in context-aware applications with intelligibility. In: Proceedings of the 12th ACM international conference adjunct papers on Ubiquitous computing-Adjunct, pp. 477–480.

Lin, Tsung-Yi; Goyal, Priya; Girshick, Ross; He, Kaiming; Dollár, Piotr (2017): Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision, pp. 2980–2988.

Liu, Hui; Gamboa, Hugo; Schultz, Tanja (2023): Sensor-Based Human Activity and Behavior Research: Where Advanced Sensing and Recognition Technologies Meet. In *Sensors* 23 (1).

Lundberg, Scott M.; Lee, Su-In (2017): A unified approach to interpreting model predictions. In *Advances in neural information processing systems* 30.

Luo, Wenjie; Li, Yujia; Urtasun, Raquel; Zemel, Richard (2016): Understanding the effective receptive field in deep convolutional neural networks. In *Advances in neural information processing systems* 29.

M. J. Betancourt; Simon Byrne; Mark Girolami (2015): Optimizing The Integrator Step Size for Hamiltonian Monte Carlo.

Maedche, Alexander; Morana, Stefan; Schacht, Silvia; Werth, Dirk; Krumeich, Julian (2016): Advanced user assistance systems. In *Bus Inf Syst Eng* 58, pp. 367–370.

Makowski, Dominique; Pham, Tam; Lau, Zen J.; Brammer, Jan C.; Lespinasse, François; Pham, Hung et al. (2021): NeuroKit2: A Python toolbox for neurophysiological signal processing. In *Behav Res* 53 (4), pp. 1689–1696.

Maratea, Antonio; Petrosino, Alfredo; Manzo, Mario (2014): Adjusted F-measure and kernel scaling for imbalanced data learning. In *Information Sciences* 257, pp. 331–341.

Mariotti, Thomas; Schweizer, Nikolaus; Szech, Nora; Wangenheim, Jonas von (2023): Information nudges and self-control. In *Management Science* 69 (4), pp. 2182–2197.

Martin, Osvaldo (2018): Bayesian Analysis with Python: Introduction to statistical modeling and probabilistic programming using PyMC3 and ArviZ: Packt Publishing Ltd.

Martin, Osvaldo A.; Kumar, Ravin; Lao, Junpeng (2021): Bayesian modeling and computation in python: CRC Press.

Martinez-Navarro, Jesus; Bigné, Enrique; Guixeres, Jaime; Alcañiz, Mariano; Torrecilla, Carmen (2019): The influence of virtual reality in e-commerce. In *Journal of Business Research* 100, pp. 475–482.

McElreath, Richard (2018): Statistical rethinking: A Bayesian course with examples in R and Stan: Chapman and Hall/CRC.

McKnight, D. Harrison; Carter, Michelle; Thatcher, Jason Bennett; Clay, Paul F. (2011): Trust in a specific technology: An investigation of its components and measures. In *ACM Transactions on Management Information Systems (TMIS)* 2 (2), pp. 1–25.

McKnight, D. Harrison; Choudhury, Vivek; Kacmar, Charles (2002): Developing and validating trust measures for e-commerce: An integrative typology. In *Information Systems Research* 13 (3), pp. 334–359.

Meißner, Martin; Pfeiffer, Jella; Peukert, Christian; Dietrich, Holger; Pfeiffer, Thies (2020): How virtual reality affects consumer choice. In *Journal of Business Research* 117, pp. 219–231.

Meißner, Martin; Pfeiffer, Jella; Pfeiffer, Thies; Oppewal, Harmen (2019): Combining virtual reality and mobile eye tracking to provide a naturalistic experimental environment for shopper research. In *Journal of Business Research* 100, pp. 445–458.

Meurisch, Christian; Mihale-Wilson, Cristina A.; Hawlitschek, Adrian; Giger, Florian; Müller, Florian; Hinz, Oliver; Mühlhäuser, Max (2020): Exploring user expectations of proactive AI systems. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4 (4), pp. 1–22.

Middlehurst, Matthew; Large, James; Flynn, Michael; Lines, Jason; Bostrom, Aaron; Bagnall, Anthony (2021): HIVE-COTE 2.0: a new meta ensemble for time series classification. In *Machine Learning* 110 (11), pp. 3211–3243.

Miles, W. R. (1929): Ocular dominance demonstrated by unconscious sighting. In *Journal of Experimental Psychology* 12 (2), pp. 113–126.

Millecamp, Martijn; Conati, Cristina; Verbert, Katrien (2021): Classifeye: classification of personal characteristics based on eye tracking data in a recommender system interface. In: Joint Proceedings of the ACM IUI 2021 Workshops co-located with 26th ACM Conference on Intelligent User Interfaces, vol. 2903. CEUR Workshop Proceedings.

Miller, George A. (1956): The magical number seven, plus or minus two: Some limits on our capacity for processing information. In *Psychological Review* 63 (2), pp. 81–97.

Moghaddasi, Masoud; Marín-Morales, Javier; Khatri, Jaikishan; Guixeres, Jaime; Chicchi Giglioli, Irene Alice; Alcañiz, Mariano (2021): Recognition of Customers' Impulsivity from Behavioral Patterns in Virtual Reality. In *Applied Sciences* 11 (10), p. 4399.

Möllering, Guido (2005): The trust/control duality: An integrative perspective on positive expectations of others. In *International sociology* 20 (3), pp. 283–305.

Morana, Stefan; Pfeiffer, Jella; Adam, Marc T. P. (2020): User Assistance for Intelligent Systems. In *Business & Information Systems Engineering* 62 (3), pp. 189–192.

Mori, Masahiro (1970): The uncanny valley: the original essay by Masahiro Mori. In *IEEE Spectrum* 6.

Moussawi, Sara; Koufaris, Marios; Benbunan-Fich, Raquel (2021): How perceptions of intelligence and anthropomorphism affect adoption of personal intelligent agents. In *Electronic Markets* 31, pp. 343–364.

Muir, Bonnie M.; Moray, Neville (1996): Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. In *Ergonomics* 39 (3), pp. 429–460.

Naruetharadhol, Phaninee; Wongsaichia, Sasichakorn; Pienwisetkaew, Teerapong; Schrank, Johannes; Chaiwongjarat, Kullanan; Thippawong, Peeranut et al. (2023): Consumer Intention to Utilize an E-Commerce Platform for Imperfect Vegetables Based on Health-Consciousness. In *Foods* 12 (6), p. 1166.

Nicoll, Derek (1999): Taxonomy of information intensive products. In *The University of Edinburgh Management School: Edimburgh, Scotland, UK*.

Novák, Jakub Štěpán; Masner, Jan; Benda, Petr; Šimek, Pavel; Merunka, Vojtěch (2023): Eye Tracking, Usability, and User Experience: A Systematic Review. In *International Journal of Human-Computer Interaction*, pp. 1–17.

Oliveira, Tiago; Alhinho, Matilde; Rita, Paulo; Dhillon, Gurpreet (2017): Modelling and testing consumer trust dimensions in e-commerce. In *Computers in Human Behavior* 71, pp. 153–164.

Orquin, Jacob L.; Holmqvist, Kenneth (2018): Threats to the validity of eye-movement research in psychology. In *Behavior Research Methods* 50, pp. 1645–1656.

Pan, Huali; Zhang, Zhijun (2021): Research on context-awareness mobile tourism e-commerce personalized recommendation model. In *Journal of Signal Processing Systems* 93, pp. 147–154.

Pan, Yue; Zhang, Jason Q. (2011): Born Unequal: A Study of the Helpfulness of User-Generated Product Reviews. In *Journal of Retailing* 87 (4), pp. 598–612.

Panigrahi, Shrikant; Azizan, Noor Azlinna; Waris, Muhammad (2018): Investigating the empirical relationship between service quality, trust, satisfaction, and intention of customers purchasing life insurance products. In *Indian Journal of Marketing• January*.

Papadopoulou, Panagiota (2007): Applying virtual reality for trust-building e-commerce environments. In *Virtual Reality* 11, pp. 107–127.

Papagiannidis, Savvas; Bourlakis, Michael A. (2010): Staging the New Retail Drama: at a Metaverse near you! In *Journal of Virtual Worlds Research* 2 (5), pp. 425–446.

Parashar, Sapna; Singh, Supriya; Sood, Gunjan (2023): Examining the role of health consciousness, environmental awareness and intention on purchase of organic food: A moderated model of attitude. In *Journal of Cleaner Production* 386, p. 135553.

Parise, Salvatore; Kiesler, Sara; Sproull, Lee; Waters, Keith (1999): Cooperating with life-like interface agents. In *Computers in Human Behavior* 15 (2), pp. 123–142.

Park, Chan-Wook; Moon, Byeong-Joon (2003): The relationship between product involvement and product knowledge: Moderating roles of product type and product knowledge type. In *Psychology & Marketing* 20 (11), pp. 977–997.

Park, Hyejune; Kim, Seeun (2021): Do augmented and virtual reality technologies increase consumers' purchase intentions? The role of cognitive elaboration and shopping goals. In *Clothing and Textiles Research Journal* 41 (2), pp. 91–106.

Patil, Anand; Huard, David; Fonnesbeck, Christopher J. (2010): PyMC: Bayesian stochastic modelling in Python. In *Journal of statistical software* 35 (4), p. 1.

Patney, Anjul; Salvi, Marco; Kim, Joohwan; Kaplanyan, Anton; Wyman, Chris; Benty, Nir et al. (2016): Towards foveated rendering for gaze-tracked virtual reality. In *ACM Trans. Graph.* 35 (6), pp. 1–12. DOI: 10.1145/2980179.2980246.

Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O. et al. (2011): Scikit-learn: Machine Learning in Python. In *Journal of Machine Learning Research* 12, pp. 2825–2830.

Perl, Oskar (2020): Distributed Multi-User VR With Full-Body Avatars. Technische Universität Wien.

Petter, Stacie; Straub, Detmar; Rai, Arun (2007): Specifying formative constructs in information systems research. In *MIS Quarterly*, pp. 623–656.

Peukert, Christian; Lechner, Jessica; Pfeiffer, Jella; Weinhardt, Christof (2020): Intelligent invocation: Towards designing context-aware user assistance systems based on real-time eye tracking data analysis. In: Information systems and neuroscience: Springer, pp. 73–82.

Peukert, Christian; Pfeiffer, Jella; Meißner, Martin; Pfeiffer, Thies; Weinhardt, Christof (2019): Shopping in virtual reality stores: the influence of immersion on system adoption. In *Journal of Management Information Systems* 36 (3), pp. 755–788.

Peukert, Christian; Weinhardt, Christof; Hinz, Oliver; van der Aalst, Wil M. P. (2022): Metaverse: How to Approach Its Challenges from a BISE Perspective: Springer. In *Business & Information Systems Engineering*.

Pfeiffer, Jella (2011): Interactive decision aids in e-commerce: Springer Science & Business Media.

Pfeiffer, Jella; Pfeiffer, Thies; Meißner, Martin; Weiß, Elisa (2020): Eye-tracking-based classification of information search behavior using machine learning: evidence from experiments in physical shops and virtual reality shopping environments. In *Information Systems Research* 31 (3), pp. 675–691.

Pham, Tam; Lau, Zen; Chen, S. H.; Makowski, Dominique (2021): Heart rate variability in psychology: a review of HRV indices and an analysis tutorial. In *Sensors* 21 (12), p. 3998.

Pietroszek, Krzysztof (2019): Raycasting in Virtual Reality: Springer International Publishing AG, part of Springer Nature 2018.

Plass, Jan L.; Moreno, Roxana; Brünken, Roland (2010): Cognitive load theory. Cambridge, New York: Cambridge University Press.

Polson, Nicholas G.; Scott, James G. (2012): On the half-Cauchy prior for a global scale parameter.

Popescu, Gheorghe H.; Valaskova, Katarina; Horak, Jakub (2022): Augmented Reality Shopping Experiences, Retail Business Analytics, and Machine Vision Algorithms in the Virtual Economy of the Metaverse. In *Journal of Self-Governance & Management Economics* 10 (2).

Pu, Pearl; Chen, Li (2006): Trust building with explanation interfaces. In: Proceedings of the 11th international conference on Intelligent user interfaces, pp. 93–100.

Quelch, John A.; Klein, Lisa R. (1996): The Internet and international marketing. In *MIT Sloan Management Review* 37 (3), p. 60.

Rabiner, Lawrence; Juang, Biinghwang (1986): An introduction to hidden Markov models. In *ieee assp magazine* 3 (1), pp. 4–16.

Rabiner, Lawrence R. (1989): A tutorial on hidden Markov models and selected applications in speech recognition. In *Proceedings of the IEEE* 77 (2), pp. 257–286.

Rammstedt, Beatrice; Kemper, Christoph J.; Klein, Mira Céline; Beierlein, Constanze; Kovaleva, Anastassiya (2013): A Short Scale for Assessing the Big Five Dimensions of Personality: 10 Item Big Five Inventory (BFI-10). In *methods, data, analyses* 7 (2), p. 17.

Raut, Aritra; Tiwari, Abhisek; Das, Subrata; Saha, Sriparna; Maitra, Anutosh; Ramnani, Roshni; Sengupta, Shubhashis (2023): Reinforcing personalized persuasion in task-oriented virtual sales assistant. In *PLOS ONE* 18 (1).

Refaeilzadeh, Payam; Tang, Lei; Liu, Huan (2009): Cross-validation. In *Encyclopedia of database systems* 5, pp. 532–538.

Regt, Anouk de; Barnes, Stuart J. (2019): V-Commerce in Retail: Nature and Potential Impact. In: Augmented Reality and Virtual Reality: Springer, Cham, pp. 17–25.

Reinartz, Werner; Wiegand, Nico; Imschloss, Monika (2019): The impact of digital transformation on the retailing value chain. In *International Journal of Research in Marketing* 36 (3), pp. 350–366.

Richthammer, Christian; Pernul, Günther (2020): Situation awareness for recommender systems. In *Electronic Commerce Research* 20 (4), pp. 783–806.

Riedl, René (2012): On the biology of technostress: literature review and research agenda. In *ACM SIGMIS database: the DATABASE for advances in information systems* 44 (1), pp. 18–55.

Rijsdijk, Serge A.; Hultink, Erik Jan; Diamantopoulos, Adamantios (2007): Product intelligence: its conceptualization, measurement and impact on consumer satisfaction. In *Journal of the Academy of Marketing Science* 35, pp. 340–356.

Rotter, Julian B. (1966): Generalized expectancies for internal versus external control of reinforcement. In *Psychological monographs: General and applied* 80 (1), p. 1.

Rotter, Julian B. (1967): A new scale for the measurement of interpersonal trust. In *Journal of personality*.

Ruiz, Alejandro Pasos; Flynn, Michael; Large, James; Middlehurst, Matthew; Bagnall, Anthony (2021): The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances. In *Data Mining and Knowledge Discovery* 35 (2), pp. 401–449.

Russo, J. Edward; Leclerc, France (1994): An Eye-Fixation Analysis of Choice Processes for Consumer Nondurables. In *Journal of consumer research* 21 (2), p. 274.

Salber, Daniel; Dey, Anind K.; Abowd, Gregory D. (1999): The context toolkit: Aiding the development of context-enabled applications. In: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 434–441.

Salvucci, Dario D.; Goldberg, Joseph H. (2000): Identifying fixations and saccades in eye-tracking protocols. In: Proceedings of the 2000 symposium on Eye tracking research & applications, pp. 71–78.

Schmutz, P.; Roth, S. P.; Seckler, M.; Opwis, K. (2010): Designing product listing pages—Effects on sales and users' cognitive workload. In *International Journal of Human-Computer Studies* 68 (7), pp. 423–431.

Schnack, Alexander; Wright, Malcolm J.; Holdershaw, Judith L. (2021): Does the locomotion technique matter in an immersive virtual store environment?-Comparing motion-tracked walking and instant teleportation. In *Journal of Retailing and Consumer Services* 58, p. 102266.

Schreiber, Jacob (2017): Pomegranate: fast and flexible probabilistic modeling in python. In *the Journal of machine Learning research* 18 (1), 5992--5997.

Schultz, Tanja; Waibel, Alex (2001): Language-independent and language-adaptive acoustic modeling for speech recognition. In *Speech Communication* 35 (1-2), pp. 31–51.

Seabold, Skipper; Perktold, Josef (2010): Statsmodels: Econometric and Statistical Modeling with Python. In: Proceedings of the Python in Science Conference: SciPy.

Seeger, Anna-Maria; Pfeiffer, Jella; Heinzl, Armin (2021): Texting with humanlike conversational agents: Designing for anthropomorphism. In *Journal of the Association for Information Systems* 22 (4), p. 8.

Seitz, Julia; Maedche, Alexander (2022): Biosignal-Based Recognition of Cognitive Load: A Systematic Review of Public Datasets and Classifiers. In *NeuroIS Retreat*, pp. 35–52.

Shen, Bingqing; Tan, Weiming; Guo, Jingzhi; Zhao, Linshuang; Qin, Peng (2021): How to promote user purchase in metaverse? A systematic literature review on consumer behavior research and virtual commerce application design. In *Applied Sciences* 11 (23), p. 11087.

Shim, J. P.; Warkentin, Merrill; Courtney, James F.; Power, Daniel J.; Sharda, Ramesh; Carlsson, Christer (2002): Past, present, and future of decision support technology. In *Decision Support Systems* 33 (2), pp. 111–126.

Shimojo, Shinsuke; Simion, Claudiu; Shimojo, Eiko; Scheier, Christian (2003): Gaze bias both reflects and influences preference. In *Nat Neurosci* 6 (12), pp. 1317–1322. Available online at https://idp.nature.com/authorize/casa?redirect_uri=https://www.nature.com/articles/nn1150&casa_token=utehjyj_ytwaaaaa:rtettuxa6c3gj4lv9z9ee8kek0kd1j78i5ovbk6tsoh0t5nnjql7y8mpuljyizhgnw8d4h6-l2fvglda7q.

Siegel, E. H.; Wei, J.; Gomes, A.; Oliviera, M.; Sundaramoorthy, P.; Smathers, K. et al. (2021): HP Omnicept Cognitive Load Database (HPO-CLD)-Developing a Multimodal Inference Engine for Detecting Real-time Mental Workload in VR. Technical report, HP Labs, Palo Alto.

Sihi, Debika (2018): Home sweet virtual home. In *JRIM* 12 (4), pp. 398–417.

Skinner, Ellen A.; Zimmer-Gembeck, Melanie J.; Connell, James P.; Eccles, Jacquelynne S.; Wellborn, James G. (1998): Individual differences and the development of perceived control. In *Monographs of the society for Research in Child Development*, i-231.

Spagnolo, Fanny; Corsonello, Pasquale; Frustaci, Fabio; Perri, Stefania (2023): Design of a low-power super-resolution architecture for virtual reality wearable devices. In *IEEE Sensors Journal* 23 (8), pp. 9009–9016.

Sriram, G. K. (2022): A Comprehensive Survey on Metaverse. In *International Research Journal of Modernization in Engineering Technology*.

Stefanidi, Zinovia; Margetis, George; Ntoa, Stavroula; Papagiannakis, George (2022): Real-Time Adaptation of Context-Aware Intelligent User Interfaces, for Enhanced Situational Awareness. In *IEEE Access* 10, pp. 23367–23393.

Steil, Julian; Hagestedt, Inken; Huang, Michael Xuelin; Bulling, Andreas (2019): Privacy-aware eye tracking using differential privacy. In: Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications. New York, NY, USA. New York, NY, USA: ACM.

Stephenson, Neal (2003): Snow crash. A Novel: Random House Publishing Group.

Sutherland, Ivan E. (1965): The ultimate display. In *Proceedings of the IFIP Congress* 2, pp. 506–508.

Sweller, John (2011): Cognitive load theory. In: Psychology of learning and motivation, vol. 55: Elsevier, pp. 37–76.

Sykes, Tracy Ann (2015): Support Structures and Their Impacts on Employee Outcomes. In *MIS Quarterly* 39 (2), pp. 473–496.

Szegedy, Christian; Liu, Wei; Jia, Yangqing; Sermanet, Pierre; Reed, Scott; Anguelov, Dragomir et al. (2015): Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9.

Szegedy, Christian; Vanhoucke, Vincent; Ioffe, Sergey; Shlens, Jon; Wojna, Zbigniew (2016): Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818–2826.

Takahashi, Shigeo; Uchita, Akane; Watanabe, Kazuho; Arikawa, Masatoshi (2022): Gaze-driven placement of items for proactive visual exploration. In *Journal of Visualization* 25 (3), pp. 613–633.

Tavakol, Mohsen; Dennick, Reg (2011): Making sense of Cronbach’s alpha. In *International journal of medical education* 2, p. 53.

