IMMERSIORAMA

immersive, yet controlled

Farbod N. Nezami

WEST DRIVE

Immersiorama
immersive yet controlled

Cognitive Science in Virtual World, An argument for virtual reality as an improvement to laboratory

Dissertation
zur Erlangung des Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)
im Fachbereich Humanwissenschaften der

Universität Osnabrück

vorgelegt von

Farbod Nosrat Nezami

Supervisors and promotion's committee:

Prof. Dr. rer nat. Peter König, Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany

Prof. Dr. rer nat. Gordon Pipa, Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany

Dr. Lewis Chuang, Institut für Informatik, Ludwig-Maximilian-Universität München, München, Germany & Abt. Ergonomie, Leibniz Institut für Arbeitsforschung, Leipzig, Germany

Osnabrück, Universität Osnabrück, 2022

Cyperspace is thought of as the ultimate virtual reality environment.
It is an alternative computer universe where data exists like cities of light.
Information workers use a special virtual reality system to enter cyberspace and to travel its data highways.

— J.M Zheng, K. W. Chan, L. Gibson 1998

-rama: noun suffix meaning "sight, view, spectacular display or instance of," 1824, abstracted from panorama (q.v.), ultimately from Greek horama "sight, spectacle, that which is seen."

Therefore, Immersiorama is our attempt to immerse the viewer in an immersive experience in order to observe the naturalistic behaviors elicited from the spectacle.

Acknowledgements

First and formost I would like to express my deepest gratitude to Prof. Dr. Peter König and Prof. Dr. Gordon Pipa for their wondrous support, supervision and advice during the period of my Ph.D. I would like to thank you for invested time and amazing support both on personal level the utter will and support for the realization the project, which are presented in this dissertation, despite any difficulties. And for providing me with such a great environment and opportunity to work, learn and grow both as a scientist and a person.

This extend as well to Dr. Sabine König. I would like to thank you for your unprecedented support and consultation which empowered me to improve myself and my writing which is reflected in this dissertation.

I also would like to thank Maximilian Alexander Wächter for your company, support and brilliant enthusiasm and energy since the start of our join master thesis and later Ph.D.on project Westdrive both as my colleague and friend, without you this endeavour would not be possible at it stands today.

Osnabrück, 17 December 2021 F.N. Nezami.

I want to offer an special thank to all members of Team Westdrive who have accompanied me during the past three years. I would like to thank you for your commitment and support which made this work possible. With this, i would like to thank:

Johannes Maximilian Pingel, Philipp Spaniol, Hristofor Lukanov, Marc Vidal dePalol, Frederik Nienhaus, Anke Haas, Lea Maria Kühne, Thomas Klein, Deniz Gün, Linus Tiemann, Sumin Kim, Nora Maleki, Lynn Keller, Richard Ruppel, Shadi Derakhshan, Stefan Balle, Lucas Essmann, Vincent Schmidt and Dr. Sylwia Kaduz - and many more people from the Neurobiopsychology and the Neuroinformatics of the IKW.

I want to further offer my sincere gratitude to my colleagues and friends Artur czeszumski and Ashima Keshava who have both supported and help me many times both on my personal journey and scientific endeavour and learning during my Ph.D..

Last but not least I would like to thank to my family and Xinzhuo Xiao withon who have supported me in every step of the way and without whom this work were not possible.

Abstract

Neuroscience, psychology, and many other fields, such as anthropology or philosophy, try to understand our cognition and cognitive processes. However, as time passed, new views on cognition emerged. One of the newest views on cognition, known as 4E cognition, refers to embedded, embodied, extended, and enacted cognition. Alternatively, to put it in simpler terms, our cognition and cognitive processes emerge from us by being in our environment, interacting with our environment, and enacting our actions within our environment. Although the need to study human cognition from a higher perspective led to the emergence of cognitive sciences, despite these advancements, our experimental methods have stayed relatively unchanged for the past centuries.

The recent trends in cognitive science and related fields lean toward real-world experimentation. The main argument for real-world experimentation is the ecological validity of our experimentation and finding. However, despite all the positive voices advertising for real-life experimentation, there are also significant concerns and voices against such a movement. Real-world is full of dynamics and sources of noises and events no one has studied in detail before. As alluring as the idea of moving out of the lab and doing experiments in real life is, the challenges of real-life experimentation should not be neglected, at least with our current methods and tool kits.

However, one does not need to entirely abandon the control of the lab environment to get closer to real-life experimentation. Immersive virtual reality experiences can offer a close to the real-life and interactive foundation for conducting cognitive science experiments. Virtual reality experiments can offer the same level of control over the conditions and precision in measurements as laboratory-based experimentation yet enable a realistic, immersive environment to simulate real-life situations.

This dissertation seeks to investigate the ecological validity of immersive virtual reality experimentation. The investigation tries to see if virtual reality experimentation can augment the lab-based experiments to simulate closer to real-life situations. The second point of focus is on the notion of ecological validity. Here we tried to investigate which factor among realistic cues, environment, or interaction with the environment plays a vital role in improving the findings of cognitive science experiments. This dissertation seeks to answer these questions with different experiments made and conducted using immersive virtual reality simulations. These studies first investigate virtual reality technologies' current state of the art. These experiments push the limits of what others previously performed in virtual reality experimentation in terms of immersion and realism. We studied ecological validity using these environments. This work examines the hypothesis that "realism" indeed matters and, more importantly, that realism in the interaction with the environment can give us more understanding regarding our observations. Finally, we will observe participants in their behavior using virtual reality experiments with minimal to no intervention to validate the effectiveness of virtual reality experimentation.

Of course, the studies presented in this work also have further research questions to answer. These research questions include Gaze behavior during tool interaction or planning while sorting objects on a shelf is an example of investigating low-level cognitive processes. The role of perspective on the moral judgments in trolley dilemma situations or change of attitude and acceptance toward self-driving vehicles is more on the psychological aspects of cognition. However, when added together, the observations gained in each study offer solid arguments toward not only the benefits of virtual reality experimentation but the importance of studying cognition within a natural context in real work with naturalistic interactions. This dissertation provides arguments in favor of virtual reality as a suitable experimentation tool and environment in the absence of standard and precise real-life experimentation methods as a way to simulate real-life experiences in our experiments.

Contents

Publication List

Journal Articles

Kallioinen, N., Pershina, M., Zeiser, J., Nezami, F. N., Pipa, G., Stephan, A., & König, P. (2019). *Moral Judgements on the Actions of Self-Driving Cars and Human Drivers in Dilemma Situations From Different Perspectives*. Frontiers in Psychology, 10.

Nezami, F. N., Wächter, M. A., Pipa, G., & König, P. (2020). *Project Westdrive: Unity city with self-driving cars and pedestrians for virtual reality studies*. Frontiers in ICT, 7, 1.

Nezami, F. N., Wächter, M. A., Maleki, N., Spaniol, P., Kühne, L. M., Haas, A., Pingel, J. M., Tiemann, L., Nienhaus, F., Keller, L., König, S. U, König, P. & Pipa, G. (2021). *Westdrive X LoopAR: An Open-Access Virtual Reality Project in Unity for Evaluating User Interaction Methods during Takeover Requests*. Sensors, 21(5), 1879.

Nezami, F. N., Wächter, M. A., Keshava, A. Lukanov, H. Vidal de Palol, M., Pipa, G. & König, P. (*in preparation*). *Talking Cars increase trust, but not the intention to use them*.

Keshava, A., Gottschewsky, N., Balle, S., Nezami, F. N., Schüler, T., & König, P. (2021). *(in preparation)Action Affordance Affects Proximal And Distal Goal-Oriented Planning*. Biorxiv. doi:10.1101/2021.07.27.454022

Keshava, A., Nezami, F. N., Neumann, H., Izdebski, K., Schüler, T., & König, P. (2021). *(in preparation)Just-in-time: Gaze guidance behavior while action planning and execution in VR*. Biorxiv. doi:10.1101/2021.01.29.428782

Preprint Articles

Nezami, F. N., Wächter, M. A., Maleki, N., Spaniol, P., Kühne, L. M., Haas, A., Pingel, J. M., Tiemann, L., Nienhaus, F., Keller, L., König, S. U, König, P. & Pipa, G. (2020, May 25). *From Interaction to Cooperation: a new approach for human-machine interaction research for closing the out-of-the-loop unfamiliarity.* https://doi.org/10.31234/osf.io/7jg3c

Invited Talks

Nezami, F.N., Wächter, M. A. (December, 2018) *Innovation and Regulation regarding Self-Driving Cars*, Workshop for the profile line 1 "Digitale Gesellschaft - Innovation - Regulierung".

Nezami, F. N., Wächter, M. A. (June, 2019) *Projekt Westdrive: Mensch-Maschine Interaktion in der virtuellen Realität*, BMBF KarliczekImpulse Wissenschaftsjahr 2019

Nezami, F. N., Wächter, M. A. (July, 2019) *Selbsterklärende künstliche Intelligenz und virtuelle Autos: Ein Forschungprojekt der Uni Osnabrück und der Stiftung Stahlwerk Georgsmarienhütte*, IdeenExpo Hannover

Nezami, F. N., Wächter, M. A., Pipa, G. (February, 2020) *Potenziale der Künstlichen Intelligenz*, Wirtschaftsvereinigung Grafschaft Bentheim

Nezami, F. N., Wächter, M. A. (July, 2021) *Virtual Reality - Von menschlichen Verhalten bis zur Städteplanung*, OSNAhack 2021, Osnabrück

Weising M. , Nezami F. N., Maleki N. (September, 2021) *Workshop on experiment design in VR*, 3rd International Neuroergonomics Conference 2021, Munich

Keshava A., Nezami F. N., Maleki N., Tiemann L., & König P. (September, 2021) *Stress testing VR Eye-tracking System Performance*, 3rd International Neuroergonomics Conference 2021, Munich

Maleki N., Mildt M., Pätzold F., Schmidt V., Tiemann L., L Walter L. J., Zerbe A. J., Gütlin C. D., Haas A., Lang A., N Nezami F. N., König P., & Czeszumski A. (September, 2021) *A framework for low-level joint action in VR*, 3rd International Neuroergonomics Conference 2021, Munich

List of Figures

List of Tables

1 General Introduction

In the current state of technology and research in cognitive science as well as related fields which try to understand human cognitive processes, there should be a middle ground that can optimize and maximize the validity and generalization of scientific findings to the real world. The 4E cognition banner tells us that cognition is embedded, embodied, extended, and enacted. Under these premises one realizes that it is not possible to study human cognition in an environment deprived of naturalistic social and environmental interactions. Furthermore, it is not crucial whether the ecological validity only refers to the experimental cues or the experimental setup. After all, the final goal in any study is to reach a valid and generalizable theory that explains the observed phenomenon and extend these finding to explain the naturalistic responses and behavior in the real world.