Tayal, Swati; Rajagopal, Kannan; Mahajan, Vaishali (2022): Virtual reality based metaverse of gamification. In: 2022 6th International Conference on Computing Methodologies and Communication (ICCMC). IEEE, pp. 1597–1604.

Thaler, Richard H.; Sunstein, Cass R. (2021): Nudge: The final edition: Yale University Press.

Todd, P. A.; Benbasat, I. (1994): The influence of decision aids on choice strategies under conditions of high cognitive load. In *IEEE Trans. Syst., Man, Cybern.* 24 (4), pp. 537–547.

Tran, Thi Ngoc Trang; Atas, Müslüm; Felfernig, Alexander; Stettinger, Martin (2018): An overview of recommender systems in the healthy food domain. In *Journal of Intelligent Information Systems* 50 (3), pp. 501–526.

Trzebiński, Wojciech; Marciniak, Beata (2022): Recommender system information trustworthiness: The role of perceived ability to learn, self-extension, and intelligence cues. In *Computers in Human Behavior Reports* 6, p. 100193.

Usmankhujaev, Saidrasul; Ibrokhimov, Bunyodbek; Baydadaev, Shokhrukh; Kwon, Jangwoo (2021): Time Series Classification with InceptionFCN. In *Sensors* 22 (1), p. 157.

Vaessen, Bram E.; Prins, Frans J.; Jeuring, Johan (2014): University students' achievement goals and help-seeking strategies in an intelligent tutoring system. In *Computers & Education* 72, pp. 196–208.

Vallat, Raphael (2018): Pingouin: statistics in Python. In *J. Open Source Softw.* 3 (31), p. 1026.

van de Schoot, Rens; Depaoli, Sarah; King, Ruth; Kramer, Bianca; Märtens, Kaspar; Tadesse, Mahlet G. et al. (2021): Bayesian statistics and modelling. In *Nat Rev Methods Primers* 1 (1), pp. 1–26.

van de Schoot, Rens; Depaoli, Sarah; others (2014): Bayesian analyses: Where to start and what to report. In *The European Health Psychologist* 16 (2), pp. 75–84.

van den Bergh, Don; van Doorn, Johnny; Marsman, Maarten; Draws, Tim; van Kesteren, Erik-Jan; Derks, Koen et al. (2019): A Tutorial on Conducting and Interpreting a Bayesian ANOVA in JASP: Center for Open Science.

van Doorn, Johnny; van den Bergh, Don; Böhm, Udo; Dablander, Fabian; Derks, Koen; Draws, Tim et al. (2021): The JASP guidelines for conducting and reporting a Bayesian analysis. In *Psychon Bull Rev* 28 (3), pp. 813–826.

van Hove, Stephanie; Herregodts, Aron-Levi; Schuurman, Dimitri; Marez, Lieven de (2017): Assessing user experience of context-aware interfaces in a retail store. In: ISPIM Innovation Symposium. The International Society for Professional Innovation Management (ISPIM), p. 1.

Vessey, Iris (1991): Cognitive fit: A theory-based analysis of the graphs versus tables literature. In *Decision sciences* 22 (2), pp. 219–240.

Virtanen, Pauli; Gommers, Ralf; Oliphant, Travis E.; Haberland, Matt; Reddy, Tyler; Cournapeau, David et al. (2020): SciPy 1.0: fundamental algorithms for scientific computing in Python. In *Nature methods* 17 (3), pp. 261–272.

vom Brocke, Jan; Riedl, René; Léger, Pierre-Majorique (2013): Application strategies for neuroscience in information systems design science research. In *Journal of Computer Information Systems* 53 (3), pp. 1–13.

Vos, Henny de; Haaker, Timber; Teerling, Marije; Kleijnen, Mirella (2009): Consumer value of context aware and location based mobile services. In *International Journal of E-Services and Mobile Applications (IJESMA)* 1 (4), pp. 36–50.

Wahl, Deborah R.; Villinger, Karoline; König, Laura M.; Ziesemer, Katrin; Schupp, Harald T.; Renner, Britta (2017): Healthy food choices are happy food choices: Evidence from a real life sample using smartphone based assessments. In *Scientific Reports* 7 (1), p. 17069.

Wang, Chen; Deng, Chengyuan; Wang, Suzhen (2020): Imbalance-XGBoost: leveraging weighted and focal losses for binary label-imbalanced classification with XGBoost. In *Pattern Recognition Letters* 136, pp. 190–197.

Wang, Qiuzhen; Yang, Sa; Liu, Manlu; Cao, Zike; Ma, Qingguo (2014): An eye-tracking study of website complexity from cognitive load perspective. In *Decision Support Systems* 62, pp. 1–10.

Weiß, Tobias (2024): Adaptive user assistance in virtual commerce dataset and source code. Available online at http://osf.io/9jcpn.

Weiß, Tobias; Merkl, Lukas; Pfeiffer, Jella (2023): Customer Decision-Making Processes Revisited: Insights from an Eye Tracking and ECG Study using a Hidden Markov Model. In: Davis, Riedl et al. (Hg.) 2023 – Information Systems and Neuroscience.

Wilcoxon, Frank (1992): Individual comparisons by ranking methods: Springer.

Wilson, John R. (1997): Virtual environments and ergonomics: needs and opportunities. In *Ergonomics* 40 (10), pp. 1057–1077.

Xi, Nannan; Hamari, Juho (2021): Shopping in virtual reality: A literature review and future agenda. In *Journal of Business Research* 134, pp. 37–58.

Xiao; Benbasat (2007): E-Commerce Product Recommendation Agents: Use, Characteristics, and Impact. In *MIS Quarterly* 31 (1), p. 137.

Xiong, Ronglong; Kong, Fanmeng; Yang, Xuehong; Liu, Guangyuan; Wen, Wanhui (2020): Pattern Recognition of Cognitive Load Using EEG and ECG Signals. In *Sensors* 20 (18), p. 5122.

Xu, Songhua; Jiang, Hao; Lau, Francis C. M. (2008): Personalized online document, image and video recommendation via commodity eye-tracking. In: Proceedings of the 2008 ACM conference on Recommender systems, pp. 83–90.

Yang, Fumeng; Huang, Zhuanyi; Scholtz, Jean; Arendt, Dustin L. (2020): How do visual explanations foster end users' appropriate trust in machine learning? In: Proceedings of the 25th international conference on intelligent user interfaces, pp. 189–201.

Yoo, Andy B.; Jette, Morris A.; Grondona, Mark: SLURM: Simple Linux Utility for Resource Management, pp. 44–60.

Yuan, Ying; MacKinnon, David P. (2009): Bayesian mediation analysis. In *Psychological Methods* 14 (4), pp. 301–322.

Zaichkowsky, Judith Lynne (1985): Measuring the Involvement Construct. In *J Consum Res* 12 (3), p. 341.

Zhang, Hong; Lu, Yaobin; Gupta, Sumeet; Zhao, Ling (2014): What motivates customers to participate in social commerce? The impact of technological environments and virtual customer experiences. In *Information & Management* 51 (8), pp. 1017–1030.

Zhang, Xiaohui; Uruchurtu, Elizabeth (2011): A user experience perspective of design for context-aware adaption. In: 2011 Seventh International Conference on Intelligent Environments. IEEE, pp. 322–325.

Zhang, Xuchao; Gao, Yifeng; Lin, Jessica; Lu, Chang-Tien (2020): Tapnet: Multivariate time series classification with attentional prototypical network. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 6845–6852.

Zhang, Zhuo; Shang, Shang; Kulkarni, Sanjeev R.; Hui, Pan (2013): Improving augmented reality using recommender systems. In: Proceedings of the 7th ACM conference on Recommender systems. New York, NY, USA. New York, NY, USA: ACM.

Zhao, Qian; Chang, Shuo; Harper, F. Maxwell; Konstan, Joseph A. (2016): Gaze prediction for recommender systems. In: Proceedings of the 10th ACM Conference on Recommender Systems, pp. 131–138.

Zhao, Wen; Cheng, Yang; Lee, Yen-I (2023): Exploring 360-degree virtual reality videos for CSR communication: An integrated model of perceived control, telepresence, and consumer behavioral intentions. In *Computers in Human Behavior* 144, p. 107736.

5 Paper D: Real agents in virtual commerce

Tobais Weiß, Asmus Eilks, Felix Putze, Jella Pfeiffer and Tanja Schultz

Abstract

Consumers begin to integrate virtual reality (VR) into their daily lives and naturalistic shopping interactions without leaving home are one of the promising use cases for this new technology. However, currently most of the interactions in VR take place in a non-commercial context. To shed light on this lack of virtual commerce adoption, our study uses an iterative software development approach. We evaluate a sales scenario with an avatar-based sales agent that is steered by a human actor. A main feature of our research is the evaluation of different avatars because they facilitate novel, immersive interactions between buyer and seller that differ from well-studied desktop-based e-commerce scenarios. Previous avatar studies have shown that striving for naturalism can lead the avatar to elicit uncanny feelings in the user. Thus, we investigate the severity of the avatar's uncanniness qualitatively and propose the uncanny valley diagram as evaluation tool. In addition to avoiding the uncanny valley effect, our focus is on the timing of the sales agent's interference with the user. We develop a simple rule set that defines when the agent appears, based on gaze patterns. Seventeen participants enter the showroom, evaluate four different 3D printers, receive decision support from our human sales agent avatar, and make a purchase decision. The participants then answer questions about their experience in an interview format. The answers indicate that young consumers value and trust the help provided by digital human agents. In terms of the uncanny valley, the study documents occurred technical challenges, such as motion tracking inaccuracies and face tracking issues, that our participants perceived as uncanny. Regarding the interference timing, participants wanted the agent to appear after they had sufficient time to get an overview of the product assortment.

Keywords: Interference timing, Motion Tracking, Uncanny Valley, Sales Agent Avatar, Virtual Commerce

5.1 Introduction

Avatars are widely understood as digital representations of humans and other entities [15]. Recently, the immersive features of state-of-the-art virtual reality (VR) headsets have added a new level of realism to avatar interactions, entailing growing popularity of social VR applications, such as VRChat [19]. Therefore, the use and impact of avatars is growing, and they are deemed to play a vital role in the transformation of today's Internet and shopping culture. The use of avatars has been studied in various forms of computer-mediated communication [8, 32, 38, 40, 54] and digital representations of the user have long been an integral part of the VR technology [42]. In VR, the communication and interaction between entities can be more naturalistic compared to representations on a desktop computer [26, 60]. With ongoing technological advancements, there are additional technical opportunities, such as face- and eye-tracking, which allow for nonverbal interaction and information exchange between digital communication partners. Recent authors, such as Hennig-Thurau et al. [33], have stimulated the scientific discussion about avatar interaction and have shown that more avatar research is needed to keep pace with technological advancements.

Further recent empirical VR studies that investigated consumer behavior have mainly focused on fast-moving consumer goods [10, 61, 62]. To answer a research call for more product variety in consumer behavior research using VR [80], our shopping scenario frames a purchase situation for a technology product (3D printers) in a virtual commerce showroom, as shown in Figure 23. We chose 3D printers as products, because the purchase decision depends on various criteria and may be rather complex (in comparison to grocery goods). Thus, it is likely that our participants have questions and require help of a sales representative.

Building on previous eye tracking research that exploits visual attention mechanisms to delineate different decision-making subphases and transitions [59, 67], we develop interference timing rules for the digital human sales agent. We choose the uncanny valley effect [50] as one of the main aspects of our qualitative evaluation because results of previous research report the impact of uncanniness on the likeability of the avatar [17]. By incrementally adjusting the technical setup, we aim for achieving a level of avatar fidelity that participants can appreciate. We refine our understanding of how to avoid the uncanny valley, and how to apply a good timing rule set for the appearance of a digital human agent.

The goal of our research is to design, evaluate, and continuously improve a sales interaction between a consumer and a digital human agent. We compare an avatar of a fully motion-tracked sales agent wearing a VR headset to an animated avatar of a sales agent who controls the avatar on a desktop computer in third-person view. Moreover, we evaluate different approaches to represent facial expressions and speech. Where applicable, we document generalizable barriers and boundary factors that limit the adoption of virtual commerce shopping and the use of digital human sales avatars. We iterate through different hard- and software setups and solicit feedback from our participants. By aggregating and presenting their sentiments, we seek to refine guidance for practitioners with similar endeavors. To summarize, we let us guide by the following research questions:

RQ1: How do participants perceive our sales agent avatar in terms of the uncanny valley effect?

RQ2: What simple rule set lets participants appreciate the interference timing of our agent?

Figure 23. Virtual commerce showroom for technology products.

5.2 Theoretical Background

5.2.1 Avatars

Avatars represent different entities in digital environments, usually users and bots [48]. Most avatar definitions assume or imply that the primary purpose of an avatar is to facilitate engagement and interaction of a user with the environment and, more importantly, with other entities [15]. Avatars play a crucial role in virtual worlds and video games because they provide the means to identify with something and allow for embodiment in a virtual space [53]. Depending on the technical effort, avatars can facilitate complex actions, such as nonverbal communication through gestures, posture, proxemics, and even haptic interactions [66].

For e-commerce, various aspects of sales avatars have been studied, such as credibility, social presence, and trust [4, 35, 45, 69, 74]. However, only recently have related avatar studies been conducted in the context of immersive virtual commerce, such as [82]. Most of these previous studies in the immersive virtual commerce context focus on avatars for the consumers themselves [34, 49] or for their peers [36, 75].

5.2.2 The uncanny valley

The concept of the uncanny valley [63], as depicted in Figure 24, refers to a phenomenon in which human-like avatars or robots elicit negative emotional responses from observers because they are not convincingly realistic. The phenomenon was first described and coined by Mori in 1970 [50]. Mori used so-called Bunraku puppets that are a part of a traditional Japanese puppetry show. These puppets are human-like but can have imperfections that might lead a viewer to perceive the puppet as uncanny, eerie, or ghostly.

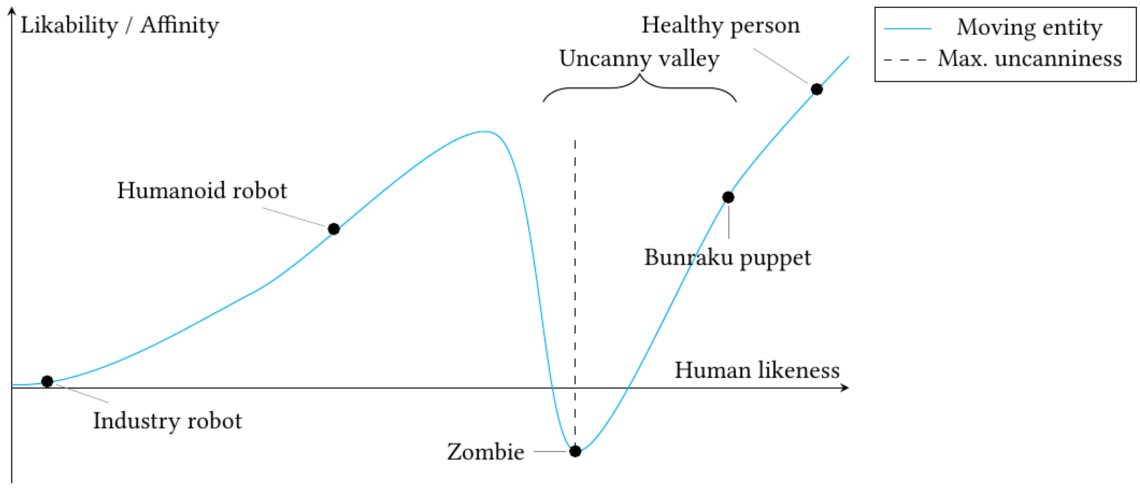

Figure 24. The uncanny valley diagram (Mori 1970).

To explain why the uncanny valley manifests, the categorical uncertainty hypothesis proposes that the discomfort or unease people experience when observing certain human-like entities is due to uncertainty about their categorization [63]. Accordingly, a cognitive conflict occurs when humans encounter entities that appear almost human but have slight imperfections. The categorical uncertainty hypothesis further suggests that the human brain has a natural tendency to categorize and classify objects and beings based on their resemblance to familiar prototypes or stereotypes. If avatars approach a high level of realism but may fall short in some ways, human brains may have difficulties placing them in one category. The avatars may not fit neatly into the human category, yet they may appear human-like enough to raise expectations about human-like behavior that are not fulfilled. This ambiguity may create a cognitive dissonance and trigger a feeling of unease or discomfort.

The categorical uncertainty hypothesis is just one of several theories proposed to explain the uncanny valley phenomenon. Other theories emphasize factors like perceptual mismatch, perceptual familiarity, or violation of human norms [81]. Recent publications argue for a change in the depicted curve and offer explanations based on evolutionary psychology theory and cognitive conflicts [24, 46].

5.3 Method

The iterative design of this qualitative study adopts ideas from agile software development [2] and design science research [56, 57]. As Figure 25 shows, we implement a feedback loop and incrementally refine our virtual environment to gain an understanding of consumer perception of the avatar's uncanniness, interference timing, and other explorative factors.

We collect the data in a multi-location lab-linking setup [71, 73] that closely resembles the technical hurdles of future virtual commerce interactions. For each design and development cycle, we modify and optimize the sales agent avatar and extend the question catalogue if new concepts emerge. In retrospective evaluation meetings after each iteration, we discuss the results, possible technical improvements, and changes needed for the next iteration.

In the feedback loop, we iterate through five design and development cycles with a total of 17 participants, with previous interviews informing the next iteration. To collect participant feedback, we choose an interview format because previous qualitative research in virtual commerce context exists, such as [23, 79], but is underrepresented [78]. To create the question catalogue, we follow the guideline by Kallio et al. [39] for a semi-structured interview. The single steps of the guideline are represented as subprocess of the initial interview design in Figure 25.

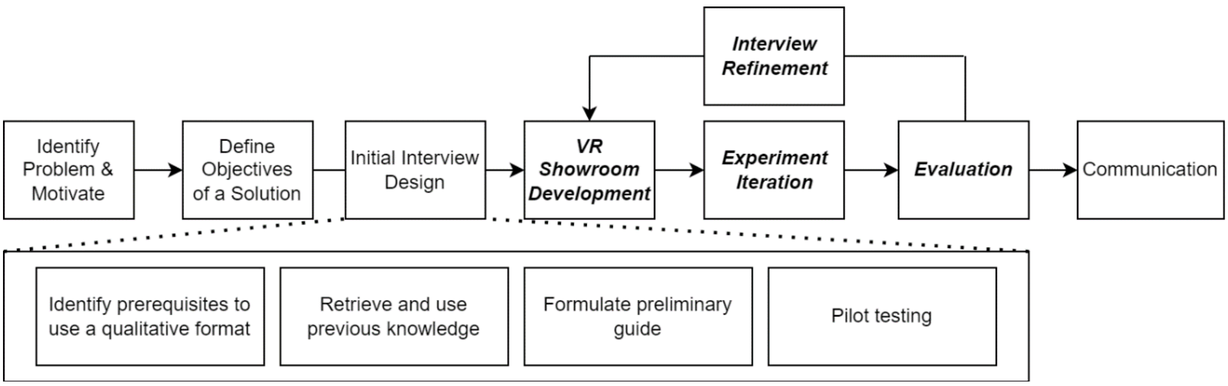

Figure 25. Research design overview (adopted from [57] and [39]).

Findings, explanations, and opinions about the uncanny valley are diverse [24, 46, 81]. We refrain from debating its correctness but use its original visual representation to help our participants during the interview. For a better insight into the participant's feelings towards the avatar, we give them a sheet of paper with the uncanny valley diagram (Figure 24). After making sure that the participants understand the concept, they mark an area on the uncanny valley diagram and describe their perception of the digital human sales agent during their conversation. We then ask our participants to verbalize how they think and feel about the avatar in the experienced scenario. With their answer, we let the participants indicate their attitude towards the avatar and reflect on their reasoning during the decision-making process.

After completing the interviews, we transcribe the audio recordings in automated manner, check for wrong or missing content, and annotate the texts with speaker labels. If words are missing or sentences are obviously wrong, we correct them manually using the audio file. In the next step, we import the corrected and formatted interview texts into the labelling tool Taguette [64] and label them. With the extracted statements grouped into categories, we inform the subsequent research iteration, and finally report our findings.

5.3.1 Showroom environment

For the spatial layout of the environment, we chose two circular rooms connected by a door (see Figure 26). One room represents the showroom where customers enter and evaluate the products; the other room is a waiting room for the agent. Participants can print example 3D objects with timelapse speed to see the visual difference in print quality. This interactive demo printing is one of the virtual showroom features that stand out, in comparison to traditional e-commerce websites.