This dissertation argues that virtual reality can offer a great degree of freedom to researchers in cognitive science and similar fields. The ability to be "present" in an environment, albeit virtual, and interact with objects presented in the environment as well as acting closer to the natural behavior satisfied the premises of the 4E cognition. Here, depending on the needs of the hypothesis, the researcher can come up with both ecologically valid stimuli and the environment under both strict or broad definitions of the term. The strict definition follows more in the direction of the original guidelines of Brunswick (Brunswik, 1955 , Brunswick, 1956) whereas the broad definition follows the suggestions of Shamay-Tsoory and Mendelsohn (Shamay-Tsoory and Mendelsohn, 2019) regarding ecological validity. The degree of freedom offered by virtual reality experiments is that one can design experiments with any desired complexity or simplification and still achieve the laboratory level of control over the stimuli presentation and the experimental procedure.

However, the benefits of settling on Virtual reality as a middle ground between laboratory and real-life in studying cognition is not limited only to the freedom it offers in experiment design. The recent innovations and advancement in the design of the virtual reality hardware as well as other sensory physiological measurement hardware such as but not limited to eye tracker, ECG, and skin conductance alongside mobile brain/body imaging techniques (MoBI), enables a greater variety of experiments to be conducted in virtual environments. Today's interconnectivity through advancements in computer networking can provide a foundation for almost instantaneous realistic social interactions in the scale possible before. Therefore one can look at virtual reality itself as the new laboratory.

However, before diving deeper into the investigation of virtual reality as a middle ground for performing experiments, one must first understand where we stand regarding our experiments. To start investigating how researchers conduct most cognitive science experiments today, one must first understand when this field emerged. The first-ever neuro-surgery on the spinal cord is estimated to go as far back as 1700 BC in ancient Egypt, evident from Edwin Smith surgical papyrus (Hughes, 1988). However, the study of brain and behavior as it became known as neuroscience and psychology today are relatively modern fields of study. It was only in the 20 century where others in science started to view the study of the brain and its function as its field of research and independent of other fields and medical research (Cowan et al., 2000). The same story goes for the field of experimental psychology. Although psychology as a field has existed since the late 18th century, the emergence of modern experimental psychology only goes back to the 1830s in Leipzig, Germany (Leahey, 1991). Although these fields are among the contemporary fields of studies, psychology and neuroscience and related areas of studies are all considered among empirical research fields today.

Since empirical research values hypothesizing and observation and later clear statistical analysis, it requires a well-designed and thought-through experiment (Harrington, 2010). Empirical research often calls for a well-controlled experimental setup. Such setup requires a set pre-determined dependant and independent variables and hypothesis. Additionally, the empirical design tries to minimize any unexpected random effects and factors that might otherwise alter the result of the observations. This clearly structured mindset in study design consequently improves the replicability of the experiments.After all the main goal of an empirical experiment is not to prove any hypothesis but to find statistically valid evidence supporting it. However, this call for controlling all factors in a study to predefined hypotheses usually means that researchers have to oversimplify and focus their experiments on a single hypothesis. Although the current methods in empirical research led to many valid discoveries, some would argue against the ecological validity of such simplifications when it concerns our complex human brain neuronal connectivity and human behavior.

The most recent view on human cognition is that our cognition is embodied, embedded, extended, and enacted (Newen et al., 2018). The embodiment of cognition means that the scope of our cognition is not just bound to our brain functionality, but it requires the body. The embedded cognition means that our cognition emerges from the interaction of our body with the environment. Extended cognition is because our brain and body do not exist in a separate reality, but the environment and the objects we interact with can become part of our cognition. Lastly, cognition enacted cognition means that our cognitive processes emerge with enacting actions within the environment (Newen et al., 2018). This view on cognition means that to understand different cognitive processes truly, we cannot separate the subject from the natural environment and objects they use for a specific cognitive process under the study. And here, virtual reality seems to be able to satisfy the premises of this view.

The empirical research cycle consists of five consecutive phases. It starts with observation of a phenomenon and starting the investigation of its causes. Afterwards is the induction phase, where the researcher tries to formulate a hypothesis that can generalize the explanation for the phenomenon. The third phase is where the researcher designs the experiment, to test the hypotheses they came up with within the induction phase for their validity. From here on, the researcher will run the experiment and will try to evaluate the data and evidence gathered in order to formulate a theory about the roots and causes of the phenomenon (Heitink, 1999). The goal of this dissertation is not to discredit the way empirical research has been conducted so far but to improve the deduction and testing phases in order to better investigate the natural causes of the phenomenon considering the points made about our cognition. One of the issues with complete real-life experimentation is in the design phase of empirical research. It is in this phase that will become more challenging for the researcher to come up with a good design.

The issue stems from the inherit hardship to control what exact cues and stimuli the participants will perceive.Moreover, the researcher cannot be sure of the environmental effects on the main observation and hypothesis under investigation. In other words the unpredictability of real world environments makes it one of the issues to design a well designed experiments to be conducted in real world.. Here virtual reality offering complete control over the cues and stimuli while simulating the real world, is therefor, a suitable tool to conduct the experience in.

Moreover, to what we learned about our cognition,it is said that humans are social beings (Shamay-Tsoory and Mendelsohn, 2019). Meaning we are not just simply an isolated brain in a jar, but rather our behavior is also entangled with our social interaction with other beings, be it other humans, animals, or robots. therefor human beings being social beings consequently means, how and under what circumstances (i.e., cooperation, competition, solo) specific phenomenon or interaction took place can dramatically affect the underlying cognitive, behavioural, and neuronal processes under investigation. (Shamay-Tsoory and Mendelsohn, 2019, Ladouce et al., 2017). Therefore, it only makes sense to also investigate human cognitive processes in the context of our natural environment replicating our naturalistic interaction with the object in the environment under realistic social circumstances. The study of cognition is far from perfect. Researchers have already been performing experiments and discovered many underlying low-level processes in our cognition. Nowadays, lab-based experiments have improved a lot from decades ago, and if fact we are performing these experiments almost at the highest possible quality.

Nevertheless, lab-based experimentation has its limitations, especially considering what was mentioned above about the new views on our cognition. Therefore, the new movement toward real-life experimentation becomes the logical next step for future research in the field. Therefore, In this introduction, we will discuss more arguments in favor of this movement. However, ecological validity as mentioned earlier regarding scientific findings also depends on very well design experiments Therefore, this dissertation will try to provide enough evidence and arguments in favor of virtual reality a well-balanced compromise to investigate human cognition in real world.

1.1 From laboratory to real life

There are already strong voices in critique of ecological validity of Lab-based experiments in neuroscience, behavioural studies and psychology (Shamay-Tsoory and Mendelsohn, 2019, Ladouce et al., 2017, Griffiths, 2015, Brunswik, 1943). Over 60 years ago, the cognitive revolution legitimized the scientific study of cognition in its form today as cognitive science(Figure 1.1) (Griffiths, 2015; G. A. Miller, 2003). As it is observable from this point of view (Figure 1.1) study of cognition requires a higher level of investigation, and it is indeed a interdisciplinary endeavor. It requires higher level understanding of previously low level independent fields of study with their own unique approaches to experimentation. Therefore, this new approach of studying cognition also requires a change in our experimentation to reflect this interdisciplinary and higher-level investigation. There new given formal models of cognition and cognitive science made it possible to investigate cognitive processes in between people's history and their actions in contrast to traditional Behaviourism's stimuli-response methods(Griffiths, 2015). However, after decades of research in the field, our methods stayed relatively similar to the previous ones in the past (Mandler, 2011, Griffiths, 2015). Therefor, to answer a question about the brain, one comes up with a set of hypothesis, brings the proper amount of people to the lab to participate in a task in order for the researcher to evaluate their hypotheses(Griffiths, 2015).

There is increasing evidence that the traditional reductionist cognitive science overlooked important aspects of cognitive science (Ladouce et al., 2017). The main issue raised is mostly a critique toward current lab-based experiments with artificial stimuli and fixed responses. Consequently these findings have lower ecological validity as compared to the real-world behaviour (Shamay-Tsoory and Mendelsohn, 2019, Ladouce et al., 2017). As mentioned in the 4E Cognition[I] movement, human cognition is intertwined with actions and environment as well as social circumstances (Newen et al., 2018). Therefore, traditional lab-based experiments often focus on investigating the cognition regarding participants as isolated agents. Researchers often conduct experiments in artificial, sensory, and socially deprived environments (Figure 1.2). This form of approach inevitably limit our understanding of the naturalistic cognitive, emotional and social phenomenon

[I] embodied, embedded, extended and enacted

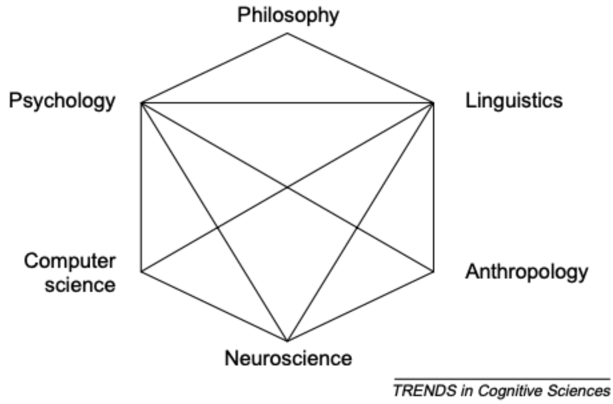

Figure 1.1: Cognitive science as proposed in 1978 where each line represent and interdisciplinary study that existed at the time.
G. A. Miller, 2003

(Shamay-Tsoory and Mendelsohn, 2019).

In this argument, real-world experiments indicate experimentation and measurements conducted in an environment that is relevant to everyday life (Shamay-Tsoory and Mendelsohn, 2019). In recent years these critiques toward lab-based experiment and the trend of moving toward real-life environments gained more and more support and argumentation (Shamay-Tsoory and Mendelsohn, 2019, Ladouce et al., 2017, Griffiths, 2015, Zaki and Ochsner, 2009, Kingstone et al., 2003). The low ecological validity of lab-based experiments has been criticized already in 1991 by one of the founders of cognitive science, Neisser, who showed his disappointment of studies with low ecological validity (Shamay-Tsoory and Mendelsohn, 2019) concerning memory writing "In the study of memory as else-where in psychology. There are certain to be more and more naturalistic studies in the years to come. Many of them will be less than outstanding in quality, but that is part of science: No one rejects evolutionary biology out of hand just because some Darwinian studies are flawed." (Neisser, 1991). With the improvement in hardware, especially that of mobile brain imaging such as mobile EEG and mobile FNIR, the so-called MoBI technology, more cognitive and neuroscience researchers are moving toward conducting their experiments in real-life environments. (Griffiths, 2015; Ladouce et al., 2017; Parada and Rossi, 2020).

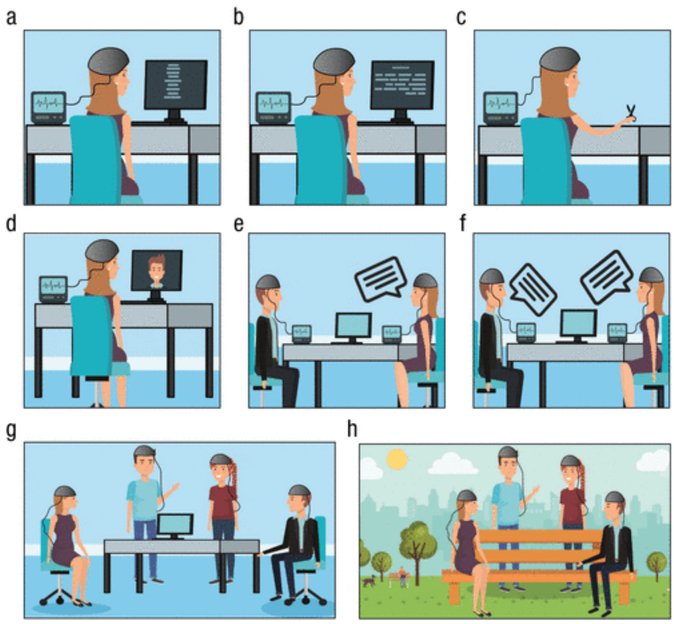

Figure 1.2: An overview of current experiment types performed in cognitive science according to the work of Shamay-Tsoory and Mendelsohn, 2019. (a) traditional lab-based experiments. It depicts isolated motionless participants with artificial stimuli as a meaningless list of words that is out of real-world context where the participant responses are limited and cannot affect the situation("person dependant limitation" Shamay-Tsoory and Mendelsohn, 2019. (b) The participant is presented with meaningful stimuli, but the rest is the same ("situation dependant limitation" Shamay-Tsoory and Mendelsohn, 2019). (c) the participant can explore an object with limited movement and therefore show a higher level of actions and cognitive activity but has no control over the context. (d) the participant is presented with artificial social stimuli and is unable to move. (e and f) lab-based unidirectional and bidirectional dyadic interaction but with limited context. (g) social interaction in order to investigate lab-based group dynamics, (h) real-life multi-directional interaction. The final approach allows for understanding and measuring social interaction in a real-life situation.

Shamay-Tsoory and Mendelsohn, 2019

Like many voices advocating the path of real-life experimentation, there is also opposition or concerns again this approach. One main critiques, although supporting the main goal of augments such as that of Shamay-Tsoory (Shamay-Tsoory and Mendelsohn, 2019), is the misuse of the term "ecological validity" as it was introduced by Egon Brunswick (Kihlstrom, 2021, Holleman et al., 2020, Brunswik, 1955, Brunswick, 1956). These research both mentions that Hammond, a student of Brunswick, already had pointed out the misuse of the term "ecological validity" decades ago (Hammond, 1998. In Brunswick's definition of the term "ecological validity" refers to the ecological validity of the cue rather than the experiments (Hammond, 1998 , Holleman et al., 2020 , Kihlstrom, 2021). The first misuse of the term happened in an article from 1962 by Martin T. Orne. The main argument of Orne in his article is that psychologists treat human subjects by mistake as "passive responders to the experimental stimuli." but actually, the perception of the subject from the stimuli might differ from the intention of the experimenters (Kihlstrom, 2021). However, arguably, in this case, even Kihlstrom, who brought up the issue with the misuse of the term, admits that the usage of the term in this context is not entirely against the original term coined by Brunswick (Kihlstrom, 2021).

Whether misused or not, nonetheless, researchers are mainly interested in the ecological validity of the experiments themselves (Kihlstrom, 2021). Furthermore, admittedly this concern does not limit to only a specific field such as social psychology. Many others, including Neissen, mentioned that "studies employing nonsense syllables and other verbal-learning paradigms lacked ecological validity and had taught us virtually nothing about memory in real life" (Kihlstrom, 2021). The arising issue here is, as pointed by Holleman, that term "ecological validity" as Brunswick defined neither mentioned that the experimental research should resemble the real life, or be close to it, nor implied that higher ecological validity would guarantee generalization to the real world (Holleman et al., 2020). Here one can argue that what matters, in the end, is the validity and generalizability of the scientific findings in explaining a real-life phenomenon.

When in the end, it comes down to the validity of scientific findings, there are different points to consider regardless of which definition of ecological validity one chooses to accept. According to Kihlstrom, in our modern scientific terminology. The original definition offered by Brunswick concerns itself with the experimental cues. The revisionist definition by Orne concerns itself with the "experiment" itself. A more modern and loose definition offered by Shamay-Tsoory and Mendelsohn, which Kilhstrom and Holleman call "mundane reality" (Kihlstrom, 2021, Holleman et al., 2020). No matter which definition we take, if the cues presented in an experiment are far from what people encounter in real life, it will lower the experiment's ecological validity. Therefore, at the end of this long discussion of a terminology, what matters is higher ecological validity, be it in its traditional sense or the modern one, which will improve and help us investigate and understand human cognitive processes better. After all, even Kihlstrom and Holleman, both writing against misuse of the terminology, supports Shamay-Tsoory and Mendelsohn in their endeavor and the point they try to portrait (Kihlstrom, 2021, Holleman et al., 2020).

However, a more valid question to ask today is if we are, if at all, ready for real-life experimentation. When moving toward real-world experimentation, there are many challenges the cognitive science community should overcome. Currently the mobile brain/body imaging community, better known as MoBI, is at the head of this movement toward real-world experimentation. They advocate using mobile brain imaging techniques to observe cognitive acts occur naturally with all their complexities (Parada, 2018). This observation can happen in either semi- or unstructured setup (Parada and Rossi, 2020). However, there is still a lack of general agreement on what constitutes a "MoBI experiment" (Parada and Rossi, 2020). Here MoBI movement refers to those who try to take the neuroimaging techniques out of the lab and bring them to the real world. The next challenge is that what physiological, behavioral, or/and neuronal data should be measured and to what extent these measurements can avoid interfering with the subject's naturalistic actions (Parada, 2018). These challenges are just the very general challenges facing those who follow the movement toward real-world experimentation.

More fundamental issues can arise from a reductionist, simplified lab experiment to a broad naturalistic observation in a real-life experiment. One study provides evidence that human attention allocation differs between the laboratory and the real world. (End and Gamer, 2017). In laboratory-based experiments, the head and body movements are most of the time restricted, Foulsham et al., 2011 has shown that the saccadic eye movements are fundamentally different in the laboratory and natural environment. Moreover, since the whole movement of real-life experimentation is relatively recent, where the MoBI community movement started around 14 years ago (Parada, 2018), there might be many more underlying differences in various cognitive processes that we are essentially unaware of. Therefore anyone who is supportive and tries to move toward real-world experimentation should be aware of the possible issues and differences arising from investigating human cognitive processes in their full complexity under a dynamic and complex real-life environment.

Nowadays the traditional experimentation in all cognition-related fields is well established. The hardware has finally caught up to the needs of conducting a complex real-life experiment. However, methods of analysis and data acquisition should also catch up in order to reach valid conclusions from real-world experimentation (Parada and Rossi, 2020). There are more and more computation methods being proposed and investigated in order to help the analysis to catch up with what the technology and real-life experimentation can offer (Griffiths, 2015). Meanwhile, all the challenges regarding real-life experimentation will be partially inherited by any method that tries to simulate real-life experimentation. However, this dissertation tries to argue that virtual reality experiments could be utilized to mitigated part of the complexity at the same time, offer and combine most of what real-life experimentation can offer such as correct understanding of our cognitive processes combined with the benefits of the well-designed strict lab-based experiments.

1.2 Virtual reality

The following sections are there to familiarize the reader with virtual reality's main technical aspects that will be discussed in more detail throughout this dissertation. In the following parts, we will briefly go over the history of virtual reality from concept to entering the consumer market and key technologies used in virtual reality hardware that enables seamless, immersive, and interactive virtual reality experiences. After which, we will shortly discuss the state of the art of virtual reality hardware and software implementation. In the final part, we will go back to the main argument of using virtual reality experiences as an environment to conduct research. By the end of this general introduction, the reader is expected to be familiarized with virtual reality and can follow the main line of argumentation and investigation in this work.

1.2.1 What is Virtual Reality

One of the broadest definitions of Virtual Reality that holds to this day is the one introduced by Frederick P. Brooks. In his paper, he defines a virtual reality experience as any experience in "which the user is effectively immersed in a responsive virtual world. This implies user dynamic control of viewpoint" (Brooks, 1999). A year earlier, Zheng Chan, and Gibson defined *virtual reality* as a form of computer-human interface that simulates a real-world, where the users are immersed in the computer-generated world and can freely move around, observe it from different angles, and even interact with the environment (Zheng et al., 1998). Although slightly different, one can observe that all attempts of defining virtual reality agree on immersion and some degree of freedom which should at least allow observing the environment from different angles.

Although with the emergence of new technologies such as augmented reality, their borders of definitions become blurry or even create new terms such as cross reality (XR). In terms of technology, augmented reality or AR extends our reality and therefore is in line with the embodiment movement. In essence, augmented reality tries to project information, in various methods and forms, onto our reality to enhance and augment it. Combining augmented reality techniques with computer-generated environments will lead to the field's ultimate goal: cross reality or XR, an attempt to perfectly blend virtual worlds and our real-world with projected augmented information. The main takeaway from all the possible definitions is that they all insist on "immersion". One can say the main difference between VR and its preceding technologies is "the sense of immediacy and control created by immersion: the feeling of **being there** or presence that comes from a changing visual display dependent on the head and eye movements" (Psotka, 1995). Therefore we can define *virtual reality* as any computer-generated virtual world that gives the user a sense of immersion and enables the users to interact with the environment in a naturalistic manner immediately.

1.2.2 History of Virtual Reality

Despite the current popularity of virtual reality experiences in the last decade, the history of such immersive experiences can stretch back as far as almost three centuries ago. The earliest example of an immersive experience is a panoramic painting by an Irish painter Robert Barker in Leicester Square in 1793(Berkman, 2018). The building itself, with its multiple floors and stairs, is an immersive experience that has been patented as such by Barker in 1796 (Figure 1.3). After which were similar techniques have been used by other artists to create immersive experiences, which a famous immersive experience Cinéorama by Grimoin-Sanson being exhibited in Paris in the year 1900 (Berkman, 2018). Cinéorama is an interesting case as the techniques used in Cinéorama are still used today in the entertainment industry. Cinéorama (1.4) featured ten synchronized videos projects on the walls of a circular room where the visitors were standing on a balloon-shaped platform simulating the experience of a balloon ride over the Paris(Berkman, 2018). These primitive early examples are far from modern-day virtual reality experience; however, they can be considered the earliest forms of virtual reality by immersion.

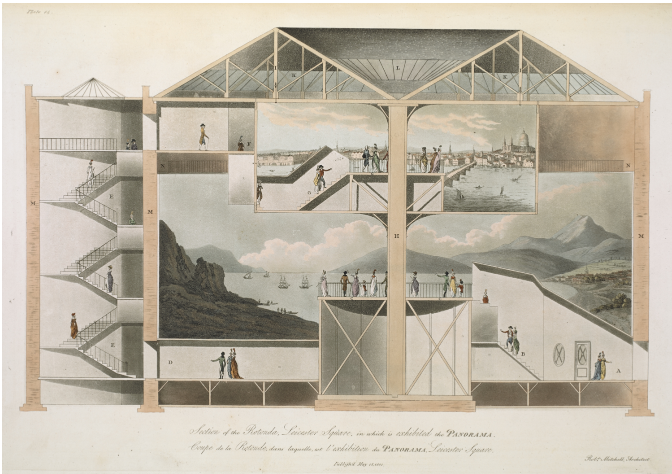

Figure 1.3: Cross-Section view of the building and paintings created by Robert Barker in 1973 and later patented by him in 1796. The Cross-Section shows a painting on a round wall stretching up for multiple floors as well as stairs leading to various sections of the building with different panoramic paintings simulating a change in the view.