Initially, the connecting door between the two rooms is closed, so that the consumer and the agent are separated. After one of the appearance criteria is met, the agent opens the door, greets the consumer, and offers help with the purchase decision. The consumer can ask questions, further evaluate the options, and finally choose one of the products. After the sales conversation, the agent guides the consumer to the checkout and helps to complete the purchase.

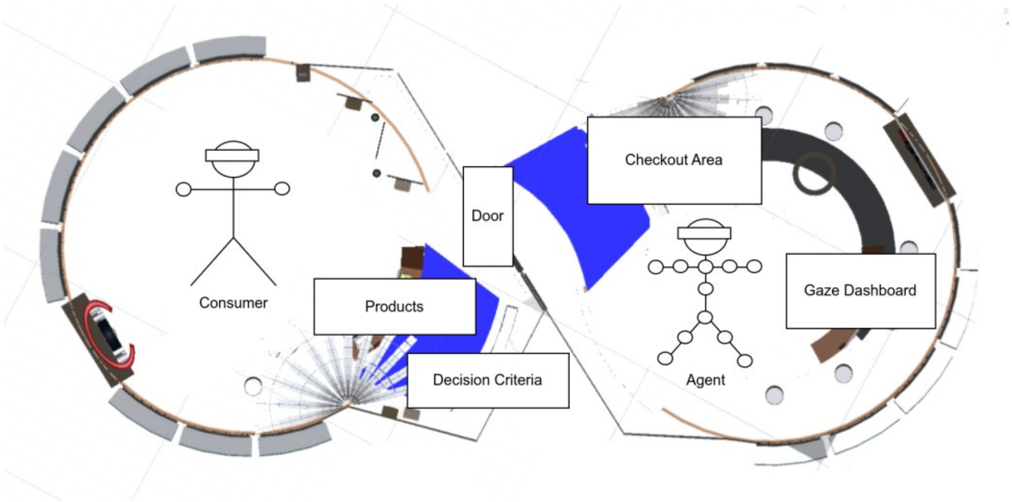

Figure 26. Showroom and agent waiting room layout.

5.3.2 Sales avatars

The sales avatar consists of several components, such as a skeleton, animations, and blend shapes (which we use to animate facial expressions). However, the visually most prominent feature of the avatar is the 3D mesh and its texture. In this study, we evaluate and compare two of the most widely used avatar frameworks currently available, Rocketbox [30] and Readyplayerme [3]. Both avatar types are humanoid and allow for facial expressions. A key difference between the two avatar providers is the customizability. The Rocketbox library only offers a pre-made set of models while Readyplayerme, on the other hand, provides a web interface that allows users to generate an avatar based on a webcam photo and customize it further according to their wishes.

5.3.3 Interference timing via eye tracking

To realize the gaze-informed agent interference timing, we introduce a gaze dashboard for the agent in the waiting room. This gaze dashboard shows the visual attention of participant and tracks the time spent looking at each of the four individual products (see Figure 27 bottom left). The Varjo VR-3 headset has eye tracking sensors that allow to identify the visual attention on the different products via ray casting [1]. We display the aggregated gaze durations on the products in real-time. Watching the accumulating durations of the different products allows the agent to identify comparison patterns and the order in which the customer evaluates the options.

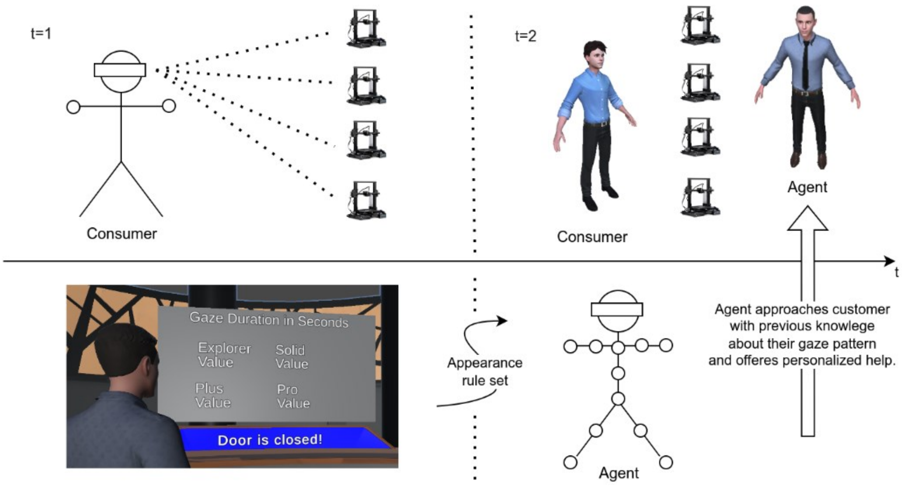

Figure 27. Interference timing of the agent based on the gaze dashboard values.

In terms of decision sub-phases (orientation, evaluation, verification, and purchase [67]), we let the agent start the interaction with the consumer shortly after entering the evaluation phase. We identify the transition from orientation to evaluation by the first pairwise product comparison [59]. In other words, our strategy lets the participants gain an overview of the assortment, and the agent interrupts only after they form an own first impression. This appearance paradigm lets the participants first read product labels and develop questions before the agent appears and offers help.

5.4 Results

The 17 interview recordings contain a total of 587 conversation sections that are relevant for the analysis, comprising a total of 22,000 words. Our labels cover 42 different concepts and the three most frequent ones in descending order are the general conversation topics "Shopping experience", “Sales agent avatar”, and "VR experience". As shown in Figure 28, the frequency for conversation parts that cover the “Uncanny valley” ranks fourth.

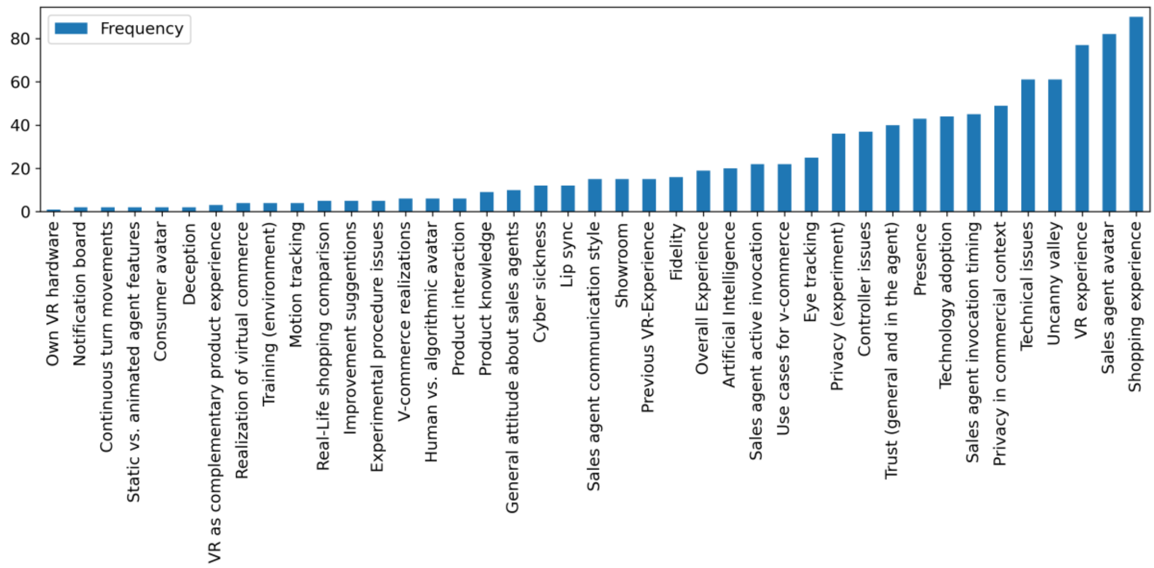

Figure 28. Bar plot of the discussed concepts during the interviews.

In Table 13, we summarize the key features of the interview iterations with the respective attributes of the environment and number of interviews. The ID consists of following abbreviations: Rocketbox (RB), Readyplayerme (RM), Full-body (FB), Third-person (TP), Static-body (SB), Static-face (SF), and Oculus-Lipsync (OC). Figure 29 depicts the categorizations of the sales avatar on the uncanny valley diagram grouped by the single iterations, where full-body motion-tracked iterations are indicated by a circle, third-person steered iterations are indicated by a square, and iteration 5a (RM-SB-OC) is indicated by a triangle.

Table 13. Interview iterations in chronological order.

Order	ID	Avatar provider	Steering mode	Facial expressions via	# Interviews
1	RB-FB-VI	Rocketbox	Full-body motion tracking VR	Vive Facial Cam	2
2	RB-TP-SF	Rocketbox	Third Person on desktop PC	Static face	3
3	RM-FB-SF	Rocketbox	Full-body motion tracking VR	Static face	4
4	RM-TP-OC	Readyplayerme	Third Person on desktop PC	Oculus lip sync	3
5a	RM-SB-OC	Readyplayerme	Static body VR	Oculus lip sync	3
5b	RM-FB-OC	Readyplayerme	Full-body motion tracking VR	Oculus lip sync	2

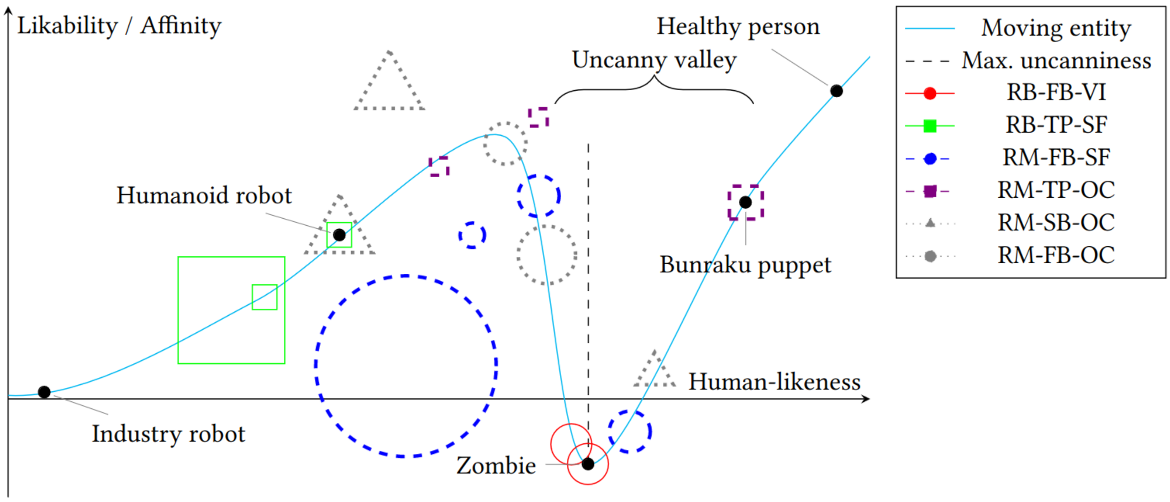

Figure 29. The uncanny valley diagram (Mori 1970) with added participant opinions about the sales agent for the interview iterations.

Overall, our design and development iterations led to a full-body motion-tracked Ready-playerme avatar with audio-based facial expressions (Iteration 5b, RM-FB-OC). Still, full-body motion tracking in VR was not satisfactory because abrupt teleportation movements of the agent drastically reduced the perceived human-likeness of the agent. Particularly at the beginning of the conversation, the agent had to be careful to get the participants' attention before coming close to them using several small distance teleports.

5.4.1 Motion-tracked VR agent with Vive Facial Tracker (RB-FB-VI)

As a starting point to represent our digital human sales agent, we used an avatar from the Rocketbox library that has been applied in previous research setups [20, 24, 45]. The avatar mimics a traditional businessman wearing a black suit and a white shirt with a tie. Both participants perceived the avatar as highly uncanny. One of the participants stated that the avatar was like a zombie, immediately after seeing the uncanny valley depiction (see Figure 29). They indicated that this feeling was mostly caused by glitches, visual imperfections, and inaccuracies in the motion capture system (because we experienced moderate fitting issues with the passive motion tracking markers of the full-body motion tracking suit). The mismatch of the markers resulted in brief periods in which the motion tracking software was unable to correctly map the skeleton or parts of it, such as a single leg. This mismatch resulted in abnormal positions that most likely contributed to the perceived eeriness. Even though both participants experienced further technical issues (wrong floor calibration and issues grabbing things), they felt present in the scene. One participant criticized the scene lighting during this iteration. They said that “[...] the shadows of the figure look pretty scary”. However, both participants perceived the virtual showroom as aesthetically fitting.

For the first participant, the interference timing of the avatar was far from ideal. The sales agent approached the participant by teleporting (too) close but outside of the participant's field of view. This led to a negative first impression and a rough start for the conversation. The participant was startled by the voice that suddenly spoke to them (even if they knew that the agent would appear at some point in time). They also mentioned that the task was not fully clear. For the second participant, the first impression was better, and the appearance and welcoming procedure went smoother compared to the first participant. Regarding the sales agent interference timing, the participant stated: "Maybe for some people, it might be too early, I think. For instance, if they are reading not that fast and they still did not figure out all about the description and what the team wants and what the model has". After explaining our appearance strategy, they added that “it would be more convenient if I have enough time for thinking and considering”.

When asked if they trusted the avatar's guidance, one of our participants said that they could imagine such an avatar as a digital shopping companion. In their opinion, it would be interesting to face a female avatar because women tend to go shopping with their friends. The other participant in this iteration stated that they preferred to inform themselves through reviews and videos instead of interacting with a sales agent.

5.4.2 Agent with static face (RB-TP-SF)

Since the uncanny perception of the agent was driven by the marker mismatch, we decided to modify the motion paradigm. We simplified our setup and animated the avatar instead of using the full-body motion suit and the facial camera. Moreover, we equipped the Rocketbox avatar with simple walking and resting animations. This had the disadvantage that the avatar could no longer use body language or display facial expressions as in the first iteration. We also changed the appearance criteria to account for the feedback of the previous session. We introduced fallback options if the gaze pattern rule was not met for too long. Starting with this iteration, the sales agent joined the consumer after at least one of the following three criteria was met:

(i) Each product viewed for more than 20 seconds, or

(ii) Total time more than two minutes, or

(iii) The participant called for help or had obvious problems.

The regions on the uncanny valley diagram that the participants of this group indicated clustered around the medium human likeness and medium affinity regions. The mostly idle avatar was perceived as quite unrealistic, and participants noted the lack of facial and body animation as one core issue. However, due to this decidedly artificial appearance, the avatar was not considered uncanny. Participants clearly stated that they perceived the avatar as “not in the problematic spectrum” and “definitely not [as] a zombie or corpse” as in the first iteration.

For two of the participants, the appearance of the sales agent was too early. One of them made the discrepancy more palpable, adding that they would have liked approximately ten more seconds on their own.

Participants felt present in the showroom and reported that they were mostly unaware of the outside world during the experience, aside from minor environmental noise that sporadically distracted them. Overall, they expressed positive general feedback and all comments about their intentions to use similar virtual commerce environments were positive. Participants also stated that they trusted the avatar because it embodied a real human agent. However, they were concerned about a potentially fully digital artificial intelligence (AI) agent. They argued that such an entity might not be fully adjusted to their personal needs and instead trained to maximize sales rather than provide ideal consultation.

5.4.3 Motion-tracked VR sales agent with static face (RM-FB-SF)

Following the improvements regarding the perceived uncanniness in the animated iteration, we decided to reintroduce full-body motion tracking, to investigate whether the removal of full-body motion tracking or facial tracking caused the reduced uncanniness. Furthermore, after analyzing the avatar feedback, we substituted the Rocketbox with Readyplayerme avatars. Apart from more compatible blend shapes, changing the avatar framework provided additional benefits: While we had to choose from a set of predefined avatars using the Rocketbox library, the Readyplayerme API allowed us to create a personalized avatar from a webcam photo, and we further customized the avatar appearance from a set of predefined outfits and accessories.

We slightly adjusted the session protocol and instructed the actor to be more aware about the startling effect of teleport movements. We advised the actor to clearly teleport into the consumer's field of view at an appropriate distance (especially not too close) to improve the consumer's first impression. As the topic came up during previous iterations, we added a question about trust in a hypothetical AI agent to our question catalogue.

Participants stated that they perceived the avatar as static, somewhere between human and robot, but closer to the human. Compared to the first iteration, participants perceived this avatar as less uncanny. None of the participants stated that they were scared, though one participant was startled by the avatar’s appearance, as they did not notice how the agent entered the room. Most participants stated that they felt present in the virtual environment during the experience, although one clearly remarked that they were always fully aware of being in a real room.

All three answers regarding the sales agent's appearance indicated that the interference timing rule set worked as intended. In one of the sessions, the actor applied rule (iii) as the consumer was asking themselves questions. One participant perceived the appearance as slightly too early and said, “I could have easily just watched by myself for another short moment.”

In terms of perceived trust, one participant stated that they would trust the avatar less than a human sales agent in the real world. They argued that they would be concerned that the person steering the avatar might be hiding their intentions. The third participant stated that they would trust an AI agent more, mentioning that the AI could eventually have a broader knowledge base than a human and could therefore be more helpful with facts.

5.4.4 Third-person sales agent with Oculus lip sync (RM-TP-OC)

Since participants criticized the static nature of the avatar's face as major issue during the previous iteration, we replaced the Vive Facial Cam that we used in the first iteration with viseme-based facial expressions using the Oculus lip sync framework. Visemes are audio based and, thus, have the benefit of not needing a facial camera but are less accurate. They can be thought as different extreme facial expressions and mouth shapes that we blend to make the avatar look like it is talking.

We applied the interference timing rule set as in the previous iterations, and again, participants perceived the salesperson's appearance as appropriately timed. They consistently reported that they were not scared or frightened by the avatar but perceived it more as robot or character in a video game. It therefore appears that while the visemes did not elicit uncanny feelings, they also did not substantially improve how realistic the avatar appeared.

As in the previous iteration, participants stated that they trusted the guidance of the agent but indicated that they would do so less in the case of an AI-based agent. One participant uttered the rationale that they value the subjective experiences a human salesperson can share compared to an AI agent. For widespread adoption of such a system, one participant had concerns that it would be easy for sellers to hide product flaws in VR.

5.4.5 Static VR sales agent with Oculus lip sync (RM-SB-OC)

Following the fourth iteration with successful lip sync for the sales agent controlled in third-person view, we evaluated the viseme-based approach in combination with full-body motion tracking. By doing so, we tried to minimize uncanny artifacts that we experienced with the Vive Facial Cam that we used in the first iteration.

A technical issue caused that the avatar stood statically without any movement despite the lip sync and teleportation. Since the planned interviews took place sequentially on the same day, we decided to continue with the sessions and collected the remaining observations for this day with the static avatar setup.

For this round, our appearance rule set yielded acceptable results because three out of four participants perceived the agent's interference timing as good. For the remaining participant, the interference timing was too early. Upon interference, they even told the sales agent to wait, and only after more than another minute they started to ask questions. However, we were not able to derive a meaningful general rule for their case.

Participants did not perceive the avatar as uncanny, stating that the static nature made it too unrealistic. Two out of three participants experienced minor controller issues but all of them stated that they felt fully present in the virtual scene and that they were not aware of the outside world anymore. Moreover, two participants stated that they trusted the agent's guidance and intentions and that they believed they would do the same if an AI controlled the agent. Interestingly, the third participant stated that the static avatar seemed too unrealistic to be trusted. So, realism and trust in virtual commerce settings may be correlated.

One participant noted that, while they liked this scenario, they would certainly not enjoy virtual commerce shopping for every product category. As an example, they stated that they would not enjoy a virtual environment where they would have to move around to pick up small products as in a grocery store. Another participant stated that, compared to e-commerce, they appreciated the ability to look at products from all angles and in the correct proportions.

5.4.6 Motion-tracked VR sales agent with Oculus lip sync (RM-SB-OC)

For this iteration, we solved the previous tracking issues and evaluated the full-body motion tracking with the viseme-based Oculus lip sync. One participant perceived the avatar as human-like and stated that the avatar's use of gestures aided their trust in the agent. The other participant differed in their opinion, stating that the avatar was not creepy, but more akin to a robot than a human. The respective mark on the diagram landed slightly closer to the uncanny region than we intended.

Still, this iteration yielded positive feedback for our appearance rule set and both participants stated that the avatar came to the showroom to assist them with good timing. The participants noted that they felt present in the scene, not noticing much of the outside world. They stated that they had minor issues with the controller handling but were overall able to navigate and interact.