London, 1801

Despite clearly successfully eliciting the sense of immersion, these techniques were still far from modern experience. The most important event which paved the way for modern-day virtual reality is undoubtedly the invention of stereoscope (Berkman, 2018). Sir Charles Wheatstone attempted to describe the essence of binocular vision and depth perception in his paper in 1838., which he called stereopsis (Wheatstone, 1838). Stereopsis, according to Wheatstone, is the fact that for any object with moderate distance from our eyes, each eye perceives a slightly different two-dimensional image of the object. This difference later enables our brain to perceive depth and three-dimensional information from observed scenes and objects. Sir Charles Wheatstone has used this knowledge to invent a

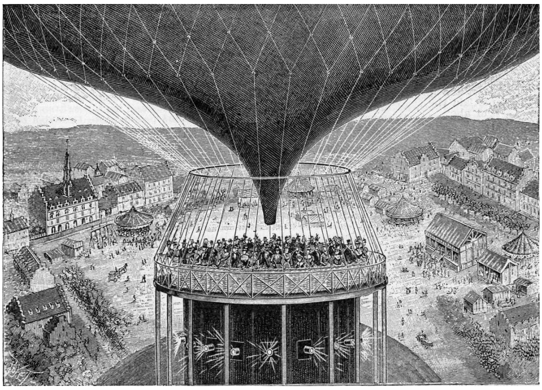

Figure 1.4: Illustration of Cinéorama. The illustration shows visitors standing on a platform in the shape of an air balloon where the cameras beneath the platform are projecting synchronized videos on circular walls surrounding the visitors.
Poyet, 1900

stereoscope in 1832 (Figure 1.5). Stereoscope uses the reverse of this phenomenon by presenting two slightly different two-dimensional images of the same scenery to each eye separately, creating a sense of depth for the observer(“Stereoscope”, 2021). Although it cannot be strictly called a virtual reality experience, it is the basic principle which all modern virtual reality devices function to create a three-dimensional environment for the user.

Between the invention of the stereoscope and the emergence of consumer-grade virtual reality devices, there were many examples of non-commercial and failed prototypes. One of the first modern-day immersive experiences was a device named Sensorama invented by cinematographer Morton Heilig in the 1950s (Virtual Reality Society, 2017). The device was an attempt to stimulated all the senses and featured a stereoscopic display as well as fans, a small generator, and a vibrating chair (1.6). Later the first virtual reality device, which was connected to a computer instead of a camera, was introduced by Ivan Sutherland and his student Bob Sproull in 1968 (Virtual Reality Society, 2017). Although the device was directly connected to a computer, the graphic processing power only allowed for illustration of wireframe object and device, which were too heavy to be worn by anyone. Therefore it was suspended from the ceiling, matching the name given

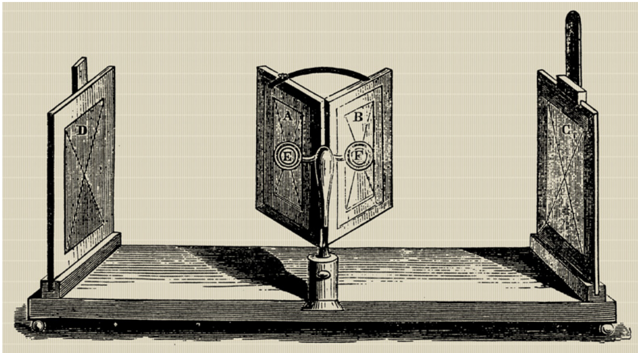

Figure 1.5: Illustration depicting the stereoscope invented by Sir Charles Wheatstone in 1832. The device utilizes two mirrors tow reflect sightly different image of the object into each of the observers eyes creating a sense of depth from two dimensional images.
Wikimedia, 2010

to it, Sword of Damocles (Virtual Reality Society, 2017). In 1986 Jaron Lanier and Thomas Zimmerman found the VPL Research (Virtual Reality Society, 2017) where Jaron Lanier coined the term virtual reality, which is commonly used today to refer to the technology ((Berkman, 2018)). Later in 1989, NASA started to use virtual reality to train astronauts in the project VIEW in partnership with VPL Research (Rosson, 2014). All these advancements slowly paved the path for today's consumer-grade VR, which was successfully introduced in 2014 by the startup company Oculus (Virtual Reality Society, 2017). Since then, with advancements in computer graphics, more companies have joined the industry offering affordable virtual reality experiences with devices that can satisfy various needs from entertainment to business to academia.

1.2.3 How does Virtual reality work

As mentioned above, stereopsis is one of the foundations of binocular depth perception (Wheatstone, 1838). When presented with a scene in real life, if far away from the eyes, the perspective lines connecting distant objects to our eyes are perfectly parallel (Wheatstone, 1838). However, in closer distances, as the eyes

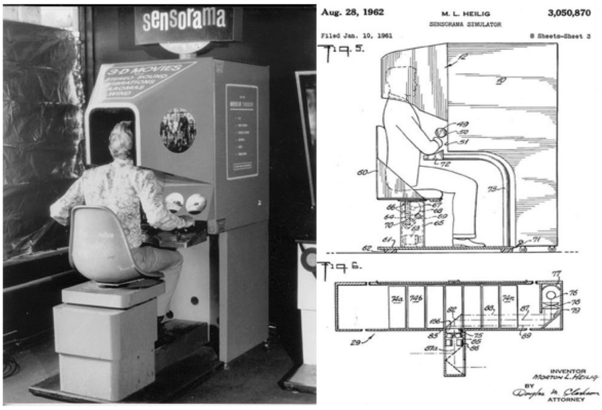

Figure 1.6: On the left an image of the device Sensorama in use. On the left an illustration of the device from cross side view by Morton Heiligg from the US3050870A patent filled by Morton Heilig in 1961. (Morton Leonard Heilig, 1962).

Parveau and Adda, 2020

converge, these lines are not on parallel planes anymore and will naturally start to converge on the fixated object (Wheatstone, 1838, Hess, 2019). This phenomenon leads to each eye receiving a slightly different image of the fixated object. Our brain is then able to pick up this slight disparity to create a sense of depth for the fixated object (Poggio and Poggio, 1984). Using two stereoscopic two-dimensional images and presenting them separately to each eye is the basis of how virtually all modern virtual reality headsets immerse their users in a three-dimensional environment.

To better understand how the virtual reality goggles work and their shortcomings, it is important to dive deeper into their structure. In order to access how well virtual reality goggles perform in simulating real-world binocular vision, we have first to explain how human binocular vision works in more detail. Here our main focus is on the physics of the light and eyes that leads to binocular vision rather than the neurological consequence of our eyes structure which leads to the unification of observed images from each eye and perception of depth. Due to having two eyes positioned with a small distance horizontally, the physics of light dictates that there will be a slightly different projection of the fixated object on the retina within the two eyes. (Harris, 2004). The geometry of binocular vision is undoubtedly complicated, especially since binocular vision can only happen in the overlapping area of both eyes' fields of view (cite). However, the projected image of the fixated object on the retina of each eye still can be mathematically calculated (Harris, 2004). Knowing the physics and the geometry of binocular disparity, there were many attempts and multiple ways for creating stereoscopic displays (Lipton, 2012).

In order to create the stereoscopic view, most virtual reality goggles use two independent LCDs, which render two different stereoscopic images near the user's eyes. A set of specific lenses then are utilized to project the images shown on the displays to a farther plane in user's view as depicted in figure 1.7(Reichelt et al., 2010,Jamali et al., 2018). However, effective in creating a stereoscopic image, simple placement of displays near eyes raises multiple issues, specifically accommodation-convergence in these displays (Reichelt et al., 2010). When fixating a near target from an afar target, accommodation is the reflex action of the eyes, where convergence is the physical inward movement of the eyes enabling single binocular vision (Jung, 2019). Currently, multiple solutions such as accommodation-invariant computation near-eye displays or adaptive focus, each with its benefits and downsides, have been proposed as a solution for the accommodation-convergence mismatch(Jamali et al., 2018). Despite these issues, modern virtual reality head-mounted displays can utilize the techniques mentioned earlier to render highly accurate and sharp three-dimensional imagery for the users.

What mentioned above, however, only explained how virtual reality experiences present a three-dimensional environment. However, for an experience to be considered immersive, one has to be able to also change the viewpoint by moving their head and realistically interact with the environment (Zheng et al., 1998). When it comes to interaction with the environment, there are multiple methods for handling users' interactions with the virtual environment. Here the main types of human-machine interfaces used for interaction with the virtual environment can be distinguished into three categories according to Mine (Mine, 1995).

- Direct user interaction: directly interacting with the virtual environment using hand tracking and gesture recognition (Mine, 1995 , Streppel et al., 2018).
- Physical controls: using physical objects to interact with the virtual environ-

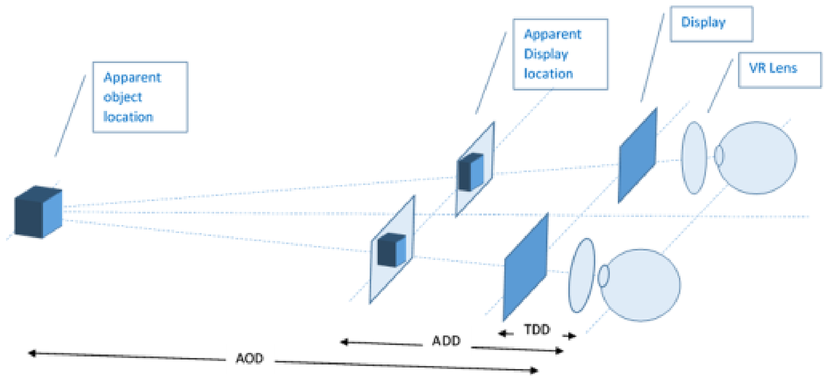

Figure 1.7: simple schematic of a typical virtual reality head mounted display where AOD indicates apparent object distance perceived by brain using the visual disparity cues of the images on the left and the right eye displays. TDD is the true display distance as contrast to the apparent display distance (ADD) achieved by using VR lenses.

Jamali et al., 2018

ment such as virtual reality controllers, steering wheels, etc... (Mine, 1995 , Streppel et al., 2018).

- Virtual controls: using virtual elements such as a virtual paint brush to interact with the virtual environment (Mine, 1995 , Streppel et al., 2018).

Currently, most virtual reality hardware uses physical controllers as the main method of interaction with the virtual world. However some manufacturers such as HTC and Oculus or with addition of extra hardware such as Leap Motion one can also use hand gestures and hand tracking for direct user interaction (Vive, 2021,Oculus, 2021,ultraleap, 2021).

In order to achieve naturalistic interaction with the virtual environment, all VR devices should be able to track essential parts of the body that matters for an immersive experience. This tracking mainly includes the tracking of the controllers or the hands based on the setup, as well as the head, which is considered a key technical component of the current VR and AR head mounted displays (Gourlay and Held, 2017). Currently, virtual reality devices use mainly two methods of tracking, namely Inside-out and Outside-in tracking (Gourlay and Held, 2017). The tracking offered by these devices is typically either 3 degrees of freedom, DoF, for tracking rotation or 6 DoF for additional positional information in three-dimensional space (Gourlay and Held, 2017). Using inside-out tracking sensors such as IMU and cameras embedded on the device is responsible for tracking the movement of the headset and controllers. The quality of the tracking is further improved if the structure of the tracking area is known (Gourlay and Held, 2017). Here the main differential factor between the two modes of tracking is that the position of the tracking equipment is outside of the headset itself (Gourlay and Held, 2017). In this scenario, one of the main methods of tracking introduced by Valve known as lighthouse tracking can be considered a hybrid approach (Figure 1.8). Non the less all methods above can offer a 6DoF tracking with a low latency of 6.71 $\pm 0.80ms$ as measured for HTC Vive (Caserman et al., 2019).

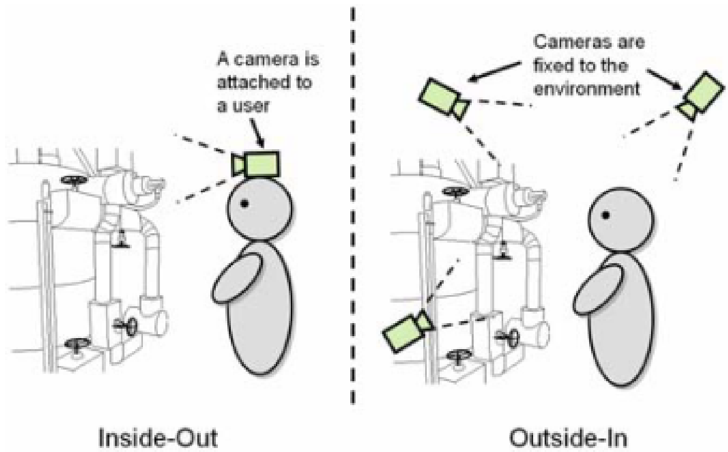

Figure 1.8: a simplified depiction of main ways of tracking in virtual reality. On the left is an example of an inside-out tracking system where the camera is attached to the head mounted display. On the right is an example of an outside-in tracking where multiple external cameras are setup and used to track user's head and bodily movements.

Ishii, 2010

As mentioned above, both methods of tracking, Inside-out and Outside-in, come with their own sets of benefits and drawbacks. In the lighthouse, tracking can be considered a form of inside-out hybrid tracking as the positional information is calculated on the head-mounted display itself (Gourlay and Held, 2017). In this method tracking the position of the HMD and subsequently the head is calculated based on the relative known position of the sensors on the device and the angular speed of infrared beams emitted by the lighthouses as horizontal and vertical sweeps (Gourlay and Held, 2017). General Outside in tracking has the benefit of precision since the tracking features, typically in light emitters, are built onto the virtual reality device itself. This method means that tracking is not dependant on environmental factors such as environmental illumination (Gourlay and Held, 2017). On the other hand, inside tracking does not require any external setup, and the device can move freely in the environment (Gourlay and Held, 2017). In conclusion, the choice of the appropriate tracking type should be decided based on the particular needs of the use case at hand.

Indeed, there is also a demand for improvement in hardware and application programming interfaces such as game engines for the virtual reality technology to be possible. As previously mentioned the process of tracking and changing the camera's viewport in virtual reality is a complicated mathematical task (Gourlay and Held, 2017, Caserman et al., 2019). In 1965 Gordon Moor stated that the number of transistors on a chip, and consequently computational power of a chip, will double in every technology generation, and this law seems to hold for now (Lundstrom, 2003). By 2008, graphic cards such as Nvidia Geforce® 8800 GTX were capable of processing 330 Giga floating-point operations per second (Owens et al., 2008). Modern graphic cards are built not only around the 3D graphics rendering pipelines which excel at the job of rendering computer-generated graphics but essentially have turned into a powerhouse that enable highly paralleled computation (Owens et al., 2008).

1.2.4 Virtual Reality, immersive yet controlled

From the definition of a virtual reality experience, one can see that being immersive and naturalistic interaction with the environment are the basis of a virtual reality experience. However, by definition, virtual reality is a computer-generated immersive experience (Zheng et al., 1998). In the real world, it is nearly impossible to control every aspect of the experimental environment and keep the environmental factors consistent between each experimental session (Gray, 2021). However, since a virtual reality experience is computer generated and controlled using programming languages, the experiment's designer controls every aspect of this immersive environment. For instance, the timing of events, a similar experience between each experiment, the world's physics, and interaction. Therefore, such an experience is explicitly controlled despite the realism or immersion.

Although this control might not be necessary for other use cases of virtual reality experiences such as entertainment, it is essential when it concerns the validity of neuroscientific experiments (Parada and Rossi, 2020). After all, one of the main reasons that most neuroscientific experiments are done inside the laboratory environment is the ability to control the environment. Keeping the environment consistent between different subjects reduces the chance of encountering random effects due to unaccounted factors and helps with the replicability of the study (Gray, 2021). Therefore, being able to create realistic and naturalistic environments and interactions and still insuring the ecological validity of the experiment means that we can get closer to investigating neuroscientific phenomenon in a closer to reality environment (Shamay-Tsoory and Mendelsohn, 2019,Holleman et al., 2020). Therefore irtual reality environments being immersive yet controlled can improve the validity of the experiments by enabling closer to real-life experiments (Fan et al., 2021).

At the beginning of this general introduction, we discussed the laboratory's ecological validity versus real-life neuroscientific experiments. Furthermore, we mentioned that our cognition is embodied, embedded, extended, and enacted (Newen et al., 2018). Meaning, it is evident that we cannot study the human brain and cognition, especially behaviors that are due to social interaction with others, interacting with the environment in the oversimplified laboratory environment. (Shamay-Tsoory and Mendelsohn, 2019). It is also important to note that certain experiments are, in essence, either impossible or immensely difficult to perform in either laboratory or the real world. As an example of such experiments, one can mention the trolley dilemma problem or body modification/perception. Consequently, an immersive virtual reality experience can help us study cognitive behaviors in a closer to reality setup and allow experimentation and performing measurements previously nearly impossible to perform (Kilteni et al., 2012; Tarr and Warren, 2002).

1.3 Short comings of virtual reality environments

We have previously touched on the most important benefits of using virtual reality environments; however, it is vital that we also discuss possible downsides and difficulties associated with creating and performing virtual reality-based experiments. One of the main difficulties in creating immersive virtual reality experiments is the level of programming skills required to create such environments (Nezami et al., 2020). Here the expertise is not only to create the task, the environment, care for proper serialization of data, Etc. Sine they all, directly affect the performance of the virtual reality experiments. As mentioned before, despite the current advancements in computation and graphics processing, virtual reality experiments are heavily demanding on the computational side. Without proper optimizations in the programming of the virtual reality environment, the frame rate and performance of the experiment can be heavily affected even on the best of today's hardware. The optimization of frame rate is also of high importance, considering the critical role frame rate plays in immersion and motion sickness induced by virtual reality hardware. After all, if the immersion is broken or the participant cannot continue the experiment due to severe motion sickness, the experiment itself or the sense of being closer to real-life is very negatively affected.

The subject of performance in virtual reality-based experiments is a serious matter to consider. Most virtual reality experiments, including those presented in this dissertation, are made with traditional Game engines such as the Unity3D game engine. However, game engines are optimized to prioritize visual fidelity over precision and persistence. This optimization means that if the game engine's longer processing time predicts a visual lag in the output rendered environment, it might discard a particular frame. Secondly, most game engines, including Unity3D does not provide a constant frame rate. There are ways around this issue, such as using the physics loop, which ensures a constant frame rate; however, the programmer of the environment has minimal control over how the graphic process will process each frame for rendering. Lastly, there is an inherent delay between the moment the processor executes a line of code until the virtual reality headset can render a frame on its displays. Scientific experiments, specifically those of psychology or neuroscience, are typically concerned with time-sensitive events such as various visual or auditory stimuli onset, reaction times, and eye gaze information of brain activity that requires precise synchronization with such events. Consequently, one should invest more time than usual optimizing the virtual reality experiment to ensure correct recording of data and presentation of the experiment. Perhaps one should develop an engine specific to scientific experiments that can address the issues mentioned above.

The other important consideration when deciding to go with a virtual reality experiment is the state of the art of measuring hardware. Currently, there are a variety of eye trackers being implemented within the most common virtual reality head-mounted displays, with Tobii and Pupils lab being the most noteworthy and commonly used eye trackers for virtual reality. However, 3D eye tracking in VR is relatively new compared to the traditional 2D eye trackers such as eye link. Not only the sample rate in these devices are considerably lower (around 120Hz), these eye trackers also have to overcome issues such as eye accommodation-convergence of head-mounted displays. Similarly, attempts for combining neuroimaging techniques such as EEG in virtual reality experiments are also recent and need more time to mature (Tauscher et al., 2019). In conclusion, despite the massive benefits of moving toward virtual reality experiments as closer to reality substitution of the Lab experiments, one has to consider the shortcomings that can directly affect the outcome of the experiments.

1.3.1 Chapters' Overview

The main goal of this thesis is to investigate Virtual Reality as a tool to improve the ecological validity of neuroscientific and psychological behavioral experiments. It tries to investigate both benefits and shortcomings of immersive virtual reality experiments and gives insight into how one can develop realistic yet high performance virtual reality experiments. Other researchers have already proved the validity of immersive virtual reality experiences in various studies and disciplines, including body perception and ownership (Peck et al., 2013; Slater et al., 2009; Slater et al., 2010), treatment of psychological disorders (North et al., 2002), trolley problem (J.-F. Bonnefon et al., 2015; Kallioinen et al., 2019), and many other psychological or neuroscientific experiments. However, as mentioned before, immersive virtual reality technologies are relatively new. There is new advancement on the software and the hardware, including but not limited to computational power and the virtual reality head mounted display itself, including various interaction methods, has the potential to enable researchers to bring their experiment even closer than ever to a genuine real-life experience.

In this thesis, the first part concerns the techniques and practices of building virtual reality experiments. In the first chapter, we conduct experiments related to self-driving cars in two relatively large virtual reality environments. However, to not limit the environments to just one purpose, we have developed the environment as generic areas so that other researchers can use them for experimentation toward a gradient of subjects such as spatial navigation or social interactions. These studies also try to reflect the best practices and concerns when designing immersive and realistic virtual reality experiments. Afterward, we briefly investigate the state of the most common eye-tracking hardware available for virtual reality experiments and try to solve the shortcomings of the current common development platform of virtual reality experiments, namely Unity3D game engine, to perform a real-time multi-participant social interaction or joint action studies. Therefore the first half of this dissertation serves as an overview of the methodical matters of developing an immersive virtual reality experiment.