One participant stated reservations about buying more expensive goods in VR, arguing that they would not trust a simulation. For higher-priced goods, they would still like to see and evaluate the devices in real life. Both participants stated that they trusted the agent's guidance and would trust an AI less than a digital human agent. One participant remarked that they would not trust the agent in general without own research about the product. Overall, both participants said that they enjoyed the experience and that they would like to use similar systems in the future.

5.5 Discussion

5.5.1 Uncanny valley effect

Regarding the uncanny valley effect (RQ1), our interviews reflect a promising improvement across the iterations. In our early attempts, technical issues were clearly a major contributor to the perceived uncanniness, and in the very first iteration we landed right at the bottom of the uncanny valley. Our final iteration contrasts this, in which our participants marked the agent’s appearance in a desirable region of the uncanny valley diagram that is close to but not within the critical uncanny valley. Participants perceived full-body motion-tracked avatars steered in VR as superior to the animated avatars. In other words, participants perceived the animated avatars that were steered in third person view on a desktop computer as robotic, and facial tracking did not change the robotic emanation. For instance, one participant said “[...] it's not exactly human […] it was a bit robotic.” As further improvement, we suggest switching to another movement paradigm for the teleport action (as instant teleportation of the agent confused some of the participants) or using continuous movements with inverse kinematics [14].

5.5.2 Interference timing

For the sales agent’s interference timing (RQ2), we first started with the simple rule to wait until the participant inspected all products for at least ten seconds. Already with a limited number of observations, it became clear that one single threshold value would hardly be enough to satisfy the preferences of a wide range of consumers. Already in the first iteration, the first participant suggested that this criterion may lead to premature interference of the sales agent. To incorporate this feedback, we implemented a slightly more advanced visual representation of the participant’s gaze. We provided a dashboard to the sales agent that showed the gaze time per product in real time. We also changed the overall timing and the question catalogue. When the actors applied this slightly more advanced rule set, most participants perceived the interference timing as good. The gaze time per product may be reduced from 20 seconds to 10 seconds because two of them indicated that the sales agent could have appeared a bit earlier ("about 10 seconds earlier", "a bit too long waiting time for me"). Overall, the interviews suggest that we defined an adequate and still simple rule set, which may be adapted and further individualized dependent on the context at hand.

5.5.3 Other hindrances and boundary factors

With exceptions, the participants stated that they would trust a human sales agent. When asked about the importance of motion and face fidelity to foster trust, participants found body language more important than lip movements. They would further trust a digital human sales agent more than an AI, but less than an in-person interaction with a salesperson. Participants were primarily concerned that a malicious sales agent could mimic persuasive body language and fake social cues. The prevailing opinion emphasized the importance of the human element that still elevates trust, both in motion fidelity and decision support. The actual perceived trust of our participants did not noticeably change between iterations, what implies that a human sales agent, fully body tracked or not, fosters consumers' trust in our showroom. It remains an open research question whether a human-like appearance and movement, that goes beyond the one implemented in this work, can overcome the uncanny valley completely but for now it remains highly recommendable to aim for its lower end. We conclude that of whether the sales agent is a real human, or a fully automated AI algorithm had the greatest impact on our participants' perceived trust, and that the preference for or against a human versus AI agent was highly individual.

Privacy awareness was mostly present, but participants had few concerns about ET data. Twelve out of the seventeen participants indicated that they either did not view ET data as important or that they did not have strong feelings about their ET data being collected. Particularly for scholastic purposes, they saw no problem in sharing their ET data and behavioral data. In the interviews of later iterations, we asked participants how they consider ET data in comparison to their spatial location. Of the eleven participants to whom we asked this question, seven responded that they considered the spatial location more critical than gaze data, while only one participant considered ET data more critical. Six participants noted that, while they were comfortable with companies collecting their ET data, it was important to them that the company stores their data securely. One participant noted that they would like to have gaze dashboard for themselves, so that they could see the data to reflect their own attention and factor it into their purchasing decision.

When asked whether participants would have preferred to call the sales representative manually, we found mixed results. With eight participants, a slight majority of our participants favored the active agent invocation via button. The preference for or against automatic interference may depend on the general sales attitudes [31] and further impact factors. We call for future quantitative studies that answer the question of whether consumers prefer active invocation via button/speech recognition or automatic (passive) interference.

5.6 Conclusion

Our interviews indicate that agents are relevant to virtual commerce, and this paper shows that we can reduce their uncanniness during five design and development iterations, without overcoming the issue completely. Our findings serve as a first indicator because they reflect the opinions of a limited number of people with similar demographic backgrounds. We outline the role of digital human sales agents as a key aspect of future virtual commerce and provides guidance for their implementation, for example by documenting the differences between motion-tracking VR and third-person desktop-controlled avatars. Our interviews shed light on consumer perceptions of digital human salespeople, and the iteration summaries can guide researchers and practitioners in designing similar environments. The results may also be of interest to other domains. For example, avatars and avoiding the uncanny valley are also applicable to tutors in educational and instructors in professional industry training. To substantiate our claims, it would be necessary to confirm the results of the interviews with a quantitative experiment. Further research opportunities are to extend the gaze dashboard or to evaluate different interference timing rules quantitatively.

References

[1] Alghamdi, N. and Alhalabi, W. 2019. Fixation detection with ray-casting in immersive virtual reality. *International Journal of Advanced Computer Science and Applications* 10, 7.

[2] Al-Saqqa, S., Sawalha, S., and AbdelNabi, H. 2020. Agile software development: Methodologies and trends. *International Journal of Interactive Mobile Technologies* 14, 11.

[3] Altundas, S. and Karaarslan, E. 2023. Cross-platform and Personalized Avatars in the Metaverse: Ready Player Me Case. In *Digital Twin Driven Intelligent Systems and Emerging Metaverse*. Springer, 317–330.

[4] Arakji, R. Y. and Lang, K. R. 2008. Avatar business value analysis: a method for the evaluation of business value creation in virtual commerce. *Journal of Economic Commerce Research*.

[5] Barthes, R. and MacDonald, S. 2013. On bunraku. In *Popular Theatre*. Routledge, 49–52.

[6] Bear, H. L. and Harvey, R. 2017. Phoneme-to-viseme mappings: the good, the bad, and the ugly. *Speech Communication* 95, 40–67.

[7] Benford, S., Bowers, J., Fahlén, L. E., Greenhalgh, C., and Snowdon, D. 1997. Embodiments, avatars, clones and agents for multi-user, multi-sensory virtual worlds. *Multimedia Systems* 5, 93–104.

[8] Beyea, D., van der Heide, B., Ewoldsen, D., Eden, A., and Meng, J. 2021. Avatar-based self-influence in a traditional CMC environment. *Journal of Media Psychology: Theories, Methods, and Applications*.

[9] Bianchini, G. E., Zanotti, L., and Meléndez, C. 2023. Using OpenAI models as a new tool for text analysis in political leaders' unstructured discourse.

[10] Bigne, E., Llinares, C., and Torrecilla, C. 2016. Elapsed time on first buying triggers brand choices within a category: A virtual reality-based study. *Journal of Business Research* 69, 4, 1423–1427.

[11] Bock, O., Baetge, I., and Nicklisch, A. 2014. hroot: Hamburg registration and organization online tool. *European Economic Review* 71, 117–120.

[12] Bourlakis, M., Papagiannidis, S., and Li, F. 2009. Retail spatial evolution: paving the way from traditional to metaverse retailing. *Electronic commerce research* 9, 135–148.

[13] Cao, Q., Yu, H., Charisse, P., Qiao, S., and Stevens, B. 2023. Is High-Fidelity Important for Human-like Virtual Avatars in Human Computer Interactions? *International Journal of Network Dynamics and Intelligence*, 15–23.

[14] Caserman, P., Achenbach, P., and Göbel, S. 2019. Analysis of inverse kinematics solutions for full-body reconstruction in virtual reality. In *2019 IEEE 7th International Conference on Serious Games and Applications for Health (SeGAH)*, 1–8.

[15] Castronova, E. 2003. *Theory of the Avatar*. CESifo.

[16] Chung, Y.-Y., Annaswamy, T. M., and Prabhakaran, B. 2022. Design of calibration module for a home-based immersive game using camera motion capture system. In *Proceedings of the 2022 ACM Symposium on Spatial User Interaction*, 1–2.

[17] Ciechanowski, L., Przegalinska, A., Magnuski, M., and Gloor, P. 2019. In the shades of the uncanny valley: An experimental study of human-chatbot interaction. *Future Generation Computer Systems* 92, 539–548.

[18] Cummings, J. J. and Bailenson, J. N. 2016. How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychology* 19, 2, 272–309.

[19] Deighan, M. T., Ayobi, A., and O'Kane, A. A. 2023. Social Virtual Reality as a Mental Health Tool: How People Use VRChat to Support Social Connectedness and Wellbeing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–13.

[20] Dwivedi, Y. K., Hughes, L., Baabdullah, A. M., Ribeiro-Navarrete, S., Giannakis, M., Al-Debei, M. M., Dennehy, D., Metri, B., Buhalis, D., Cheung, C. M. K., and others. 2022. Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management* 66, 102542.

[21] Engel, J. F., Kollat, D. T., and Blackwell, R. D. 1968. A model of consumer motivation and behavior. *Research in Consumer Behavior, Holt, Rinehart and Winston, Inc., New York, NY*, 3–20.

[22] Evans, P. and Wurster, T. S. 1999. Getting real about virtual commerce. *Harvard Business Review* 77, 84–98.

[23] Farah, M. F., Ramadan, Z. B., and Harb, D. H. 2019. The examination of virtual reality at the intersection of consumer experience, shopping journey and physical retailing. *Journal of Retailing and Consumer Services* 48, 136–143.

[24] Ferrey, A. E., Burleigh, T. J., and Fenske, M. J. 2015. Stimulus-category competition, inhibition, and affective devaluation: a novel account of the uncanny valley. *Frontiers in psychology* 6, 249.

[25] Fox, J., Ahn, S. J., Janssen, J. H., Yeykelis, L., Segovia, K. Y., and Bailenson, J. N. 2015. Avatars versus agents: a meta-analysis quantifying the effect of agency on social influence. *Human-Computer Interaction* 30, 5, 401–432.

[26] Freeman, G. and Maloney, D. 2021. Body, avatar, and me: The presentation and perception of self in social virtual reality. *Proceedings of the ACM on human-computer interaction* 4, CSCW3, 1–27.

[27] Freeman, G., Zamanifard, S., Maloney, D., and Adkins, A. 2020. My body, my avatar: How people perceive their avatars in social virtual reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–8.

[28] Gidlöf, K., Wallin, A., Dewhurst, R., and Holmqvist, K. 2013. Using Eye Tracking to Trace a Cognitive Process: Gaze Behaviour During Decision Making in a Natural Environment. *Journal of Eye Movement Research* 6, 1.

[29] Gil-López, C., Guixeres, J., Mar\'\in-Morales, J., Torrecilla, C., Williams, E., and Alcañiz, M. 2023. Is mixed reality technology an effective tool for retail? A vividness and interaction perspective. *Frontiers in virtual reality* 4, 1067932.

[30] Gonzalez-Franco, M., Ofek, E., Pan, Y., Antley, A., Steed, A., Spanlang, B., Maselli, A., Banakou, D., Pelechano, N., Orts-Escolano, S., and others. 2020. The rocketbox library and the utility of freely available rigged avatars. *Frontiers in virtual reality* 1, 20.

[31] Guo, R., Izzouzi, L., and Steed, A. 2023. User Motion Accentuation in Social Pointing Scenario. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 575–576.

[32] Han, E., Miller, M. R., DeVeaux, C., Jun, H., Nowak, K. L., Hancock, J. T., Ram, N., and Bailenson, J. N. 2023. People, places, and time: a large-scale, longitudinal study of transformed avatars and environmental context in group interaction in the metaverse. *Journal of Computer-Mediated Communication* 28, 2, zmac031.

[33] Hennig-Thurau, T., Aliman, D. N., Herting, A. M., Cziehso, G. P., Linder, M., and Kübler, R. V. 2023. Social interactions in the metaverse: Framework, initial evidence, and research roadmap. *Journal of the Academy of Marketing Science* 51, 4, 889–913.

[34] Holzwarth, M., Janiszewski, C., and Neumann, M. M. 2006. The influence of avatars on online consumer shopping behavior. *Journal of Marketing* 70, 4, 19–36.

[35] Hwang, S. and Shin, S. 2013. The Effects of the Virtual Avatar Fitting Models for Apparel e-Commerce in Consumer’s Purchasing Behavior: Comparing Traditional Model with Virtual Avatar Model. *Journal of Fashion Business* 17, 5, 57–69.

[36] Izadi, B., Dong, L., and Esfidani, M. R. 2021. Stay home and shop together. *Journal of Electronic Commerce Research* 22, 1, 59–75.

[37] Izzouzi, L. and Steed, A. 2022. Integrating rocketbox avatars with the ubiq social vr platform. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 69–70.

[38] Jacobsen, L. F., Mossing Krogsgaard-Jensen, N., and Peschel, A. O. 2022. Shopping in Reality or Virtuality? A Validation Study of Consumers' Price Memory in a Virtual vs. Physical Supermarket. *Foods* 11, 14, 2111.

[39] Kallio, H., Pietilä, A.-M., Johnson, M., and Kangasniemi, M. 2016. Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *Journal of advanced nursing* 72, 12, 2954–2965.

[40] Kang, H.-S. and Yang, H.-D. 2006. The visual characteristics of avatars in computer-mediated communication: Comparison of Internet Relay Chat and Instant Messenger as of 2003. *International Journal of Human-Computer Studies* 64, 12, 1173–1183.

[41] Kourtesis, P., Linnell, J., Amir, R., Argelaguet, F., and MacPherson, S. E. 2023. Cybersickness in Virtual Reality Questionnaire (CSQ-VR): A Validation and Comparison against SSQ and VRSQ. *Virtual Worlds* 2, 1, 16–35.

[42] Lanier, J. 2017. *Dawn of the new everything: A journey through virtual reality*. Random House.

[43] Lee, Y. J. and Dubinsky, A. J. 2017. Consumers' desire to interact with a salesperson during e-shopping: development of a scale. *International Journal of Retail & Distribution Management* 45, 1, 20–39.

[44] Lewis, J. P., Anjyo, K., Rhee, T., Zhang, M., Pighin, F. H., and Deng, Z. 2014. Practice and theory of blendshape facial models. *Eurographics (State of the Art Reports)* 1, 8, 2.

[45] Liew, T. W., Tan, S.-M., and Ismail, H. 2017. Exploring the effects of a non-interactive talking avatar on social presence, credibility, trust, and patronage intention in an e-commerce website. *Human-centric Computing and Information Sciences* 7, 1–21.

[46] Mathur, M. B., Reichling, D. B., Lunardini, F., Geminiani, A., Antonietti, A., Ruijten, P. am, Levitan, C. A., Nave, G., Manfredi, D., Bessette-Symons, B., and others. 2020. Uncanny

but not confusing: Multisite study of perceptual category confusion in the Uncanny Valley. *Computers in Human Behavior* 103, 21–30.

[47] Meier, M., Mason, C., Porzel, R., Putze, F., and Schultz, T. 2018. Synchronized Multimodal Recording of a Table Setting Dataset. In *IROS 2018: Workshop on Latest Advances in Big Activity Data Sources for Robotics & New Challenges*, Madrid, Spain.

[48] Miao, F., Kozlenkova, I. V., Wang, H., Xie, T., and Palmatier, R. W. 2022. An emerging theory of avatar marketing. *Journal of Marketing* 86, 1, 67–90.

[49] Moon, J. H., Kim, E., Choi, S. M., and Sung, Y. 2013. Keep the social in social media: The role of social interaction in avatar-based virtual shopping. *Journal of Interactive Advertising* 13, 1, 14–26.

[50] Mori, M. 1970. The uncanny valley: the original essay by Masahiro Mori. *IEEE Spectrum* 6.

[51] Mukashev, D., Kairgaliyev, M., Alibekov, U., Oralbayeva, N., and Sandygulova, A. 2021. Facial expression generation of 3D avatar based on semantic analysis. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, 89–94.

[52] Nagymáté, G. and Kiss, R. M. 2018. Application of OptiTrack motion capture systems in human movement analysis: A systematic literature review. *Recent Innovations in Mechatronics* 5, 1, 1–9.

[53] Nardi, B. 2010. *My life as a night elf priest: An anthropological account of World of Warcraft*. University of Michigan Press.

[54] Nowak, K. L. and Fox, J. 2018. Avatars and computer-mediated communication: a review of the definitions, uses, and effects of digital representations. *Review of Communication Research* 6, 30–53.

[55] Parger, M., Mueller, J. H., Schmalstieg, D., and Steinberger, M. 2018. Human upper-body inverse kinematics for increased embodiment in consumer-grade virtual reality. In *Proceedings of the 24th ACM symposium on virtual reality software and technology*, 1–10.

[56] Peffers, K., Rothenberger, M., Tuunanen, T., and Vaezi, R. 2012. Design science research evaluation. In *Design Science Research in Information Systems. Advances in Theory*

and Practice: 7th International Conference, DESRIST 2012, Las Vegas, NV, USA, May 14-15, 2012. Proceedings 7, 398–410.

[57] Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. 2007. A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems* 24, 3, 45–77.

[58] Perl, O. 2020. *Distributed Multi-User VR With Full-Body Avatars,* Technische Universität Wien.

[59] Peukert, C., Lechner, J., Pfeiffer, J., and Weinhardt, C. 2020. Intelligent Invocation: Towards Designing Context-Aware User Assistance Systems Based on Real-Time Eye Tracking Data Analysis. In *Information Systems and Neuroscience*. Springer, Cham, 73–82.

[60] Peukert, C., Pfeiffer, J., Meißner, M., Pfeiffer, T., and Weinhardt, C. 2019. Shopping in virtual reality stores: the influence of immersion on system adoption. *Journal of Management Information Systems* 36, 3, 755–788.

[61] Peukert, C., Pfeiffer, J., Meißner, M., Pfeiffer, T., and Weinhardt, C. 2019. Shopping in virtual reality stores: The influence of immersion on system adoption. *Journal of Management Information Systems* 36, 3, 755–788.

[62] Pfeiffer, J., Pfeiffer, T., Meißner, M., and Weiß, E. 2020. Eye-tracking-based classification of information search behavior using machine learning: evidence from experiments in physical shops and virtual reality shopping environments. *Information Systems Research* 31, 3, 675–691.

[63] Ramey, C. H. 2006. An inventory of reported characteristics for home computers, robots, and human beings: Applications for android science and the uncanny valley. In *Proceedings of the ICCS/CogSci-2006 Long Symposium 'Toward Social Mechanisms of Android Science', Vancouver, Canada*.

[64] Rampin, R. and Rampin, V. 2021. Taguette: open-source qualitative data analysis. *Journal of Open Source Software* 6, 68, 3522.

[65] Reysen, S. 2005. Construction of a new scale: The Reysen likability scale. *Social Behavior and Personality: an international journal* 33, 2, 201–208.

[66] Rogers, S. L., Broadbent, R., Brown, J., Fraser, A., and Speelman, C. P. 2022. Realistic motion avatars are the future for social interaction in virtual reality. *Frontiers in virtual reality* 2, 750729.

[67] Russo, J. E. and Leclerc, F. 1994. An eye-fixation analysis of choice processes for consumer nondurables. *Journal of Consumer Research* 21, 2, 274–290.

[68] Ruttan, R. L. and Lucas, B. J. 2018. Cogs in the machine: The prioritization of money and self-dehumanization. *Organizational Behavior and Human Decision Processes* 149, 47–58.

[69] Scarle, S., Arnab, S., Dunwell, I., Petridis, P., Protopsaltis, A., and Freitas, S. de. 2012. E-commerce transactions in a virtual environment: virtual transactions. *Electronic commerce research* 12, 379–407.

[70] Schroeder, R. 2002. Social interaction in virtual environments: Key issues, common themes, and a framework for research. In *The social life of avatars: Presence and interaction in shared virtual environments*. Springer, 1–18.

[71] Schultz, T., Putze, F., Fehr, T., Meier, M., Mason, C., Ahrens, F., and Herrmann, M. 2021. Linking Labs: Interconnecting Experimental Environments. *arXiv preprint arXiv:2102.03684*.

[72] Schultze, U. and Avital, M. 2011. Designing interviews to generate rich data for information systems research. *Information and Organization* 21, 1, 1–16.

[73] Steed, A., Archer, D., Brandstätter, K., Congdon, B. J., Friston, S., Ganapathi, P., Giunchi, D., Izzouzi, L., Park, G. W. W., Swapp, D., and others. 2023. Lessons learnt running distributed and remote mixed reality experiments. *Frontiers in Computer Science* 4, 966319.