In the second part, this work will focus on the scientific outcome of virtual reality experiments. In the second chapter, we investigate task planning using virtual reality experiments. The studies in chapter three are regarding immediate planning in sorting tasks or gaze pattern when interacting with familiar or unfamiliar tools. While the first study discovers the just-in-time nature of planning in a sorting task, it also shows some clear evidence of the effect of the need to naturally interact with an actual size shelf in the sorting patterns of participants: the second study and a replication of an in-lab experiment confirming the same results. However, indicate that different effects were not visible in an interaction using virtual hands instead of the controller. Chapter three returns to the premise of using virtual reality experiments to perform experiments impossible in real life, investigating the trolley dilemma and acceptance of self-driving cars. Here the experiments are closest to the big picture of this thesis that a well-crafted immersive virtual reality experiment can be used to elicit and subsequently measure desired behaviors.

2 Virtual Environments and techniques

2.1 Layman's summary

How can one design a scientifically valid experiment using virtual reality and virtual environments. Virtual reality experiences are computer-generated simulations. Here it is important to emphasize that simulated is a synonym to programmed. For a virtual reality experiment to be valid, in other words for it to work and be understood by the participant precisely as the experimenter intended it to, there are many challenges to overcome. First of all, such simulations, namely virtual reality environments, are not just a simple programming problem. The programmer needs to have a deep understanding of the currently available programming techniques, any game engine or programming language that supports VR, computer graphics , which evolves around 3d geometry, and even perhaps fundamentals of physics. Every asset, every 3d model, needs to be purchased and optimized or made out of scratch. Every behaviour should be programmed in advance. These are just the part of challenges of doing any virtual reality experiment.

In light of the mentioned complexity of creating virtual environments the Westdrive and Westdrive LoopAR environments were built. Initially, these projects were meant to be an environment to investigate trust in self-driving cars and the role of audiovisual warning in reducing the driver's reaction time in a critical take-over request. However, we soon realized the gap and the need for realistic virtual environments that one can modify with little programming knowledge. These environments offer generic, large realistic environments containing the most common driving environment, such as mountain roads or country roads. However, one can also utilize these environments for other research, such as face perception or spatial navigation. These environments offer dynamic scenery with cars and pedestrians with naturalistic behavior. But the base environments can be used in any experiments that need a naturalistic outdoor environment.

Nevertheless, projects Westdrive and Westdrive LoopAR are not only simulated environments. They offered a great learning experience throughout their development process as to the possible limitations and challenges that need to be addressed while creating virtual reality experiments. From synchronization of different data, acceptable frame rate to overcome motion sickness induced by wearing virtual reality headsets, to creating stable environments that provide laboratory-grade controlled experiments yet can be run without any interference gathering behavioral data for months from thousands of people. These projects offered an understanding of the current state of the art and what is possible for future research using virtual reality environments.

At the same time, creating these large-scale environments clarified the shortcomings of the current state-of-the-art hardware and software for virtual reality experimentation. The lack of such computer networking platforms that is reliable enough which can be used for simultaneous study of more than one participant in the same environment is an example of one of the major shortcomings. The variety of offered measurement hardware is also another concern that is of importance for scientific research. One of the most common integrated measurement devices in the modern virtual reality glasses are eye trackers. However, the specifics of their performance are not transparent or clear for researchers. That is why we tried to compare two different integrated eye trackers on the market and develop a lightweight yet practical framework for multi-participant experimentation in virtual reality.

2.2 Project Westdrive: Unity city with self-driving cars and pedestrians for virtual reality studies

This section was submitted as a peer reviewed paper in the Frontiers in computer science together with Maximilian Alexander Wächter, Gordon Pipa, and Peter König. See Publication List for details.

2.2.1 Abstract

Virtual environments will deeply alter the way we conduct scientific studies on human behavior. Possible applications range from spatial navigation over addressing moral dilemmas in a more natural manner to therapeutic applications for affective disorders. The decisive factor for this broad range of applications is that virtual reality (VR) is able to combine a well-controlled experimental environment together with the ecological validity of the immersion of test subjects. Until now, however, programming such an environment in Unity requires profound knowledge of C# programming, 3D design, and computer graphics. In order to give interested research groups access to a realistic VR environment which can easily adapt to the varying needs of experiments, we developed a large, open source, scriptable, and modular VR city. It covers an area of 230 hectare, up to 150 self-driving vehicles and 655 active and passive pedestrians and thousands of nature assets to make it both highly dynamic and realistic. Furthermore, the repository presented here contains a stand-alone City AI toolkit for creating avatars and customizing cars. Finally, the package contains code to easily set up VR studies. All main functions are integrated into the graphical user interface of the Unity Editor to ease the use of the embedded functionalities. In summary, the project named Westdrive is developed to enable research groups to access a state-of-the-art VR environment that is easily adapted to specific needs and allows focus on the respective research question.

2.2.2 Introduction

With the opening of the consumer market in recent years, VR has penetrated many areas of everyday life: there are e.g., applications for marketing, the games industry and for educational purposes (Anthes et al., 2016; Burke, 2018; A. Miller, 2018). Research on human behavior is also beginning to take an interest in experiments in virtual reality (de la Rosa & Breidt, 2018; Rus-Calafell et al., 2018; Wienrich et al., 2018). For instance, it is possible to embed ethical decision making in a seemingly real context in order to achieve a higher validity of experiments (Faulhaber et al., 2019; Sütfeld et al., 2017). Further, studies based on VR techniques address questions regarding spatial navigation, such as neurological correlations of human navigation (Epstein et al., 2017), as well as gender differences in navigation tasks in a well-controlled environment (Castelli et al., 2008). Although there are already available tools for creating virtual cities, these applications have not yet been designed for experiments on human behavior, but rather for planning and simulating urban development (Botica et al., 2015; CityEngine, 2013; Dong et al., 2019; "VR Design Studio | FORUM8 | 3D VR & Visual Interactive Simulation", n.d.). Furthermore, it is possible to use VR in a variety of psychotherapeutic and clinical scenarios (A. Li et al., 2011; Riva, 2005). Not only is this cost-efficient and more interactive than classical psychotherapy (Bashiri et al., 2017), it also offers the possibility to use this treatment at home, as VR becomes more widespread in the future. This means that VR has the potential to increase access to insights of human behavior as well as to psychological interventions (Freeman et al., 2018; Slater & Wilbur, 1997). Finally, VR can be combined with further technologies, such as EEG (Bischof & Boulanger, 2003)and fMRI, facilitating research of clinical disorders (Reggente et al., 2018). In summary, VR techniques have the potential to heavily advance research in the human sciences.

Still, compared to classical screen experiments, VR-based experiments are complex and require extensive programming, which is an intricate task by itself (Freeman et al., 2018). This causes VR experiments in behavioral research to lag behind their actual potential (Faisal, 2017). Even if already existing experiments are transferred to VR, knowledge of software and hardware must be acquired, meaning a larger expenditure of time and content (Pan & Hamilton, 2018). Westdrive is developed to eliminate these obstacles in the context of studies on spatial navigation and ethical aspects. It shortens the time required for the setup of or the transfer to VR experiments by a considerable magnitude either by enabling researchers to use the project scene directly, or indirectly by letting them use only the provided assets and code.

2.2.3 Results

2.2.4 Key Features

Probably the most crucial features of Westdrive are size, modularity and the simple handling of complex environments, since all components of the City AI toolkit can be used independently even without any programming knowledge.

Size is often a critical factor for virtual environments. This is the case with e.g., navigation tasks within VR (S. U. König et al., 2019). A distinction is made here between room-sized vista space and large environmental space. Small rooms are easier to grasp and therefore it is only possible in large environments to distinguish between test subjects who navigate using snapshots of landmarks only and those who have learned a true map of their environment (Ekstrom & Isham, 2017).

The modularity of a virtual environment is of equal importance. Not only does building realistic cities require the consideration of many different aspects, but different research projects also depend upon distinctive dynamic objects. For example, an experiment on the trolley dilemma requires driving vehicles and pedestrians (Faulhaber et al., 2019). A therapeutic application for fear of heights requires high buildings and animated characters to make the environment appear real (Freeman et al., 2018). Project Westdrive offers a wide variety of applications due to its modularity, both of the static environment which comprises trees, pavements, buildings, etc. and the dynamic objects like pedestrians or self-driving cars.

Additionally, the aforementioned managers of the City AI toolkit enable a simple handling of the project. The City AI toolkit, which facilitates implementation of paths, pedestrians and cars, which are all usable within the Unity GUI without any experience in coding. All of the components are accessible within the Unity Editor. All managers can be edited separately according to the respective requirements of an experiment. In this sense, these separately adjustable components also support modularity as only adjustments for the necessary components have to be made.

To use the project, only a powerful computer, VR glasses and the free Unity program are needed1. If the aforementioned requirements are met, the scene presented here can be changed or manipulated at will. It is also possible to make alterations exclusively in the GUI of the Unity editor without writing any code. This project offers not only the templates for static models, but also the functions integrated into the GUI for paths, character creation, and the creation of moving cars.

Westdrive and the City AI have been created with having simplicity in mind to relieve users from as much time-consuming preparations and programming as possible. Yet, as an open source project under constant development, we also encourage future researchers to further improve the project or change the codes based on their specific needs. Westdrive gives the user the possibility to carry out a multitude of investigations on human behavior through the key features. For example, the simple routing of pedestrians and cars makes it possible to carry out studies on trolley dilemmas or the human-machine interaction. Also, due to the realism of the avatars (Fig. 2.1) it is possible to build therapeutic applications for the treatment of fear of heights or social phobias. However, this is only a very small part of the possible applications.

Figure 2.1: Overview of all used Fuse CC Avatars in the virtual city

2.2.5 Methods

Project Structure

The Westdrive virtual environment is built in Unity 2018.3.0f2 (64 bit), a game engine platform by Unity Technologies. This engine is used together with a graphical user interface (GUI) called the Unity editor, which supports 2D and 3D graphics as well as scripting in JavaScript and C# to create dynamic objects inside a simulation. Unity runs on Windows and Mac and a Unity-built project can be run on almost all common platforms including mobile devices like tablets or smartphones. We have chosen this software due to many available application programming interfaces (APIs) and good compatibility with a variety of VR headsets (Juliani et al., 2020). Moreover, the use of Unity grants access to an asset store, which offers the option to purchase prefabricated 3D objects or scripts which only need to be imported into an already existing scene. Thus, Westdrive is a modular virtual environment, making it easy to integrate other software now and in the future.

The Westdrive repository contains a city as one completed game scene. All associated assets including driving cars, walking characters, buildings, trees, plants, and a multitude of smaller 3D objects such as lanterns, traffic lights, benches etc. are included and offer a high level of detail (Fig. 2.2). It also contains the relevant code that executes interactions and animations of the mentioned objects. Thus, users have all desirable components for an experiment in one consistent package. Westdrive can be divided into two sub-areas. On the one hand there is the static environment and on the other hand there is the code for interactions between dynamic objects. Both will be explained in the following.