[74] Tan, S.-M. and Liew, T. W. 2020. Designing embodied virtual agents as product specialists in a multi-product category E-commerce: The roles of source credibility and social presence. *International Journal of Human-Computer Interaction* 36, 12, 1136–1149.

[75] Tawira, L. and Ivanov, A. 2023. Leveraging personalization and customization affordances of virtual try-on apps for a new model in apparel m-shopping. *Asia Pacific Journal of Marketing and Logistics* 35, 2, 451–471.

[76] Viksne, K., Salkovska, J., Gaitniece, E., and Puke, I. 2016. Comparative analysis of customer behaviour models. In *Proceedings of the 2016 International conference on Economic Science for Rural Development*, 231.

[77] Weiss, T., Merkl, L., Pfeiffer, J., and Weinhardt, C. 2023. Customer Decision-Making Processes Revisited: Insights from an Eye Tracking and ECG Study using a Hidden Markov Model. In *Information Systems and Neuroscience: NeuroIS Retreat 2023*.

[78] Xi, N. and Hamari, J. 2021. Shopping in virtual reality: A literature review and future agenda. *Journal of Business Research* 134, 37–58.

[79] Xue, L., Parker, C. J., and Hart, C. A. 2020. How to engage fashion retail with virtual reality: a consumer perspective. *Augmented Reality and Virtual Reality: Changing Realities in a Dynamic World*, 23–35.

[80] Yoo, K., Welden, R., Hewett, K., and Haenlein, M. 2023. The merchants of meta: A research agenda to understand the future of retailing in the metaverse. *Journal of Retailing*.

[81] Zhang, J., Li, S., Zhang, J.-Y., Du, F., Qi, Y., and Liu, X. 2020. A literature review of the research on the uncanny valley. In *Cross-Cultural Design. User Experience of Products, Services, and Intelligent Environments: 12th International Conference, CCD 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19-24, 2020, Proceedings, Part I 22*, 255–268.

[82] Zhu, S., Hu, W., Li, W., and Dong, Y. 2023. Virtual Agents in Immersive Virtual Reality Environments: Impact of Humanoid Avatars and Output Modalities on Shopping Experience. *International Journal of Human-Computer Interaction*, 1–23.

5.7 Supplementary material

5.7.1 Cover story

Please imagine the following scenario:

You and a team of fellow students develop a board game idea. The team decides to put the idea into practice and builds a prototype. In a collaborative effort, you design the game pieces in a 3D software. Now you want to evaluate producing these models.

For the production, your team decides to purchase a 3D printer. However, an abundance of different printer variants exists. We offer you the opportunity to evaluate different 3D printers in a virtual environment. Now, you put on VR glasses and enter our showroom. In the virtual environment, you see decision criteria that your team considers important. Furthermore, you see several 3D printers with their properties. Your task is to choose the right product. Before the experience, we asked a group of people to agree on which printer is the right choice for your team. Do you have any questions?

5.7.2 Agent knowledge

During the experience, we ask the consumer to choose one of the products, given the criteria shown in Figure 30. The products look as depicted in Figure 31. The agent approaches the consumer, welcomes them to the environment, and offers to answer questions if the consumer has some. Then, the agent moves next to the product table (right, in sight of the consumer when evaluating the products) and waits for requests. When the participant has made their decision, the agent accompanies the consumer to the checkout and explains how to finalize the purchase.

Figure 30. Decision criteria.

Figure 31. Products.

The agent has the following knowledge regarding the product criteria:

- Easy device setup: We offer two different types, ready-made and self-assembly kits. If participants ask, advise them to go for the ready-made variant. Self-assembly kits are rather for hobbyists and the time it takes is not economical.
- PETG material printable: Three types of print material (so called “filament”) are common: ABS, PLA, and PETG. The full material names are not important. Instead, it is important that the printed miniatures are durable, i.e., do not break during fierce gaming sessions. First, ABS requires very high printing and heat-bed temperatures. It is also recommended to use an additional cover to prevent the model from deforming during the print process. Overall, the cost-benefit ratio is not great. Second, PLA, which is the cheapest and easiest to work with material, is not an option because it is not durable enough. Figures would tend to break on frequent usage. The third option, PETG, combines the strengths of ABS and PLA. While it only requires a moderate printing temperature, the resulting objects are durable and robust.
- High print quality: The print quality can be judged by the printed models. A high quality is represented by a high-poly rabbit and a low quality is represented by a low-poly rabbit. The two models “Explorer” and “Solid” have low quality. The agent should encourage the participants to try out the printers themselves by pressing the print button. The agent can also press the button themselves if required.
- Fast print speed: All models, except the “Explorer” model print in fast mode (approx. 10 sec.). Again, the agent should encourage participants to try it themselves.

- The device should not catch fire: Fire protection certification after the DIN norm is not necessarily required. The certified “Pro” product still can catch fire. Certification is no 100\% guarantee. For all products, consumers are strongly advised to place the printer on a non-flammable underground, i.e., concrete. For the participants, certification should not be a major decision-making criterion. They should make sure that the device does not catch fire by the mentioned safety measure.
- Large model print size: The miniatures should be relatively/possibly large, which speaks for the “Plus” and “Pro” versions. Roughly 30cm vs. 20cm (Explorer and Solid) maximum model size is a noticeable difference. However, the additional 2 centimeters in each dimension of the “Pro” variant in comparison to the “Plus” variant is not that important.
- Good value for money: Finally, the cost effectiveness is another (striking) argument to buy the “Plus” variant. The “Plus” model is more than 20% cheaper and if it is placed on a fire-proof surface (which is suggested for all models), it offers no relevant disadvantages over the “Pro” variant. Thus, the “correct” choice is the WS 3D Plus model. The agent should try to recommend the features of this model without revealing that this is the “right/desired” outcome.

5.7.3 Question catalogue

- Can you please introduce yourself and tell us something about your VR experience so far.
- How often did you have a VR headset on before the experience?
- There is a concept called Uncanny Valley. It says that very humanized representations seem ``uncanny''. When you look at the Uncanny Valley scale, where do you place the avatar of the experience? [Participant is given a sheet with the uncanny valley, see Figure 29]
- Were there any technical problems with the experience?
- How present did you feel?
- Could you imagine going shopping in such an environment?
- If VR shopping would become really popular, would you trust the avatar and his advice?

- Now about the avatar of the advisor. How did it feel to face a human in the form of an avatar?
- How did you feel about the timing of the appearance?
- Did the agent come at exactly the right time?
- If not, when would it have been better for him to appear?
- Would you have preferred to actively call the agent yourself?
- How real did the avatar feel?
- What thoughts come to your mind when you think about your gaze data being analyzed?
- What measures would you like to see in a commercial product to protect your personal data?
- You have now experienced our interpretation of a virtual commerce scenario.
- Any final general words about the VR experience?

6 Paper E: Adaptive product comparison assistance in virtual reality

Tobias Weiß, Jella Pfeiffer and Martin Meißner

Abstract

The exponential growth of online transactions and the proliferation of e-commerce platforms have led to the necessity of effective user assistance mechanisms. As retail evolves into the digital realm, the role of User Assistance Systems (UAS) is pivotal because useful adaptations help facing the challenges associated with consumer's shopping experiences, ranging from product discovery and selection to payment and post-purchase support. New interaction paradigms demand for experimental evidence underpinning that the adaption really helps the consumer with their decision-making process. We investigate if the adaptive UAS is too smart for its own good and focus on the critical moment of system appearance and its impact on consumers trust in a virtual reality (VR) retail scenario. In our laboratory experiment with 120 participants, who all made three different purchase decisions for muesli products, a comparison matrix UAS was either present from the beginning or appeared after the participants began comparing two products. This product comparison was determined by means of eye tracking. Our data analysis unveils the impact of context-awareness and explanations about the adaption mechanism on consumer trust. We find a negative relation between context-awareness and trust with competing mediations via perceived control over the UAS and intelligence of the UAS. For practitioners, our findings suggest that offering product comparison UAS in VR retail environments immediately outperforms context-aware interference timing in terms of building trust.

Keywords: Consumer Behavior, Eye Tracking, User Assistance System, Virtual Commerce, Virtual Reality

6.1 Introduction

In retail, gaining and maintaining consumer trust is a clear success factor, and vendors seek to foster trusting beliefs of their clients by paying close attention to their individual needs (Bauman & Bachmann, 2017; Gomez et al., 2004). With the transformation of retail from brick-and-mortar stores to e-commerce, decision support for users has become ubiquitous (Maedche et al., 2016). With the recent attempts to shift to virtual commerce (i.e., shopping in immersive virtual 3D environments), the importance of User Assistance Systems (UAS) as an interface between buyer and seller has increased (Acar & Tekinerdogan, 2020). Although the visual acuity of modern headsets has improved, reading text when interacting with 3D objects in VR can still be cumbersome. Thus, it makes sense to support consumers with a comfortable tool that allows for side-by-side comparisons in 2D as on e-commerce platforms. Following Friemel et al. (2018), we deem adaptivity and interference timing as elemental for successful customer interactions with such an UAS.

Traditionally, UAS rely on deliberate user input, such as pressing a button, a gesture, or voice activation. In VR, manual UAS activation can be cumbersome and inefficient because users cannot see the buttons on the controllers. The limited availability of different buttons on VR controllers intensifies the issue and, thus, the user experience may improve if activation happens automatically. In contrast to basic manual system invocation, we propose to analyze the consumer's gaze, track product comparison patterns, and use them to determine the UAS interference timing.

However, it is not clear how an adaptive UAS impacts the consumer's perceived trust. A UAS with gaze-based appearance may be more effective than a UAS that is present from the very beginning because it avoids any negative impact on the overall first impression by distracting and occluding the products. On the other hand, consumers may lack control over the system and feel patronized by an automatic appearance paradigm. Furthermore, as explainability is an important factor for regulators, it is a valuable insight whether an explanation to the consumer (about how and why the UAS appears) influences their perceived trust in the system (Angerschmid et al., 2022).

We compare two passive interference paradigms: UAS present from the beginning (NoCtx) and context-aware UAS interference (Ctx). As second dimension of comparison, we either provide explanations about how and why the UAS interferes (Expl) or we do not provide any explanation (NoExpl). As Figure 32 shows, the UAS helps to compare multiple products in a tabular format. In the context-aware condition, the system appears when consumers start to compare two products. We detect this moment by means of eye tracking and use the first gaze pattern that goes back and forth between two different products. The experiment design aims to identify differences in perceived trust between the basic and the context-aware UAS interference that may be moderated by perceived intelligence of the system and perceived control over the system.

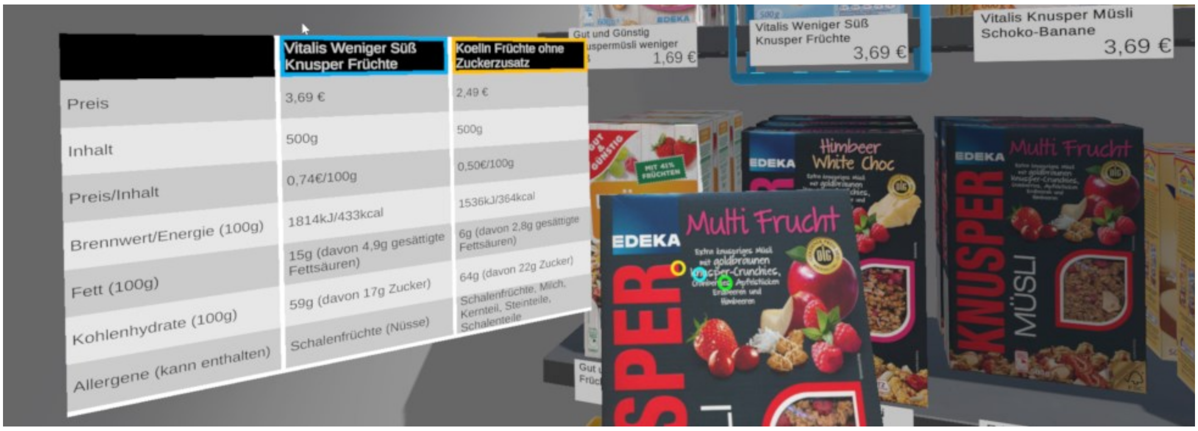

Figure 32. The UAS shows a comparison matrix to inform the consumer's purchase decision.

This paper contributes to the literature body in following ways: (i) It strengthens the theoretical understanding of the relation between context-awareness of a UAS and consumers intention to trust the system. (ii) It provides guidance if an explanation about the interference timing improves the consumer's trusting intentions, and (iii) it presents a data-driven comparison of different mediation models and an alternative moderated mediation approach. Moreover, the presented artifact design informs practitioners who want to implement similar adaptive UAS.

6.2 Background

6.2.1 User Assistance Systems

At the nexus of consumer behavior and decision support, user assistance is "an intelligent system's capability to assist users while performing their task by means of human-, task-, and/or context-dependent augmentation of […] human-computer interaction." (Morana et al. 2020, p. 189). Examples for common types of UAS in modern computer software are help systems, tutorials, contextual menus and tooltips. By providing consumers with relevant information, UAS can make technology more accessible and user-friendly (Olenberger, 2023). Among the possible benefits that UAS offer are improved user experience, reduced support costs, and increased technology adoption (Friemel et al., 2018; Olenberger, 2023).

6.2.2 Context-awareness of the UAS

Context-awareness of computer systems has been subject to investigation for more than two decades (Abowd et al. 1999; Barkhuus and Dey, 2003, Chittaro and Ration 2000; Capurso et al. 2018; Lallemand and Koenig 2020). Schilit et at. (1994) have been first to introduce the concept of context-aware applications and analyzed it in the domain of mobile distributed computing. In their categorization, they list proximate selection (the category in which our UAS fits in) and describe it as “a user-interface technique where the objects located nearby are emphasized or otherwise made easier to choose” (Chen and Kotz 2000, p. 3). Since then, further definitions emerged that describe different levels of interactivity and delineate active and passive context-awareness (Barkhuus and Dey 2003). An active system adapts to the context independently where a passive system asks to user to do so.

In the domain of human-computer interaction, context-aware interfaces are a frequent subject of investigation (Stefanidi et al. 2022). Several studies mention the potential of context-awareness to improve the user experience by tracking and adapting to the user’s state (Carrera-Rivera et al. 2022; van Hove et al. 2017; Zhang and Uruchurtu 2011). However, previous research has also studied cases in which UAS failed to be beneficial (Dey 2009). The probably most famous negative example is the Microsoft Office Assistant “Clippy”. Such failed assistance approaches may have left traces in the minds of future virtual commerce users, increasing the importance of thoroughly understanding the impact of context-aware systems on consumer trust.

For smart user assistance, tracking the user’s state and then quickly adapting to it sounds like a good strategy. The user state refers to the current condition or situation, which can include their location, other entities they are with, and the interactable objects that are nearby (Schilit et at. 1994). A related study has investigated interference timing as a form of context-awareness and concluded “that a small delay in the delivery of information could result in a large mitigation of disruption” (Bailey and Konstan 2006, p. 705).

6.2.3 Perceived trust in the UAS

On a basic level, trust can be seen as the willingness to hand over control to another entity and give up own agency (Berg et al. 1995). The concept of trust plays an important role in decision-making and has different notions, such as organizational and interpersonal trust (Rotter 1967). McKnight et al. (2002) emphasize trust-building as essential factor when adapting to new technology. A user's trust in an unfamiliar trustee (the assistance system) is referred to as initial trust (Kim and Prabhakar 2004; McKnight et al. 2011; McKnight et al. 2002). For interpersonal trust, recent research has identified the dimensions competence, benevolence, and predictability as constituting elements (Deljoo et al. 2018; Afzal et al. 2010). In a business context, the level of trust has implications for consumer satisfaction and their intention to reuse (Panigrahi et al. 2018; Ginting et al. 2023). The literature suggests that perceived intelligence of the trustee (Trzebiński and Marciniak 2022) and perceived control over the trustee (Arcand et al. 2007; Huang et al. 2014) act as potential impact factors on the relationship between context-awareness and trust.

"[T]rust is a critical factor in stimulating purchases over the Internet" (Quelch and Klein 1996, p. 61) and thus it is of relevance for e- and virtual commerce. In e-commerce, trust plays a central role because of its high relevance not only for web stores but also for online platforms and marketplaces with a large number of buyers and sellers (Corbitt et al. 2003; Jones and Leonard 2008). Empirical evidence from a desktop-based experiment suggests that consumers with high overall trust in a particular vendor also have a higher intention to purchase an offered product (Oliveira et al. 2017). Further authors have investigated how VR may foster trust and the results suggest that offering an immersive virtual commerce outlet may facilitate a vendor's overall trustworthiness (Papadopoulou 2007; Gupta et al. 2020).

6.2.4 Explanations of UAS actions

As regulators are currently shaping ethical guidelines and laws for future virtual commerce applications, it is a relevant question how an explanation about the adaptive behavior can alter consumers' perceived trust in the system (Angerschmid et al. 2022). Explanations can be categorized in how and why statements (Liao et al. 2020) and specific research streams may have an own focus. For example, explainable artificial intelligence research concentrates on why explanations, potentially because of the black-box behavior of deep learning algorithms that makes it difficult to explain how the system operates (Bauer et al.

2023). Yet, even in the in the field of explainable artificial intelligence, how explanations about the system‘s overall logic are discussed (Liao et al. 2020). Several studies have shown that explanations can enhance trust in a system (Rader et al. 2018; Dodge et al. 2019; Yang et al. 2020). However, other authors provide evidence that users do not follow the algorithm‘s advice if it is transparent (Poursabzi-Sangdeh et al. 2021) or that users are even less willing to trust a system when explanations are provided (Erlei et al. 2020). Overall, there is controverse empirical evidence about explanations and their impact on the perception of a context-aware UAS.

6.2.5 Perceived intelligence of the UAS

Johnson et al. (2008) define a user’s perception of overall system intelligence as the sum of its intelligence, knowledge, and purpose. Likewise, in the context of human-robot interaction, previous studies have shown that perceived intelligence depends on the perceived competence, knowledge, responsibility as well as sensibleness (Bartneck et al. 2009; Parise et al. 1999).

Paralleling the research on intelligence perception of technology, the marketing literature focuses on determining key dimensions of perceived product intelligence. Rijsdijk et al. (2007) identify six key dimensions for perceived product intelligence: autonomy, ability to learn, reactivity, ability to cooperate, human-like interaction, and personality. They validate a now widely adopted scale by comparing non-intelligent with intelligent products, such as autonomous versus manual lawnmowers, regarding their impact on perceived trust. For non-human systems, only the dimensions autonomy, ability to learn, and reactivity apply. In the following, we give a brief overview of these relevant sub-dimensions: Autonomy is the degree to which the UAS acts independently and goal-directed (Baber 1996). The second dimension is the ability to learn, and it refers to the degree to which the UAS can use prior information and adapt to the consumer’s needs (Nicoll 1999). The third relevant intelligence dimension is reactivity and it refers to the ability of the UAS to react to changes in its environment and respond to stimuli (Bradshaw 1997). In line with Moussawi et al. (2021), who investigated personal intelligent agents, we argue that perceptions of intelligence may play a role for the trust in a system.

6.2.6 Perceived control over the UAS

The concept of control describes whether an individual perceives a feedback mechanism contingent on their own behavior or independent of it. Rotter (1966) builds up on theory about the human cognitive reward system and describes control as an individual’s perception of the causal link between their actions and the outcome. Perceived control has many facets and involves different constructs and theoretical ideas, such as locus of control, causal attributions, learned helplessness, and self-efficacy (Skinner et al. 1998). The level of perceived control seems to be highly dependent on cultural and individual differences (Hornsey et al. 2019; Skinner et al. 1998) but there are certain general tendencies. If an outcome is consistently contingent on the preceding behavior, a notion of perceived control is present while, on the other hand, if the outcome has a chance component or is independent, the feeling of control may be weaker or absent.

In the context of adaptive UAS in virtual commerce, perceived control is an integral part of the consumer experience (Hu 2023). A recent study showed that perceived control and purchase intention go hand in hand (Zhao et al. 2023). However, the used 360° videos do not allow for movement and interaction. Thus, the results in an immersive VR setup may vary.

6.2.7 Eye tracking in VR

UAS can leverage the capabilities of bio-sensors, such as eye tracking (ET) cameras, to record and respond to the state of the consumer (Gellersen et al. 2002). Gaze patterns are suitable for tracking visual attention (Duchowski 2017) but ET research heavily relies on the eye-mind hypothesis (Just and Carpenter 1980). The eye-mind hypothesis only holds if individuals do not intentionally direct their attention and visually focus on a certain object while thinking about something completely different. Even though there are further threads to the validity of ET research (Orquin and Holmqvist 2018), experimental findings indicate the robustness and replicability of ET results in numerous scenarios (Holmqvist et al. 2011). Previous research combined ET and VR and showed that visual attention and pupillometry can help to learn about the user state (Pfeiffer et al. 2020; Meißner et al. 2019; Wang et al. 2014; Novák et al. 2023) and thus may inform gaze-based adaptive features of a UAS. Further previous ET studies built upon the Engel-Kollat-Blackwell decision phase model that subdivides decision processes into different phases, such as orientation, evaluation, and validation (Engel et al. 1968; Russo and Leclerc 1994). They determined the transition between different phases by simple rules based on ET features, like refixations on products. For the right timing of user assistance, this shift between orientation and evaluation may be of interest. Peukert et al. (2020) pursued an on-the-fly attempt to determine the phases and used ET to identify the first pairwise product comparison that indicates a transition from orientation to evaluation phase. That argue that after starting to evaluate two of the buying options in detail, help may be appreciated by the user.

6.3 Hypotheses

6.3.1 Context-awareness

Barkhuus and Dey (2003) report that context-awareness facilitates a smooth interaction between humans and information technology. Pointing to the same direction, Richthammer and Pernul (2020) present results indicating that context-awareness positively influences the purchasing behavior of consumers when using a recommender system. A further study shows that context-awareness allows for an increased consumer value for location-based mobile services (Vos et al. 2009).

Following these positive reports on effects of context-awareness in related domains, we transfer the idea to a virtual commerce shopping scenario and, compare a context-aware UAS with one that is immediately present. In our case, context-awareness means that the consumer’s gaze patterns activate the UAS after the first comparison of two products. The effect may be negative as an adaptive interference of the UAS can appear intrusive and the lack of control may diminish the user experience. On the other hand, if a UAS is present from the very beginning, it could occlude part of the products and negatively impact the user experience as well. We expect the first effect to be stronger and hypothesize that perceived trust increases when users are given time to oversee the shelf before displaying the UAS that helps comparing the products. Thus, we formulate following hypothesis:

H1: A context-aware appearance of the UAS increases perceived trust in the UAS.

6.3.2 Explanation effects

Explanations are revealing the system's internal mechanisms to its users and that plays an important role in fostering trust (Nunes and Jannach 2017). Related studies indicate that explanation interfaces can foster trust building (Pu and Chen 2006) and that they impact confidence in assisted decision making (Zhang et al. 2020). A further recent study found that "[e]very explanation improves users' appropriate trust in […] the human-machine collaboration" (Yang et al. 2020, p. 197). Even though there is also contrary experimental evidence that reports negative effects of explanations (Poursabzi-Sangdeh et al. 2021; Erlei et al. 2020), we pose following hypothesis:

H2: The explanations about how and why the UAS appears increase perceived trust in the UAS.

6.3.3 Mediations

On one hand, trust in the system and control over the system form a tight bond (Castelfranchi and Falcone 2000; Bijlsma-Frankema and Costa 2005) and Möllering (2005) even sees them as duality. Experiments suggest that giving up control leads to a decrease in trust (Muir and Moray 1996; Lee and Moray 1992). Trust in the system and the intelligence of this system form an equally tight bond (Haring et al. 2013) and empiric evidence suggests that perceiving an entity as intelligent leads to an increase in trust (Haring et al. 2013; Moussawi et al. 2021).

We believe in a positive sentiment in favor of the context-aware UAS. When comparing instant and adaptive UAS interference both the perception about control over the UAS and intelligence of the UAS may change. The adaptive UAS offers only little control while users may perceive timely interference as more intelligent in comparison to the instantly available UAS. We expect to observe a positive effect of the adaptive UAS interference, even though the indirect effects of perceived control and system intelligence may neutralize each other. Overall, we conjecture that the effect of context-awareness on trust is mediated by two latent constructs with opposite signs: control over the system and the intelligence of the system. In other words, we expect opposed impacts of perceived control over the system and the perceived intelligence of the system on the perceived trust level of the user. Our hypotheses regarding these mediations read as follows:

H3a: Context-awareness effects on perceived trust are mediated by the opposing influences of perceived control over the UAS and perceived intelligence of the UAS.

H3b: Explanation effects on trust are mediated by the opposing influences of perceived control over the UAS and perceived intelligence of the UAS.

6.4 Method

6.4.1 Experimental design

We manipulate the context-awareness of a gaze-based UAS (Ctx) and the explanations about its behavior (Expl), as shown in Figure 33. The manipulation of context-awareness consists of two interference paradigms: (a) the UAS is present from the beginning and (b) context-aware interference of the UAS. We argue that a consumer's willingness to trust a UAS depends on the perceived intelligence (Int) of the UAS and the perceived control over the UAS (Ctrl). Therefore, we see intelligence of the UAS and control over the UAS as latent constructs that mediate the relationship between context-awareness and trust.

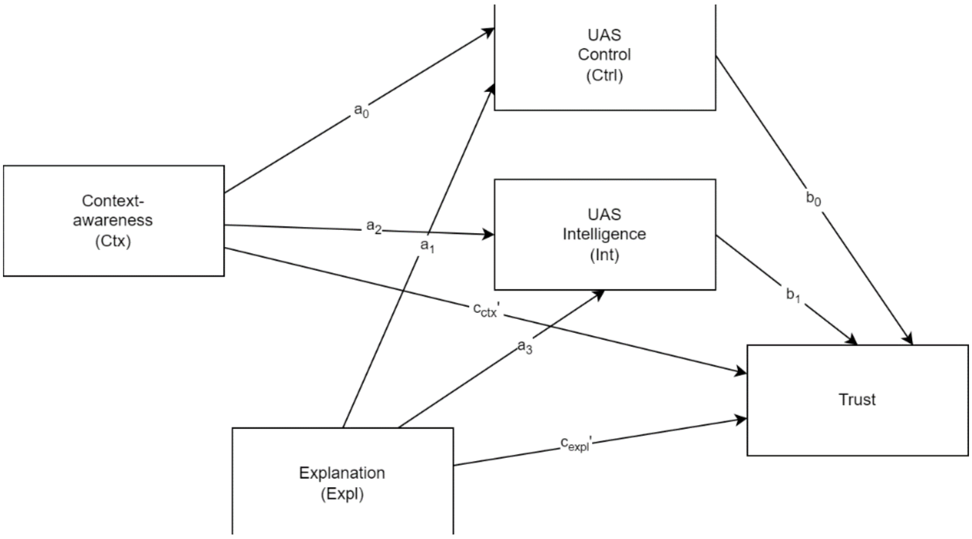

Figure 33. Base model with two parallel mediations.

The experiment follows a 2x2 between-subjects design; it has an ethics approval, and a pre-registration (AsPredicted #135337). The manipulation of context-awareness means that participants either see the UAS from the beginning of a trail or that the UAS interferes right after the first product comparison. The second treatment dimension is whether the participant receives an explanation of the UAS behavior before completing the questionnaire (see the supplementary material for the explanation texts).

We created the VR scenes in Unity; they consist of an onboarding environment and the showroom in which participants perform three purchasing tasks sequentially. The experiment takes place in a laboratory room with a 3x4m VR area, a VR computer, and a survey computer (as illustrated in supplemental Figure 37). We use an HTC Vive head mounted display (HMD) that has integrated ET cameras with 250Hz. Overall, the experiment is designed to last approximately one hour.

6.4.2 Manipulation of context-awareness

To detect the moment when the context-aware UAS appears, we use ET and determine the first X-Y-X product comparison as an indicator (Russo and Leclerc 1994; Peukert et al. 2020). To this end, fixations and saccades are determined in run-time using a saccadic velocity-based algorithmic approach (I-VT) as described by Salvucci and Goldberg (2000) and the gaze targets are determined using ray casting (Pietroszek 2019). For saccades, we set 100°/second as the lower angular speed threshold (Holmqvist et al. 2011), and we limit fixation durations to 0.1 seconds as the lower threshold and 10 seconds as the upper threshold (Duchowski 2017). If the algorithm detects a fixation, we store the event in a buffer that keeps the events of the past 10 seconds. With every new fixation, we check if an X-Y-X product comparison pattern occurred within this buffer window.

6.4.3 Measurements and constructs

In the survey, which takes place immediately after the VR experience, we ask the participants questions about their perceived control over the UAS (Kidwell and Jewell, 2003; Armitage et al., 1999), perceived intelligence of the UAS (Rijsdijk et al. 2007), and their trust in the UAS (Thatcher et al. 2011; McKnight et al. 2002). We use multiple items for autonomy, ability to learn, and reactivity to constitute the perceived intelligence construct. We do not evaluate the human-likeness or personality of the UAS. As a comparison matrix, the UAS does not have an avatar or other humanoid traits. The construct for perceived control also consists of multiple items for benevolence, competence, and predictability.

As exploratory control variables, we assess the participants' affinity for technology using the ATI-S scale (Wessel et al. 2019) and their disposition to trust technology (Lankton et al. 2015). We measure all constructs on 7-point Likert scales and adapt all questions to fit our experiment (what includes providing German translations).

A spreadsheet with the exact wording of all items is available in the accompanying online repository (Weiß 2024).

6.4.4 Procedure

On arrival, we randomly assigned participants to one of the four conditions and started the corresponding questionnaire. Our consent form informed participants about the ethical standards and asked them to agree to pseudonymized publication of their data. After the participants accepted these terms, we determined the participants dominant eye (Miles 1929) and measured their interpupillary distance to adjust the HMD accordingly. Participants watched a video that explained the upcoming scenario, tasks, controller usage, and how to interact with the UAS. Then, the experimenter helped them to fit the HMD to their head. After a 5-point ET calibration and a reading test, participants entered the training scene and practiced the interactions that they previously saw in the video, guided by the experimenter.

The training environment consisted of the same shelf scene that we used for the subsequent decision task, but the products were baking powder instead of muesli. To familiarize the participants with the controllers, the experimenter asked them to pick up a product, activate the shopping list, use the binocular function to read details on the packaging, activate the UAS for three products, compare the products, and place one product in a shopping cart next to the shelf. After viewing and canceling the confirmation dialogue, the participants had the chance to ask last questions before moving on to the experimental decision tasks.

For each of these decisions, the shopping shelf was filled with 24 different muesli products on randomized shelf positions. We designed the tasks in such a way that only one product met the set of criteria specified in the task (see the online supplementary material for the task texts). For example, participants had to search for a chocolate muesli with a low fat and sugar content. They also had to consider a nut allergy and, thus, avoid products containing nuts. Participants were able to check these criteria at any time using the shopping list. The tasks were incentivized in that participants received a fixed reward of 12 Euro for the entire experiment and had the opportunity to earn an additional 1 Euro for each product selected correctly.

After completing all three tasks, half of the participants received an explanation describing how and why the UAS appeared. We displayed the explanation as text in VR and additionally provided it on paper after our participants detached the HMD (see supplementary material for the explanation texts).

6.4.5 Statistical modelling and sample size

In our Bayesian analysis, we report distributions instead of point estimators(Kruschke 2014; McElreath 2018; van Doorn et al. 2021; Martin 2018). With a small sample and when assumptions about the population are hard to fulfill, a Bayesian modeling approach offers advantages over other traditional approaches (van de Schoot et al. 2021). Bayesian sample size considerations are about achieving desired credibility intervals and posterior accuracy, and there is no closed formula to calculate the needed number of observations based on expected effect sizes (van de Schoot et al. 2014; McElreath 2018). Taking the substantial cost of sequential experimental VR sessions into consideration, we pragmatically aim for the minimum number of observations that allows us to assume a normal distribution of the drawn sample. Overall, we collected 120 clean observations, 30 for each treatment group.

With the small dataset and coarse sample size determination, we deem it advisable to perform a model comparison that evaluates different prior distributions and model variants. In our analysis, we relax the assumption for a normal prior distribution by comparing a model that utilizes normal prior distributions with models that utilize less informative Student-t prior distributions. The student-t distribution allows to express more uncertainty, as it has heavier tails compared to the normal distribution and assigns higher probability to more extreme values.

6.4.6 Participants

The participants (72 females and 48 males, mean age = 26.2, SD = 5.5) were recruited on our campus and were mostly students. Our questionnaire and procedure were bilingual, enabling 97 (80.8%) German speakers and 23 (19.2%) English speakers to participate. We excluded one participant who could not wear the VR headset (due to vision problems) and two participants who demonstrated insufficient language proficiency in the selected survey language (German or English).

6.5 Results

We perform the following analysis with JASP (van Doorn et al. 2021), python 3.10 with additional libraries, especially the package PyMC for the Bayesian models (Patil et al. 2010). The aggregated data and code are available in the accompanying online repository (Weiß 2024). For the sampling mechanism, PyMC relies on a No U-turn sampler (NUTS) implementation (Lao and Louf 2020). First, we provide an overview of the main constructs in Figure 34 which shows the responses for the constructs control, intelligence, and trust on a per-group basis.

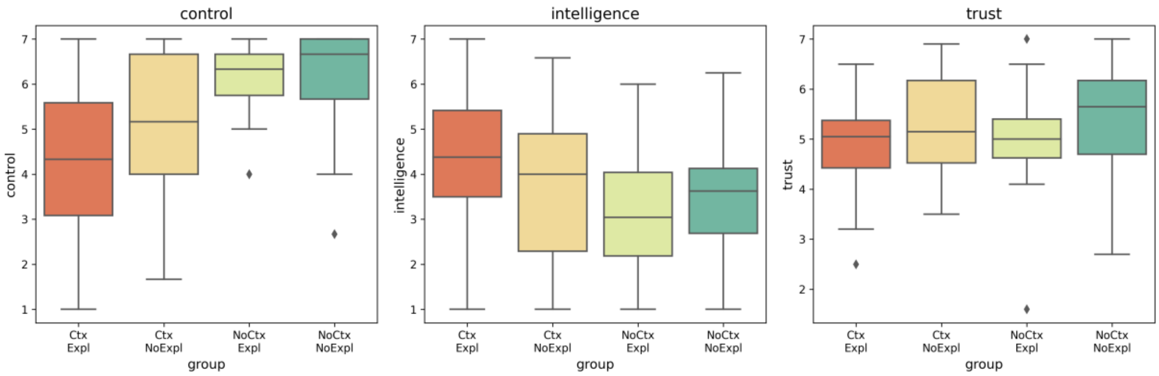

Figure 34. Boxplots per treatment group for control, intelligence, and trust.

6.5.1 Reliability and control variables

Supplementary Table 17 shows the reliability of the constructs. With a Cronbach's α > 0.8, most constructs display good reliability (Petter et al. 2007). For the three benevolence items, the value of α = 0.791 is slightly below the commonly applied threshold of 0.8, but we still deem the construct valid. To check whether the questionnaire items map to the theoretical constructs, we perform a confirmatory factor analysis that is shown in Table 14. The items for the theoretical sub constructs benevolence, competence, and predictability map correctly and the correlation between the factors is moderate (see supplementary Table 20).

Table 14. Factor loadings for perceived trust with fixed number of three factors.

	Factor 1	Factor 2	Factor 3	Uniqueness
Predictability [TP002]	0.921			0.168
Predictability [TP001]	0.853			0.311
Predictability [TP004]	0.809			0.330
Predictability [TP003]	0.561			0.621
Competence [TC002]		0.957		0.137
Competence [TC001]		0.804		0.337
Competence [TC003]		0.729		0.312
Benevolence [TB001]			0.944	0.251
Benevolence [TB002]			0.806	0.254
Benevolence [TB003]			0.501	0.667

Note. Applied rotation method was promax.

To check for confounds in the experimental data, we analyze the control variables. We conduct Bayesian ANOVAs (van den Bergh et al. 2019) for the participants' affinity for technology (measured by the ATI-S scale) and their disposition to trust technology. For both constructs the ANOVA indicates no significant differences between the four experimental groups and favored the respective null model. We omit the tables but provide the JASP file in the accompanying online repository (Weiß 2024).

6.5.2 Mediations

To address our hypotheses, we use Bayesian mediation models (Hayes 2017; Yuan and MacKinnon 2009) that utilize different prior distributions and control variables. We compare these models and select the model with the highest expected log pointwise predictive density (ELPD). The ELPD provides a common measure for the generalization capability of the model at hand (Martin et al. 2021). For the prior distributions, we apply Gaussian distributions with μ = 0, σ = 3 and Student-t distributions with μ = 0, σ = 3, and ν = 15. For the variance terms ($\sigma_{ctrl}, \sigma_{int}, and\ \sigma_{trust}$), we use Half-Cauchy distributions with β = 1 (Polson and Scott 2012). We fit the models using four Markov chains with 4000 samples each, with an acceptance rate threshold value of 0.8 (M. J. Betancourt et al. 2015), and monitor the stability of the chains (van de Schoot et al. 2014). The base model, as shown in Figure 33, is constituted of linear functions and can be denoted using following equations, where i represents the intercept and a and b are the respective coefficients:

$$Control \sim \mathrm{Prior}(i_{ctrl} + a_0 \cdot Ctx + a_1 \cdot Expl\ , \sigma_{ctrl}),$$

$$Intelligence \sim \mathrm{Prior}(i_{int} + a_2 \cdot Ctx + a_3 \cdot Expl\ , \sigma_{int}),$$

$$Trust \sim \mathrm{Prior}\big(i_{trust} + \mathrm{c}'_{\mathrm{ctx}} \cdot Ctx + \mathrm{c}'_{\mathrm{expl}} \cdot Expl + b_0 \cdot Control + b_1 \cdot Int,\ \sigma_{trust}\big).$$

We evaluate model variants with Student-t prior distributions that utilize disposition to trust (Dsp), technology affinity (Ati) or both as control variables, and a variant of the best model that does not consider explanations. Additionally, we evaluate an alternative moderated mediation approach (Muller, 2005) which models the explanation about the how and why of the UAS appearance as moderator for the mediation of context-awareness on trust (see Figure 35). This alternative moderated mediation model can be denoted using following equations:

$$Control \sim \text{Prior}(i_{ctrl} + a_0 \cdot Ctx + a_2 \cdot Ctx \cdot Expl + a_4 \cdot Expl, \sigma_{ctrl}),$$

$$Intelligence \sim \text{Prior}(i_{int} + a_1 \cdot Int + a_3 \cdot Ctx \cdot Expl + a_5 \cdot Expl , \sigma_{int}),$$

$$Trust \sim \text{Prior}\begin{pmatrix} i_{trust} + \mathrm{c}'_{\mathrm{ctx}} \cdot Ctx + \mathrm{c}'_{\mathrm{expl}} \cdot Exp + b_0 \cdot Ctrl + b_1 \cdot Int \\ + a_6 \cdot Dsp + a_7 \cdot Ati, \sigma_{trust} \end{pmatrix}.$$

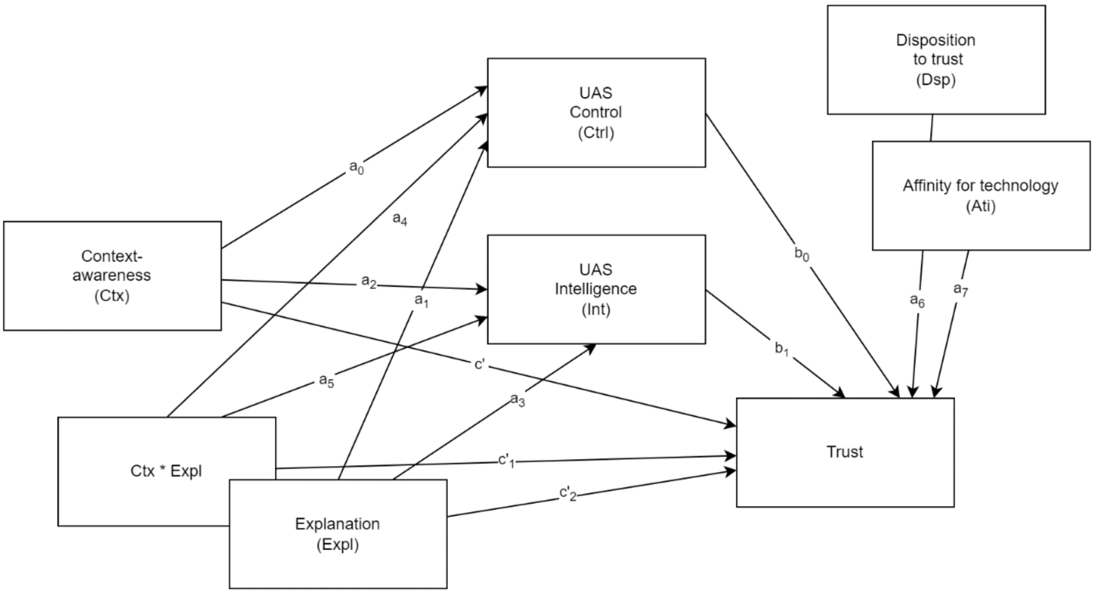

Figure 35. Moderated mediation model as alternative approach.