Static Environment

The static environment models a large urban area. It includes 93 houses, several kilometers of roads and footpaths, about 10,000 small objects and about 16,000 trees, and plants on a total area of about 230 hectares. A large part of the 3D objects used for this purpose are taken from the Unity asset store for free. A list of used assets and their licenses can be found in the specified repository. However, the design of the city presented here can be varied at will in the editor and an

Figure 2.2: Overview of the level of detail in the simulated city of the project Westdrive in a completed scene.

included mesh separating tool. It is possible to change the size, shape and amount of individual buildings, streets, cars, and pedestrians in the GUI of the Unity Editor. The same applies to all other assets presented here. The static environment alone can thus be used indirectly for the development of further VR simulations as the project provides a large number of prefabricated assets (prefabs) that do not have to be created again. Consequently, it is possible to easily develop a broad range of scenarios for realistic VR experiments by simply manipulating the static environment to match respective needs.

Scripting Dynamic Objects

To use Westdrive to its full extent, the code described here is of essential importance. The code is written entirely in C# based on Microsoft's Net 4.0 API level and envelopes all functions for the stand-alone City AI toolkit (Fig. 2.5). This includes six components developed by us: a Path Manager to create and manipulate paths for pedestrians and cars, a Car Engine script that allows cars to move independently, and a Car Profile Manager to create different profiles for different cars (e.g., the distance maintained to other vehicles, engine sound and car color). Additionally, there is the Pedestrian Manager and the Character Manager, that control animations and spawn points for moving characters along their defined route and an Experiment Profile Manager, which defines the experimental context, like routes, audio files, and scripted events along the path. The City AI works as a stand-alone toolkit in the GUI of the Unity editor. In short, it is possible to define fixed routes with spawn points for pedestrians and cars along which the non-playable characters (NPCs), such as pedestrians and cars, will move. Only if visual change of characters is desired an external tool is necessary2.

Figure 2.3: Impressions of cars in the highly realistic city scene.

To enable well-controlled movements of cars and pedestrians, we developed a path creation toolkit inside the City AI which incorporates mathematical components of Bezier Splines (Prautzsch et al., 2002). This results in a deterministic and accurate path following system which is only dependent on units of time in a non-physics-based simulation. The users can themselves change the control points of the path inside the editor (see Fig. 2.4). It is also possible to define the duration of the route or the circuit in the Unity editor. Furthermore, the kinematic path creation facilitates the creation of spawn points for different asset types (cars and pedestrians at the moment, see Fig. 2.3, 2.1) for each path. All these functions work without programming knowledge. The following components of the City AI are also depicted in Fig. 2.4 to give a better overview of interactions and possibilities within Westdrive:

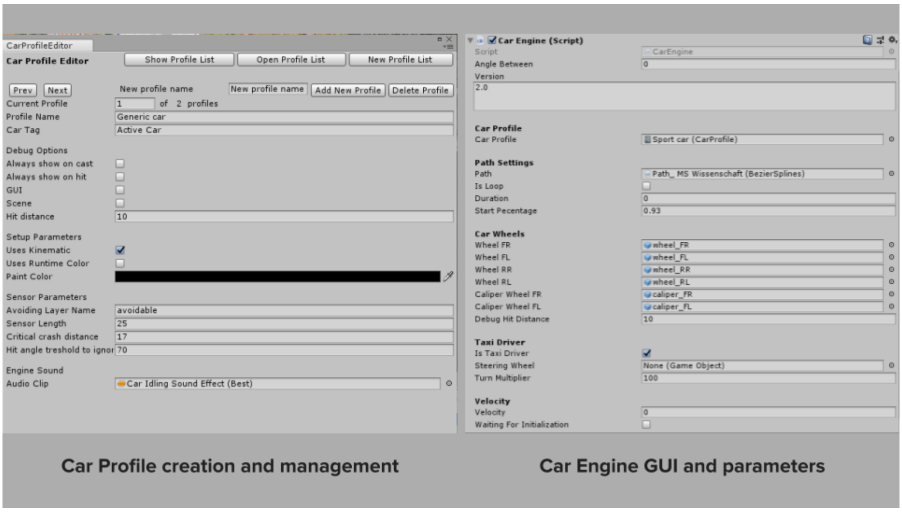

Figure 2.4: Overview in the Editor of the Car Profile Manager, the Car Engine and the according parameter bar. These functions allow users to use different types of vehicles in the city. The Car Profile changes the appearance of the vehicles, such as color, engine noise and sensor length. Car Engine allows the vehicles to move independently on the defined routes through the city and to accelerate, brake and steer independently. For each of these functions defaults are provided. An adjustment of these parameters is therefore only necessary for new vehicles.

Path Manager: This is the basis for all moving objects in Westdrive. With just a few clicks in the editor, the user can create new routes for pedestrians and cars or change existing routes. To do so, the control points of the already mentioned Splines can be moved using the mouse only. Afterwards it is possible to set the speed for objects on this route.

Car Engine: This component enables vehicles to steer, brake and accelerate independently both at traffic lights and in the event of an imminent collision with other road users.

Car Profile Manager: This component allows users to create and manage multiple independent profiles for cars. It enables creation of various types of cars with different characteristics such as engine sound, color, or a different spacer for preceding vehicles.

Pedestrian and Car Manager System: These systems take care of automatic spawn, restart, and re-spawn of all pedestrians and cars in the scene. They have the ability to load resources in an either synchronous or asynchronous manner, to ensure a smooth-running experiment.

Experiment Profiles and Procedure Controller: These scripts enable users to create different experiments within the environment. These profiles set up parameters for e.g., the routes that cars will follow. They also trigger the beginning and the end of the experiment; the end of the experiment blocks and they disable dynamic objects not necessary in the scene if needed. The Procedure Controller uses the Experiment Profile to automatize the experimental procedure e.g., by ending blocks, altering the appearance of or completely excluding dynamic objects.

All of these managers assign the correct scripts to objects and move them to a resources folder in order for them to be spawned in runtime when the experiment starts. These toolkits ensure that cars and pedestrians have all the necessary components attached to them.

Implementation

As head-mounted display (HMD), the HTC Vive Pro is used at our department. At the time of writing, this virtual reality device is the most advanced technology available (Ogdon, 2019). In order to transfer the player’s head movements into the virtual reality, HTC utilizes two passive laser-emitting “lighthouses” that have to be attached to the ceiling in two opposing corners of the room. The two handheld controllers and the headset use no <70 combined sensors to calibrate the positions of controllers and headset, measuring the time difference in sending and receiving the emitted signal (Ahmad, 2020). To use the HTC Vive Pro and the HTC Setup Software, an account at the online gaming platform Steam is necessary. This requires a stable internet connection, as both Steam and the HTC Setup software are free to use. Since this device is one of the most expensive ones on the market, it is used mainly for academic or industrial research rather than private gaming.

It is also worth mentioning, that although Westdrive has been developed for the HTC Vive Pro, it can easily be transferred to other virtual reality HMDs. The last component for the implementation is the Unity software. Unity can also be used free of charge as long as a project is not used commercially. Licenses are free for students and researchers. The Unity editor can be downloaded from the Unity website. Now it is possible to create a project order and convert the files from the repository presented here into Unity.

A more detailed description of how to set up Westdrive as well as an example of the functionalities can be found as tutorial videos in the repository and in the Supplementary Materials.

2.2.6 Discussion

Current Limitations

Due to the complexity of the project and the differences between a deterministic simulation and a computer game, there are still many possible improvements to be implemented. With current enhancements like occlusion culling where, objects are not rendered when they are not seen by the player, simplified shadows, and mesh combining, an acceptable frame rate of at least 30 fps can be achieved using an NVidia GeForce RTX 2080 TI in combination with an Intel(R) Xenon E5-1607 v4. The desired goal in the course of further research will be to reach the stable 90 Hz suggested by virtual reality technology providers such as HTC and Oculus.

It is important to note that the code does not calculate the mentioned objects physically, but kinematically, so no physically simulated forces are applied to any moving objects. There are several reasons for this: on the one hand, the computational requirements of the computer on which Westdrive is used on are kept as low as possible. On the other hand, an exact control bar of the visual stimuli can be guaranteed, because each object is spatially located exactly at the same place at the same time. Furthermore, it makes potential directed changes easy, as no physical interactions have to be reverse engineered.

Another point is that there is currently no structured software architecture. So far, the priority has been on the simple handling of all functionalities within the editor to facilitate the creation of own experiments. A structured architecture is still under development.

Outlook

Concluding, we again want to emphasize the impact Westdrive can have on future VR research. Already over a decade ago, the potential of combining VR with physiological measurements has been discussed (Bischof and Boulanger, 2003), but only in the past years, when software became affordable, there was a renewed interest in VR in science (Interrante et al., 2018). The main advantage of the project is a simple implementation of a versatile project which, despite its complexity, can be altered quickly and easily without programming knowledge. Likewise, the experiment in its basic form doubles as an eye-tracking study. The code for the implementation is not included in this version, mainly because it was not written by the two authors, but by the Seahaven research group, investigating spatial navigation in a virtual environment (König et al., 2019). However, the repository will be constantly updated, thus it will also contain the required eye tracking code for Pupil Labs in the future. Westdrive as a city environment offers many areas of application. Nevertheless, the project is constantly in development and extension. At least two more scenes are currently planned in order to allow for an even wider application, for example the investigation of trolley dilemmas (Thomson, 1984) using a railway track or possible applications of the acceptance of new mobility concepts. All improvements and added scenes will be released via GitLab. Additionally, we are going to further clear up old parts of code and unused assets as code janitor, as well as fixing any possible typo or mistake in the code. At the same time, we will expand the comments and wiki section to have a user guide on how to use the project.

Since we are constantly improving the code and add functionalities, this cleanup is an ongoing process.

In this work, particular importance was attributed to a comprehensible formulation in order to ensure an understandable documentation of the work performed. There is an almost unlimited number of application possibilities for the extension of this project. The authors are looking forward to the many great ideas for the continuation of Westdrive.

Author Contribution

MW and FN wrote this paper. Initial Idea to Westdrive began as a joint Master thesis. Both authors were building and designing the project. PK and GP supervised the project.

Funding

We acknowledge support by Deutsche Forschungsgemeinschaft (DFG) and Open Access Publishing Fund of Osnabrück University. This contribution is part of the research training group “va-eva: Vertrauen und Akzeptanz in erweiterten und virtuellen Arbeitswelten” of the University of Osnabrück.

2.2.7 Supplementary Material

The Supplementary Material can be found online at: https://www.frontiersin.org/articles/10.3389/fict.2020.00001/full#supplementary-material. Unity 3D learning: www.unity.com/learn. Online Character animation: www.mixamo.com.