Table 15 shows the comparison results of all evaluated models. The base model with Student-t prior distributions has a better ELPD score than the base model with Gaussian priors. This initial finding motivates the evaluation of further model variants with Student-t prior distributions. The overall best-performing model with Student-t prior distributions uses disposition to trust as single control variable. The runner-up model also uses Student-t prior distributions but does not take the explanations into account. The difference in ELPD between these two models is only 0.461 but the difference in standard error (SE) of the model without explanation path is considerably higher. In our comparison, adding the Ati control variable decreases model ELPD performance. It is also noteworthy that the moderated mediation approach ranks fourth and has a difference in ELPD 1.456 in comparison to the best model.

Table 15. Model comparison for combined mediations.

Models	**Rank**	**ELPD LOO**	**P LOO**	**ELPD Diff**	**SE**
Student-t Dsp	0	-153.426	7.261	0	8.883
Student-t Dsp without explanation path	1	-153.888	6.250	0.461	9.018
Student-t Dsp Ati	2	-154.232	8.234	0.806	8.821
Student-t Dsp moderated mediation	3	-154.883	7.305	1.456	8.959
Student-t	4	-156.908	6.039	3.480	8.684
Gaussian	5	-157.640	7.038	4.214	8.604
Student-t Ati	6	-157.657	6.238	4.231	9.059

We show the best model, Student-t Dsp, in Figure 36 with mean values for the posterior distributions of all individual parameters and with the corresponding 94% highest density interval (HDI) threshold values. The 94% HDI is a common choice and the default value set in the PyMC package, although some authors prefer other values (McElreath 2018). The dotted arrows in Figure 36 indicate effects that are not significantly different from zero (according to the chosen 94% HDI threshold). For the parameters a_1 and c_1', the value zero is only on the tail of the distribution, and the respective 94% HDI almost does not cover it. Thus, although not significant, there is a negative tendency for the explanations regarding their effect on perceived control over the UAS and the direct effect on trust.

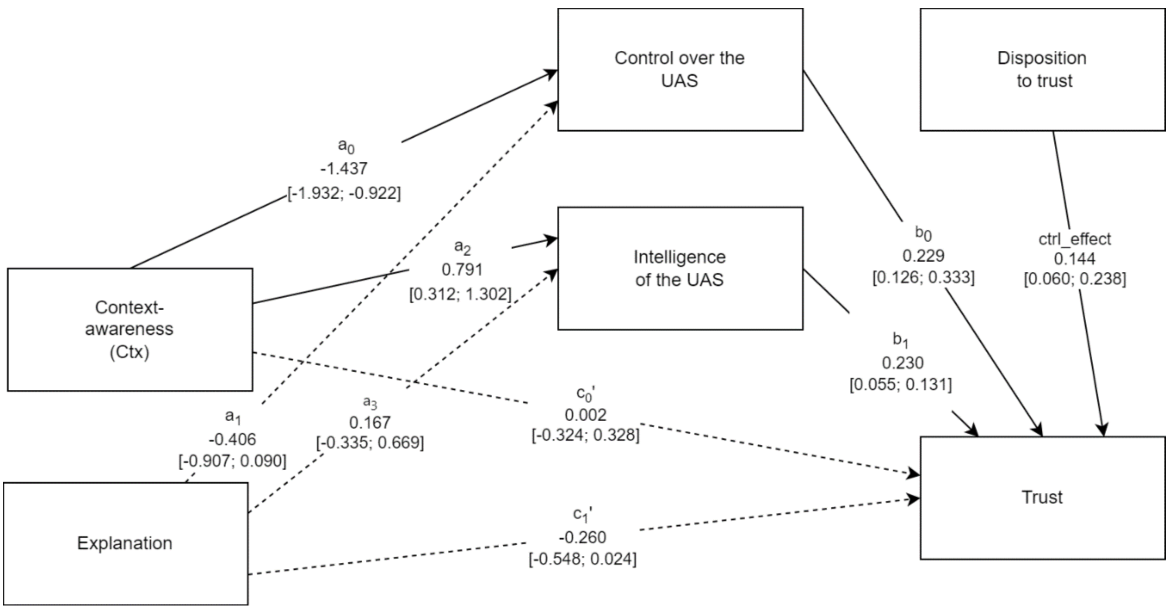

Figure 36. Diagram of the best combined Bayesian mediation model with parameter mean values and 94% HDI.

In

Table *16*, we list the total, direct, and indirect effects for the best combined mediation model. The 3% and 97% HDI columns represent the lower and upper thresholds for the 94% HDI that are also shown in Figure 36. Supplementary Figure 39 shows the corresponding posterior distributions with 94% HDI interval indicators. Regarding the total effects, the posterior distribution for context-awareness (mean = -0.420, SD = 0.185) is significantly different from zero. This means the context-aware condition had a negative effect on trust, what is opposed to what we expected. While a total negative effect is present for context-awareness, the total effect of the explanations is not significant (as it has a negative sign but is close to zero). Thus, the best model does not provide support for both H1 and H2.

We find that the best model supports both a negative indirect path from context-awareness via perceived control over the UAS (mean = -0.329, SD = 0.100) and a positive path from context-awareness (Ctx) via perceived intelligence of the UAS (mean = -0.184, SD = 0.077), and thus H3a is supported. However, parameters for the explanation effects (Expl) are not significant and therefore H3b is not supported.

Table 16. Posterior parameter distributions for the Student-t Dsp mediation model.

Effect	Mean	SD	HDI 3%	HDI 97%
Total effect Cxt (a_0*b_0+a_1*b_1+c_0’)	-0.420	0.185	-0.764	-0.073
Total effect Expl (a_2*b_0+a_3*b_1+c_1’)	-0.037	0.183	-0.373	0.315
Direct effect Ctx → Trust (c_0’)	0.002	0.172	-0.324	0.328
Direct effect Expl → Trust (c_1’)	-0.260	0.152	-0. 548	0.024
Indirect effect Ctx → Control → Trust (a_0*b_0)	-0.329	0.100	-0.519	-0.147
Indirect effect Expl → Control → Trust (a_1*b_0)	-0.093	0.066	-0.222	0.026
Indirect effect Ctx → Intelligence → Trust (a_2*b_1)	0.184	0.077	0.048	0.333
Indirect effect Expl → Intelligence → Trust (a_3*b_1)	0.039	0.065	-0.083	0.164

6.6 Discussion

The main insight of our study is that participants trusted the context-aware UAS less than the static variant. The experiment could not confirm findings from other domains that reported more consumer trust for the context-aware system (H1). Instead, the best mediation model suggests a negative total effect of the context-aware UAS (which consists of gaze-based interference of the comparison matrix).

The model comparison in Table 15 indicates that disposition to trust is a useful control variable while using affinity for technology has no positive impact on model performance. The Student-t model that solely uses affinity for technology as control variable performs even worse than the base model with Gaussian priors. Adding the affinity for technology control variable in combination with disposition to trust also reduces the ELPD score. Likewise, the comparison shows that our data-driven approach favors two parallel mediations over a moderated mediation, even though the difference in model performance according to the ELPD metric is small.

We acknowledge that different notions of context-awareness are possible, and it is likely that other context-aware UAS implementations lead to different results. We recall that our context-aware UAS appeared after the first X-Y-X product comparison pattern. During the experiments, we observed that invoking the UAS by this single rule was not sufficient for all participants. We conclude that the simple gaze patterns that we used were not ideal as an interference criterion. To further generalize, we propose to introduce a band filter to ascertain a time interval with minimum and maximum values for the UAS appearance.

The Student-t Dsp mediation model suggests no significant explanation effect about how and why the UAS appeared on the participants perceived trust, what is contrary to what we expected (H2) and what the literature suggests. We conclude that control over the UAS and intelligence of the UAS are likely not the sole key mediators for the effect of an explanation on perceived trust. As our results are averse to previous findings, future research may focus and investigate this dimension separately.

As expected, the indirect mediated explanation effects (via control over the UAS and intelligence of the UAS) have different sign, but they are also not significantly different from zero. It may make a difference at what point in time the explanation is provided. During the experiment procedure, we showed the explanation after the three purchase decisions were made, just before answering the questionnaire. The manipulation may be strengthened by providing the explanation to the participants prior to the decision tasks. We note that giving the explanation prior to the tasks bears the risk that participants play with the gaze activation and introduce certain bias. Moreover, participants in the context-aware UAS group received the information that the UAS activates based on ET data, but they did not specifically know about the X-Y-X gaze pattern. Providing this explicit information about how to trigger the UAS may enhance trust in the system and alter the parameters of the presented mediation. Another viable approach could be to introduce intelligibility as a mediator and let the UAS introduce its capabilities by itself (Lim 2010).

While both indirect effects for explanations are not significant, the indirect effects *Context-awareness → Control → Trust* (a0*b0) and *Context-awareness → Intelligence → Trust* (a2*b1) are both significantly different from zero and have a different sign. In other words, we can report competing indirect effects for context-awareness but not for our explanation manipulation. When additionally considering the distributions for *Explanation → Control → Trust* (a1*b0) and *Explanation → Intelligence → Trust* (a3*b1) in supplementary Figure 39, it is easy to verify the relatively small mean effect sizes in the proximity to zero. Thus, we can further support H3a but not H3b.

6.7 Conclusion

Neither did we expect to find a negative relationship between the context-awareness of the UAS and participants trust, nor did we expect the explanation to have only a slight direct negative effect. We must acknowledge that for our scenario, the explanation effect is unlikely mediated by perceived control and intelligence of the UAS. Still, the study confirms that the effect of context-awareness on perceived trust is mediated by competing paths via perceived control over the UAS and perceived intelligence of the UAS.

For the comparison matrix in our scenario, it seems advisable to refrain from context-aware interference timing. For muesli products and a relatively small economic incentive in our experiment, the basic alternative was perceived as more trustworthy, which in turn is likely leading to a higher intention to reuse and overall satisfaction (Acharya et al. 2022). However, the product category (especially more valuable goods) and other impact factors (such as product involvement and saving potential) may alter the situation. For instance, a comparison matrix for high-end designer furniture may introduce different relationships, and context-awareness may be appreciated as experience-enhancing feature of the sales environment for these products. In any case, a button-press to toggle the UI might also be a simple but effective means to improve the user experience. Future experiments on virtual commerce UAS, and particularly comparison matrices, should therefore consider evaluating manual activation paradigms as additional baseline.

A limitation of our study is our working definition of context-awareness and the question if we really manipulated it. We argue that interference timing may work different than other adaptations. This issue points to a future research avenue which could continue investigating different notions of context-awareness. To provide another kind of context-aware UAS, researchers could leverage latest artificial intelligence developments, use prompt engineering, and fine-tune large language models to provide more salient manipulations. Applied to the virtual commerce scenario at hand, future research may introduce a human-like agent that users can ask about the best muesli, given a set of criteria, instead of solving the search task on their own. As agents and avatars come into play, the intelligence sub-dimension which we did not consider in this work (ability to cooperate, human-like interaction, and personality) seem to be relevant mediators and should be incorporated into respective models. Eventually, virtual agents may be able to provide individualized support which is far superior to what consumers are currently used to in terms of advertisements, decision support, and recommendations.

References

Abowd, G. D., Dey, A. K., Brown, P. J., Davies, N., Smith, M., & Steggles, P. (1999). Towards a better understanding of context and context-awareness. In *Handheld and Ubiquitous Computing: First International Symposium, HUC'99 Karlsruhe, Germany, September 27-29, 1999 Proceedings 1.* Symposium conducted at the meeting of Springer.

Acar, M., & Tekinerdogan, B. (2020). Analyzing the Impact of Automated User Assistance Systems: A Systematic Review. *ArXiv Preprint ArXiv:2003.05782*.

Acharya, N., Sassenberg, A.-M., & Soar, J. (2022). Consumers' Behavioural Intentions to Reuse Recommender Systems: Assessing the Effects of Trust Propensity, Trusting Beliefs and Perceived Usefulness. *Journal of Theoretical and Applied Electronic Commerce Research*, *18*(1), 55–78.

Afzal, H., Khan, M. A., ur Rehman, K., Ali, I., & Wajahat, S. (2010). Consumer's trust in the brand: Can it be built through brand reputation, brand competence and brand predictability. *International Business Research*, *3*(1), 43.

Angerschmid, A., Zhou, J., Theuermann, K., Chen, F., & Holzinger, A. (2022). Fairness and Explanation in AI-Informed Decision Making. Mach. Learn. Knowl. Extr., 4, 556-579.

Baber, C. (1996). Humans, servants and agents: human factors of intelligent domestic products. In *IEE Colloquium on Artificial Intelligence in Consumer and Domestic Products.* IEE.

Bansal, S., Gupta, C., & Sinha, A. (2017). Clickstream & behavioral analysis with context awareness for e-commercial applications. In *2017 Tenth International Conference on Contemporary Computing (IC3)*.

Bauman, A., & Bachmann, R. (2017). Online consumer trust: Trends in research. *Journal of Technology Management & Innovation*, *12*(2), 68–79.

Barkhuus, L., & Dey, A. (2003, October). Is context-aware computing taking control away from the user? Three levels of interactivity examined. In *International Conference on Ubiquitous Computing* (pp. 149-156). Berlin, Heidelberg: Springer Berlin Heidelberg.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122–142.

Bijlsma-Frankema, K., & Costa, A. C. (2005). Understanding the trust-control nexus. *International Sociology*, *20*(3), 259–282.

Bradshaw, J. M. (1997). An introduction to software agents. *Software Agents*, *4*, 3–46.

Capurso, N., Mei, B., Song, T., Cheng, X., & Yu, J. (2018). A survey on key fields of context awareness for mobile devices. *Journal of Network and Computer Applications*, *118*, 44–60.

Carrera-Rivera, A., Larrinaga, F., & Lasa, G. (2022). Context-awareness for the design of Smart-product service systems: Literature review. *Computers in Industry*, *142*, 103730.

Castelfranchi, C., & Falcone, R. (2000). Trust and control: A dialectic link. *Applied Artificial Intelligence*, *14*(8), 799–823.

Chen, Y.-C., Thaipisutikul, T., & Shih, T. K. (2020). A learning-based POI recommendation with spatiotemporal context awareness. *IEEE Transactions on Cybernetics*, *52*(4), 2453–2466.

Chen, G., & Kotz, D. (2000). A survey of context-aware mobile computing research.

Chittaro, L., & Ration, R. (2000). Adding adaptive features to virtual reality interfaces for e-commerce. In *International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems.* Symposium conducted at the meeting of Springer.

Corbitt, B. J., Thanasankit, T., & Yi, H. (2003). Trust and e-commerce: a study of consumer perceptions. *Electronic Commerce Research and Applications*, *2*(3), 203–215.

Dasgupta, A., Manuel, M., Mansur, R. S., Nowak, N., & Gračanin, D. (2020). Towards real time object recognition for context awareness in mixed reality: a machine learning approach. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW).* Symposium conducted at the meeting of IEEE.

Deljoo, A., van Engers, T., Gommans, L., & Laat, C. de (2018). The impact of competence and benevolence in a computational model of trust. In *Trust Management XII: 12th IFIP WG 11.11 International Conference, IFIPTM 2018, Toronto, ON, Canada, July 10-13, 2018, Proceedings 12.* Symposium conducted at the meeting of Springer.

Dey, A. K. (2009). Explanations in Context-Aware Systems. In *ExaCt*.

Dodge, Jonathan; Liao, Q. Vera; Zhang, Yunfeng; Bellamy, Rachel K. E.; Dugan, Casey (2019): Explaining models. In: Proceedings of the 24th International Conference on Intelligent User Interfaces. New York, NY, USA. New York, NY, USA: ACM.

Drif, A., Zerrad, H. E., & Cherifi, H. (2021). Context-awareness in ensemble recommender system framework. In *2021 International Conference on Electrical, Communication, and Computer Engineering (ICECCE).* Symposium conducted at the meeting of IEEE.

Duchowski, A. T. (2017). *Eye tracking methodology: Theory and practice*. Springer.

Engel, J. F., Kollat, D. T., & Blackwell, R. D. (1968). A model of consumer motivation and behavior. *Research in Consumer Behavior, Holt, Rinehart and Winston, Inc., New York, NY*, 3–20.

Erlei, Alexander; Nekdem, Franck; Meub, Lukas; Anand, Avishek; Gadiraju, Ujwal (2020): Impact of algorithmic decision making on human behavior: Evidence from ultimatum bargaining. In: Proceedings of the AAAI conference on human computation and crowdsourcing, vol. 8, pp. 43–52.

Friemel, C., Morana, S., Pfeiffer, J., & Maedche, A. (2018). On the role of users' cognitive-affective states for user assistance invocation. In *Information Systems and Neuroscience: Gmunden Retreat on NeuroIS 2017* (pp. 37-46). Springer International Publishing.

Gellersen, H. W., Schmidt, A., & Beigl, M. (2002). Multi-sensor context-awareness in mobile devices and smart artifacts. *Mobile Networks and Applications*, *7*, 341–351.

Ginting, Y., Chandra, T., Miran, I., & Yusriadi, Y. (2023). Repurchase intention of e-commerce customers in Indonesia: An overview of the effect of e-service quality, e-word of mouth, customer trust, and customer satisfaction mediation. *International Journal of Data and Network Science*, *7*(1), 329–340.

Gomez, M. I., McLaughlin, E. W., & Wittink, D. R. (2004). Customer satisfaction and retail sales performance: an empirical investigation. *Journal of Retailing*, *80*(4), 265–278.

Grubert, J., Langlotz, T., Zollmann, S., & Regenbrecht, H. (2016). Towards pervasive augmented reality: Context-awareness in augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, *23*(6), 1706–1724.

Gupta, K., Hajika, R., Pai, Y. S., Duenser, A., Lochner, M., & Billinghurst, M. (2020). Measuring human trust in a virtual assistant using physiological sensing in virtual reality. In *2020 IEEE Conference on virtual reality and 3D user interfaces (VR).* Symposium conducted at the meeting of IEEE.

Haring, K. S., Matsumoto, Y., & Watanabe, K. (2013). How do people perceive and trust a lifelike robot. In *Proceedings of the world congress on engineering and computer science*.

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Publications.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

Hornsey, M. J., Greenaway, K. H., Harris, E. A., & Bain, P. G. (2019). Exploring cultural differences in the extent to which people perceive and desire control. *Personality and Social Psychology Bulletin*, *45*(1), 81–92.

Hu, X. (2023). Empowering consumers in interactive marketing: examining the role of perceived control. In *The Palgrave Handbook of Interactive Marketing* (pp. 117–147). Springer.

Jones, K., & Leonard, L. N. K. (2008). Trust in consumer-to-consumer electronic commerce. *Information & Management*, *45*(2), 88–95.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354.

Kruschke, J. (2014). Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan.

Kulkarni, S., & Rodd, S. F. (2020). Context Aware Recommendation Systems: A review of the state of the art techniques. *Computer Science Review*, *37*, 100255.

Lallemand, C., & Koenig, V. (2020). Measuring the contextual dimension of user experience: development of the user experience context scale (UXCS). In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*.

Lao, J [J.], & Louf, R. (2020). Blackjax: A sampling library for JAX. *URL Http://github. Com/blackjax-Devs/blackjax*.

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, *35*(10), 1243–1270.

Luo, Y., Liu, F., She, Y., & Yang, B. (2023). A context-aware mobile augmented reality pet interaction model to enhance user experience. *Computer Animation and Virtual Worlds*, *34*(1), e2123.

Liao, Q. Vera; Gruen, Daniel; Miller, Sarah (2020): Questioning the AI: informing design practices for explainable AI user experiences. In: Proceedings of the 2020 CHI conference on human factors in computing systems, pp. 1–15.

M. J. Betancourt, Simon Byrne, & Mark Girolami. (2015). *Optimizing The Integrator Step Size for Hamiltonian Monte Carlo*.

Maedche, A., Morana, S., Schacht, S., Werth, D., & Krumeich, J. (2016). Advanced user assistance systems. *Business & Information Systems Engineering*, *58*, 367–370.

Martin, O. (2018). *Bayesian Analysis with Python: Introduction to statistical modeling and probabilistic programming using PyMC3 and ArviZ*. Packt Publishing Ltd.

Martin, O. A., Kumar, R., & Lao, J [Junpeng]. (2021). *Bayesian modeling and computation in python*. CRC Press.

McElreath, R. (2018). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC.

Meißner, M., Pfeiffer, J., Pfeiffer, T., & Oppewal, H. (2019). Combining virtual reality and mobile eye tracking to provide a naturalistic experimental environment for shopper research. *Journal of Business Research*, *100*, 445–458.

Miles, W. R. (1929). Ocular dominance demonstrated by unconscious sighting. *Journal of Experimental Psychology*, *12*(2), 113–126.

Möllering, G. (2005). The trust/control duality: An integrative perspective on positive expectations of others. *International Sociology*, *20*(3), 283–305.

Morana, S., Pfeiffer, J., & Adam, M. T. P. (2020). User Assistance for Intelligent Systems. *Business & Information Systems Engineering*, *62*(3), 189–192.

Moussawi, S., Koufaris, M., & Benbunan-Fich, R. (2021). How perceptions of intelligence and anthropomorphism affect adoption of personal intelligent agents. *Electronic Markets*, *31*, 343–364.

Muir, B. M., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, *39*(3), 429–460.

Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of personality and social psychology*, *89*(6), 852.

Nicoll, D. (1999). Taxonomy of information intensive products. *The University of Edinburgh Management School: Edimburgh, Scotland, UK*.

Novák, J. Š., Masner, J., Benda, P., Šimek, P., & Merunka, V. (2023). Eye Tracking, Usability, and User Experience: A Systematic Review. *International Journal of Human-Computer Interaction*, 1–17.

Nunes, Ingrid; Jannach, Dietmar (2017): A systematic review and taxonomy of explanations in decision support and recommender systems. In *User Model User-Adap Inter* 27 (3-5), pp. 393–444.

Olenberger, C. (2023). The effects of user assistance systems on user perception and behavior.

Oliveira, T., Alhinho, M., Rita, P., & Dhillon, G. (2017). Modelling and testing consumer trust dimensions in e-commerce. *Computers in Human Behavior*, *71*, 153–164.

Orquin, J. L., & Holmqvist, K. (2018). Threats to the validity of eye-movement research in psychology. *Behavior Research Methods*, *50*, 1645–1656.

Pan, H., & Zhang, Z. (2021). Research on context-awareness mobile tourism e-commerce personalized recommendation model. *Journal of Signal Processing Systems*, *93*, 147–154.

Panigrahi, S., Azizan, N. A., & Waris, M. (2018). Investigating the empirical relationship between service quality, trust, satisfaction, and intention of customers purchasing life insurance products. *Indian Journal of Marketing• January*.

Papadopoulou, P. (2007). Applying virtual reality for trust-building e-commerce environments. *Virtual Reality*, *11*, 107–127.

Patil, A., Huard, D., & Fonnesbeck, C. J. (2010). PyMC: Bayesian stochastic modelling in Python. *Journal of Statistical Software*, *35*(4), 1.

Petter, S., Straub, D., & Rai, A. (2007). Specifying formative constructs in information systems research. *MIS Quarterly*, 623–656.

Peukert, C., Lechner, J., Pfeiffer, J., & Weinhardt, C. (2020). Intelligent Invocation: Towards Designing Context-Aware User Assistance Systems Based on Real-Time Eye Tracking Data Analysis. In *Information Systems and Neuroscience* (pp. 73–82). Springer, Cham.

Pfeiffer, J., Pfeiffer, T., Meißner, M., & Weiß, E. (2020). Eye-Tracking-Based Classification of Information Search Behavior Using Machine Learning: Evidence from Experiments in Physical Shops and Virtual Reality Shopping Environments. *Information Systems Research*, *31*(3), 675–691.

Pietroszek, K. (2019). *Raycasting in Virtual Reality* [Encyclopedia of Computer Graphics and Games]. Springer International Publishing AG, part of Springer Nature 2018.

Polson, N. G., & Scott, J. G. (2012). On the half-Cauchy prior for a global scale parameter.

Poursabzi-Sangdeh, Forough; Goldstein, Daniel G.; Hofman, Jake M.; Wortman Vaughan, Jennifer Wortman; Wallach, Hanna (2021): Manipulating and measuring model interpretability. In: Proceedings of the 2021 CHI conference on human factors in computing systems, pp. 1–52.

Pu, P., & Chen, L. (2006). Trust building with explanation interfaces. In *Proceedings of the 11th international conference on Intelligent user interfaces*.

Quelch, J. A., & Klein, L. R. (1996). The Internet and international marketing. *MIT Sloan Management Review*, *37*(3), 60.

Rader, Emilee; Cotter, Kelley; Cho, Janghee (2018): Explanations as Mechanisms for Supporting Algorithmic Transparency. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. New York, NY, USA. New York, NY, USA: ACM.

Richthammer, C., & Pernul, G. (2020). Situation awareness for recommender systems. *Electronic Commerce Research*, *20*(4), 783–806.

Rijsdijk, S. A., Hultink, E. J., & Diamantopoulos, A. (2007). Product intelligence: its conceptualization, measurement and impact on consumer satisfaction. *Journal of the Academy of Marketing Science*, *35*, 340–356.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied*, *80*(1), 1.

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*.

Russo, J. E., & Leclerc, F. (1994). An Eye-Fixation Analysis of Choice Processes for Consumer Nondurables. *Journal of Consumer Research*, *21*(2), 274.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*.

Schilit, B., Adams, N., & Want, R. (1994). Context-aware computing applications. In *1994 first workshop on mobile computing systems and applications* (pp. 85-90). IEEE.

Skinner, E. A., Zimmer-Gembeck, M. J., Connell, J. P., Eccles, J. S., & Wellborn, J. G. (1998). Individual differences and the development of perceived control. *Monographs of the Society for Research in Child Development*, i-231.

van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., Vannucci, M., Gelman, A., Veen, D., Willemsen, J., & Yau, C. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers*, *1*(1), 1–26.

van de Schoot, R., Depaoli, S., & others (2014). Bayesian analyses: Where to start and what to report. *The European Health Psychologist*, *16*(2), 75–84.

van den Bergh, D., van Doorn, J., Marsman, M., Draws, T., van Kesteren, E.-J., Derks, K., Dablander, F., Gronau, Q. F [Quentin Frederik], Kucharský, Š., Raj, A., Sarafoglou, A., Voelkel, J. G., Stefan, A. M., Ly, A., Hinne, M., Matzke, D., & Wagenmakers, E.-J. (2019). *A Tutorial on Conducting and Interpreting a Bayesian ANOVA in JASP*. Center for Open Science.

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F [Quentin F.], Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Gupta, Akash R. Komarlu Narendra, Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E.-J. (2021). The JASP guidelines for conducting and reporting

van Hove, S., Herregodts, A.-L., Schuurman, D., & Marez, L. de (2017). Assessing user experience of context-aware interfaces in a retail store. In *ISPIM Innovation Symposium.* Symposium conducted at the meeting of The International Society for Professional Innovation Management (ISPIM).

Vos, H. de, Haaker, T., Teerling, M., & Kleijnen, M. (2009). Consumer value of context aware and location based mobile services. *International Journal of E-Services and Mobile Applications (IJESMA)*, *1*(4), 36–50.

Wang, Q., Yang, S., Liu, M., Cao, Z., & Ma, Q. (2014). An eye-tracking study of website complexity from cognitive load perspective. *Decision Support Systems*, *62*, 1–10.

Weiß, T. (2024). *Adaptive user assistance in virtual commerce dataset and source code*. http://osf.io/9jcpn

Yang, F., Huang, Z., Scholtz, J., & Arendt, D. L. (2020). How do visual explanations foster end users' appropriate trust in machine learning? In *Proceedings of the 25th international conference on intelligent user interfaces*.

Yuan, Y., & MacKinnon, D. P. (2009). Bayesian mediation analysis. *Psychological Methods*, *14*(4), 301–322.

Zhang, X., & Uruchurtu, E. (2011). A user experience perspective of design for context-aware adaption. In *2011 Seventh International Conference on Intelligent Environments.* Symposium conducted at the meeting of IEEE.

Zhang, Y., Liao, Q. V., & Bellamy, R. K. E. (2020). Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*.

Zhao, W., Cheng, Y., & Lee, Y.-I. (2023). Exploring 360-degree virtual reality videos for CSR communication: An integrated model of perceived control, telepresence, and consumer behavioral intentions. *Computers in Human Behavior*, *144*, 107736.

6.8 Supplemental material

6.8.1 Supplemental figures

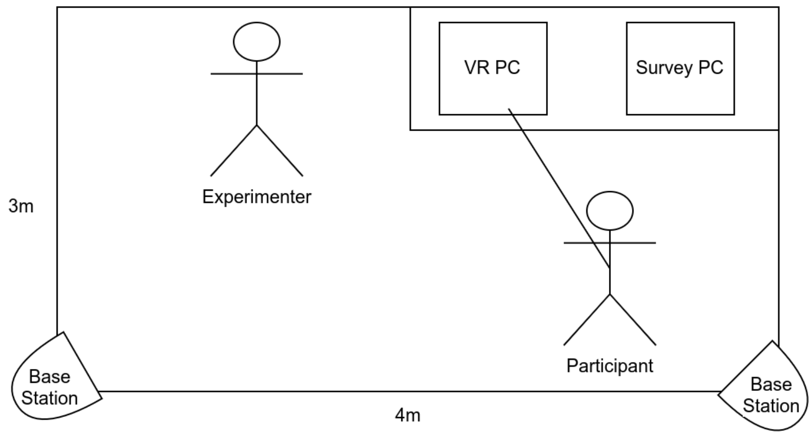

Figure 37. The room layout. A 3x4m laboratory environment with a dedicated VR computer and a survey computer.

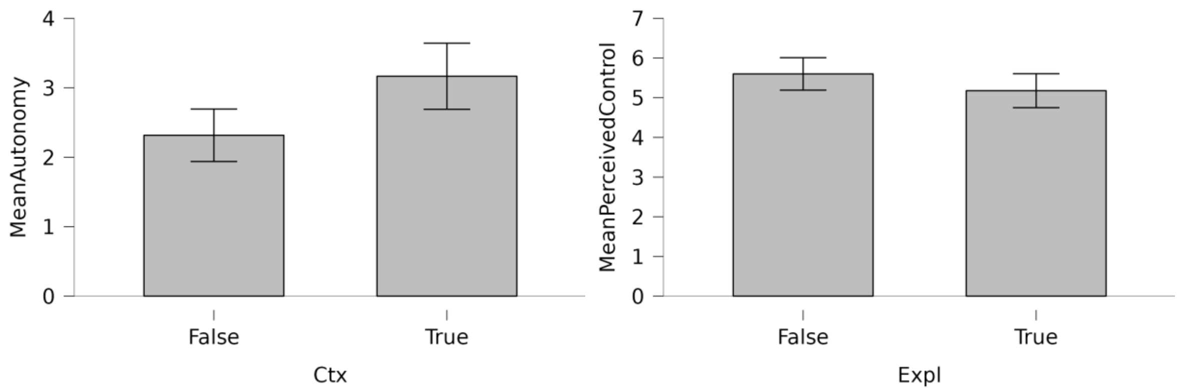

Figure 38. Left: Mean perceived autonomy conditioned on the UAS type (context-aware versus basic) visually interesting. Right: Mean perceived control conditioned on the Explanation visually less strong, but still interesting.

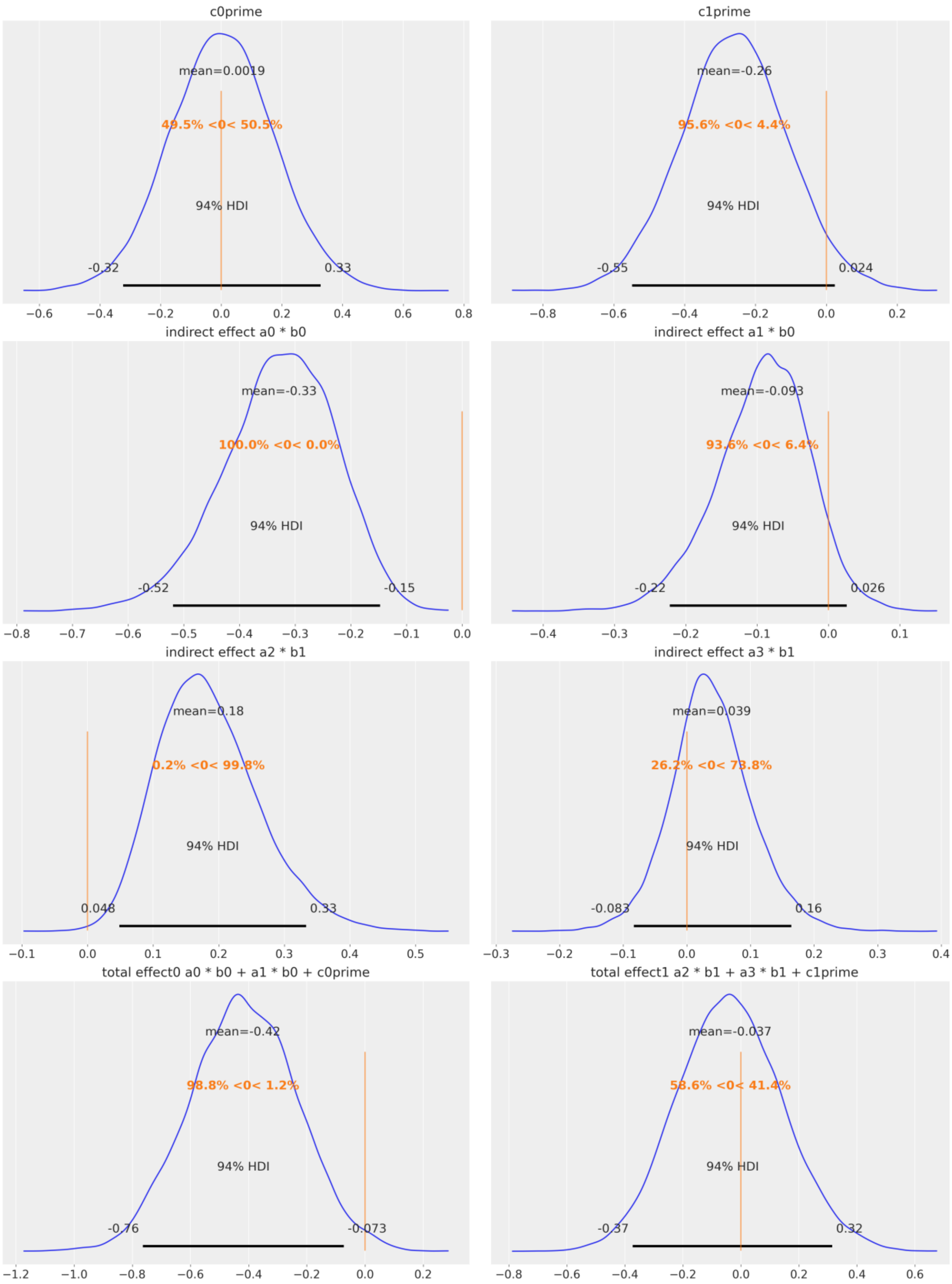

Figure 39. Posterior distributions for the Student-t Dsp mediation model.

6.8.2 Supplemental tables

Table 17. Reliability of the latent constructs.

Construct	Cronbachs α	95% CI
Intelligence	0.927745	[0.907 0.945]
Reactivity	0.867878	[0.825 0.903]
Ability to learn	0.958906	[0.946 0.969]
Autonomy	0.835089	[0.776 0.88]
Perceived control	0.851598	[0.799 0.892]
Intrusiveness	0.832579	[0.778 0.877]
Trusting Beliefs	0.814369	[0.761 0.86]
Predictability	0.864595	[0.82 0.9]
Benevolence	0.79075	[0.716 0.848]
Competence	0.879306	[0.836 0.912]
Disposition to trust	0.864862	[0.806 0.906]
User Experience	0.829567	[0.779 0.872]

Table 18. Factor loadings for perceived intelligence items.

	Factor 1	Factor 2	Factor 3	Uniqueness
Abilitytolearn[IL003]	1.046			0.077
Abilitytolearn[IL002]	0.930			0.169
Abilitytolearn[IL004]	0.889			0.251
Abilitytolearn[IL001]	0.816			0.213
Autonomy[IA002]		0.899		0.375
Autonomy[IA003]		0.837		0.316
Autonomy[IA001]		0.742		0.372
Reactivity[IR004]			0.947	0.219
Reactivity[IR003]			0.844	0.439
Reactivity[IR002]			0.530	0.266
Reactivity[IR001]			0.440	0.337

Note. Applied rotation method is promax.

Table 19. Factor loadings for perceived trust.

	Factor 1	Factor 2	Uniqueness
Benevolence[TB002]	0.815		0.371
Competence[TC003]	0.793		0.342
Competence[TC001]	0.722		0.430
Competence[TC002]	0.713		0.340
Benevolence[TB001]	0.691		0.533
Benevolence[TB003]	0.576		0.686
Predictability[TP002]		0.933	0.168
Predictability[TP001]		0.833	0.332
Predictability[TP004]		0.803	0.340
Predictability[TP003]		0.588	0.618

Note. Applied rotation method is promax.

Table 20. Factor correlation with manually adjusted number of three factors.

	Factor 1	Factor 2	Factor 3
Factor 1	1.000	0.306	0.050
Factor 2	0.306	1.000	0.637
Factor 3	0.050	0.637	1.000

Table 21. Analysis of effects on trust.

Effects	P(incl)	P(excl)	P(incl|data)	P(excl|data)	BFincl
Ctx	0.600	0.400	0.216	0.784	0.184
Expl	0.600	0.400	0.589	0.411	0.954
Ctx ✻ Expl	0.200	0.800	0.029	0.971	0.120

6.8.3 Explanations

6.8.3.1 Basic UAS

How: The UAS has been available at all times, so that the help could be used immediately from the beginning. It has not adapted to your behavior.

Why: The UAS was primarily intended to help you compare products (product comparison matrix). Once you started comparing, you could directly use the UAS to support you.

6.8.3.2 Context-aware UAS

How: The UAS decided when to offer you help based on real-time analysis of eye tracking data. It did this by analyzing your eye movements. Once you made a pairwise comparison between two products, the UAS was provided to you. The system has therefore behaved adaptively.

Why: The UAS was primarily intended to help you compare products (product comparison matrix), which is why the help was only offered to you when you made the first comparison between two products. Only from this point on was the UAS helpful for you. Before, the comparison matrix could have hindered you from gaining the overview by blocking your field of view.

6.8.4 Tasks

6.8.4.1 Task 1

A good friend is coming to visit you over the weekend, which is why you would like to get some muesli for breakfast. You know that your guest prefers muesli with chocolate. Your guest also likes muesli from the Koelln brand. Specifically, your guest prefers a pure chocolate muesli without any other flavors like cookies or crunch. Additionally, you are aware that your guest has a peanut allergy, so the product must not contain it.

Finally, the muesli should be as cheap as possible.

Figure 40. Solution for Task 1.

6.8.4.2 Task 2

A good friend is coming to visit you over the weekend, so you would like to get some muesli for breakfast. You know that your guest prefers muesli with fruits. Therefore, the muesli should not only contain one fruit but should include several different fruits (at least two different ones). The muesli should also be low in sugar or have no added sugar. Lastly, the muesli should contain as little fat as possible (per 100g).

Figure 41. Solution for Task 2.

6.8.4.3 Task 3

A good friend is coming to visit you over the weekend, so you would like to get some muesli for breakfast. You know that your guest prefers crunchy muesli. Additionally, since your guest enjoys eating chocolate, the muesli should be a crunchy muesli with chocolate (at least partially chocolate as an ingredient). You also want to buy a muesli with as much content in the packaging as possible. Lastly, the muesli should have as few calories (kJ/kcal per 100g) as possible.

Figure 42. Solution for Task 3.

Declaration of Authorship / Selbstständigkeitserklärung

I hereby declare that I have prepared the submitted essays independently and only with the help specified for the respective essay. As stated, I was partially involved in the collaboration with the co-authors listed. In the investigations I carried out and mentioned in the articles, I adhered to the principles of good scientific practice, as laid down in the statutes of the Justus Liebig University of Gießen to ensure good scientific practice.

Hiermit erkläre ich, dass ich die vorgelegten Aufsätze selbstständig und nur mit den Hilfen angefertigt habe, die für den jeweiligen Aufsatz angegeben sind. In der Zusammenarbeit mit den angeführten Koautoren war ich wie angegeben anteilig beteiligt. Bei den von mir durchgeführten und in den Aufsätzen erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis eingehalten, wie sie in der Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis niedergelegt sind.

17.09.2024

Date / Datum

Signature / Unterschrift