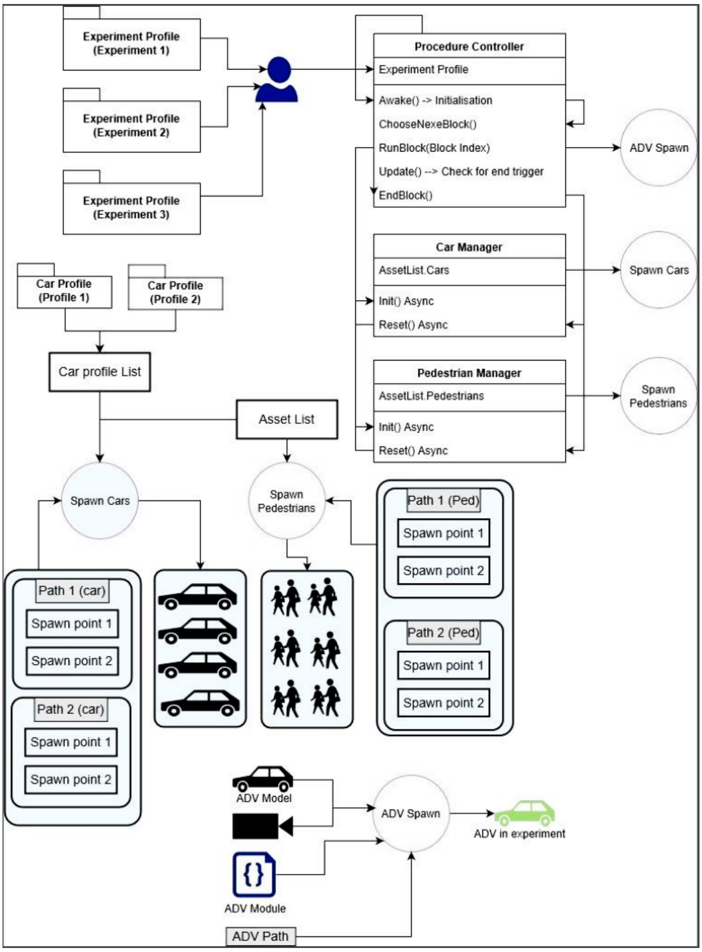

Figure 2.5: Scheme of the City AI features in Westdrive. This illustrates the interaction of the different managers of the toolkit to enable spawned cars and pedestrians as well as different experimental setups saved in one scene. These experimental profiles trigger the procedure controller, which takes care of the onset and ending of the experiment and creates the subject's car or avatar. This also triggers the car and pedestrian manager, which are responsible for the spawning of passive cars and pedestrians. In combination with the Car Profiles and the Asset List, the various cars and pedestrians required for the experiment are created in the experiment.

2.3 Westdrive X LoopAR: An Open-Access Virtual Reality Project in Unity for Evaluating User Interaction Methods during Takeover Requests

This section was submitted as a peer reviewed paper in MDPI Sensors together with Maximilian A Wächter, Nora Maleki, Philipp Spaniol, Lea M Kühne, Anke Haas, Johannes M Pingel, Linus Tiemann, Frederik Nienhaus, Lynn Keller, Sabine U König, Peter König, Gordon Pipa. See Publication List for details.

2.3.1 Abstract

With the further development of highly automated vehicles, drivers will engage in non-related tasks while being driven. Still, drivers have to take over control when requested by the car. Here, the question arises, how potentially distracted drivers get back into the control-loop quickly and safely when the car requests a takeover. To investigate effective human–machine interactions, a mobile, versatile, and cost-efficient setup is needed. Here, we describe a virtual reality toolkit for the Unity 3D game engine containing all the necessary code and assets to enable fast adaptations to various human–machine interaction experiments, including closely monitoring the subject. The presented project contains all the needed functionalities for realistic traffic behavior, cars, pedestrians, and a large, open-source, scriptable, and modular VR environment. It covers roughly 25 km2, a package of 125 animated pedestrians, and numerous vehicles, including motorbikes, trucks, and cars. It also contains all the needed nature assets to make it both highly dynamic and realistic. The presented repository contains a C++ library made for LoopAR that enables force feedback for gaming steering wheels as a fully supported component. It also includes all necessary scripts for eye-tracking in the used devices. All the main functions are integrated into the graphical user interface of the Unity editor or are available as prefab variants to ease the use of the embedded functionalities. This project's primary purpose is to serve as an open-access, cost-efficient toolkit that enables interested researchers to conduct realistic virtual reality research studies without costly and immobile simulators. To ensure the accessibility and usability of the mentioned toolkit, we performed a user experience report, also included in this paper.

2.3.2 Introduction

What defines the user-friendly design of automated systems has been the subject of scientific discussion for decades (Bengler et al., 2020; Norman, 1990). Especially in the upcoming years, when automated vehicles of SAE (society of automotive engineers) automation levels 3 and 4 will emerge, the demands on the driver's cognitive system will alter radically, as the role of humans as continuously active decision-makers in vehicles is replaced by automated systems (S. Li, Blythe, et al., 2019; Lindgren et al., 2020). Such techniques include the Audi traffic jam pilot (Audi, 2017) or Tesla's full self-driving beta (Tesla, 2020). Airlines' experiences, where automated systems are already widely integrated, clearly state that such systems' safety and reliability cannot be achieved by optimizing technical components alone (Masalonis et al., 1999). Instead, the reliability of highly automated systems is primarily determined by the driver's cognitive processes, meaning how fast a safe transition to manual drive is possible (Zeeb et al., 2015) .

The need for a fast and safe transition applies particularly to situations where humans have the task of taking over system control in the event of sensor failures or malfunctions (Abe et al., 2011; Maurer, 2015). Thus, investigating the fluent integration of the takeover request (ToR) is crucial for the safety of any system with even partially automated driving features (Marberger et al., 2018). During a takeover request, the human driver most likely has to take over control in under 10 s, even when not engaged in driving-related activities (Dogan et al., 2019; C. Gold et al., 2013; Melcher et al., 2015). Naturally, an orientation phase follows as the human driver has to assess the traffic situation (C. Gold et al., 2013). Unfortunately, the driver's reaction is often too slow in critical situations, potentially resulting in an accident in the small time frame (<4 s) before an impact occurs (Green, 2000; Summala, 2000). Even in the case of fast reactions within a time frame under 10 s, studies with prolonged driving have shown hectic responses by human drivers, which of course neither improved the reaction time nor the situational outcome (Endsley & Kiris, 1995; Jarosch et al., 2019).

This manuscript presents a new toolset for human–machine interaction research apart from typical screen-based simulators. Existing simulators are often based on actual car interior designs. Therefore, they offer only limited possibilities for human–machine interaction (HMI) research (Morra et al., 2019). A very similar problem is posed by research on prototype cars in the real world, where realistic accident scenarios are costly and can only be generated to a minimal extent without endangering the test person involved. The project, called LoopAR, provides not only all the needed assets and an environment but also all the needed code to display the information of a takeover request as a freely programmable augmented reality (AR) feature in the windshield. The developed HMI displays the takeover request and highlights critical traffic objects to enable participants to take over more quickly and precisely. Our research is aimed toward safe and effective communication between car and driver. This is not only beneficial in terms of safety for the passengers but could also increase customer acceptance of highly automated vehicles, since up until now, malfunctions have been vital concerns of possible customers (Howard & Dai, 2014a). Since LoopAR is based on the project Westdrive (Nezami et al., 2020), all the code needed and designed scenes are available in a Github repository. Project Westdrive is an open science VR project that tries to enable many researchers to conduct VR studies. It provides all the necessary code and assets in a public repository to set up VR studies. LoopAR is an extension of the Westdrive toolkit, focusing on the human–machine interaction. To fully use the project presented here, only a powerful computer, VR glasses, a simulation steering wheel and pedals, as well as Unity as a development program are required.

2.3.3 Methods and Main Features

The main focus of the presented project is versatility and modularity, which allows the fast adjustment of the environmental and functional objects via prefab and the provided code in the toolkit. Research on the interactions between humans and cars is mostly done with stationary simulators. Here, a whole car chassis is used, or only the interior is set inside a multi-screen setup. However, these classical setups are often expensive, and adjustments or graphical improvements of the stimuli used in an experiment are often not possible (Cruden, n.d.). In the past few years, there has been a significant shift in research toward virtual environments. This is reflected by applications like Cityengine and FUZOR (CityEngine, 2013; Kallotech, n.d.) and by the software for driving environments (Dosovitskiy et al., 2017). Still, experimental designs on human–machine interaction, in terms of specific car interior adjustments, are not possible yet. Therefore, the presented project enables the user to create experimental conditions and stimuli freely. All functionalities that are mentioned in the following are independent and can be adjusted at will. Additionally, the presented project does not need a specific hardware setup, making it easily adaptable and future-proof. New components, e.g., new GPUs and new VR devices, can be easily integrated into the setup displayed in Figure 2.6. The current requirements only apply to the VR devices used and are not bound to the toolkit. The following figure depicts an overview of the default experimental procedure, environmental structure, and data flow of the toolkit. Again, all of these defaults can be adjusted at will. The configurations presented here are intended to allow for a quick adaptation to other experiments.

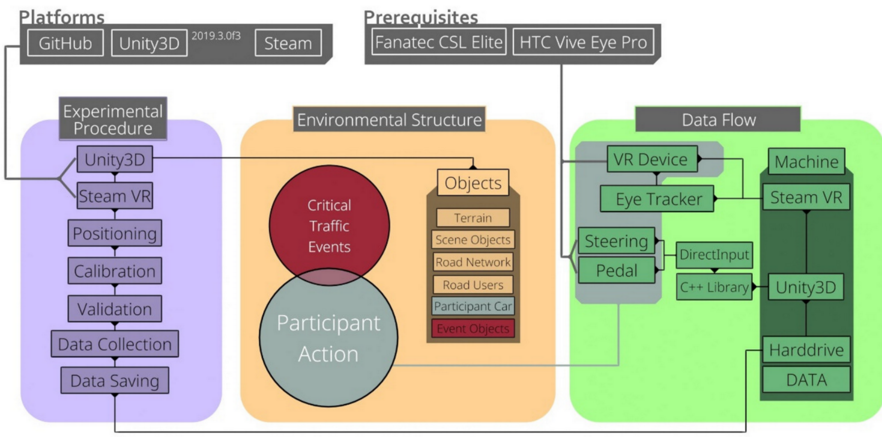

Figure 2.6: A simplified overview of the toolkit structure. It includes the default experimental procedure, a possible example of how the environmental structure can be used, and the standard data flow of the toolkit.

Platform

Project LoopAR is made with the Unity editor $2019.3.0f3$ (64bit). This software is a widely used game engine platform based on C# by Unity Technologies, supporting 2D, 3D, AR, and VR applications. The Unity editor and the Unity Hub run on Windows, Mac, and Linux (Ubuntu and CentOS), and built applications can be run on nearly all commercially usable platforms and devices. Unity also provides many available application programming interfaces and is compatible with numerous VR and AR devices (Juliani et al., 2020).

The backend code of the project LoopAR was developed entirely using C# within Unity3D Monobehaviour scripting API. The backend comprises functionalities including dynamic loading of the environment, AI car controls, pedestrian controls, event controls, car windshields augmented reality controller, data serialization, and eye-tracking connection. Additionally, the presented project contains a C++ library enabling the force feedback for Microsoft DirectX devices that enables various force feedback steering wheels to function as controllers altogether. LoopAR code has been developed with modularity in mind to avoid complicated and convoluted code. All functionalities can be enabled or disabled individually using the Unity editor's graphical interface based on need.

Figure 2.7: LoopAR map preview: mountain road (3.4 km), city (1.2 km), country road (2.4 km), and highway (3.6 km).

Virtual Environment

To test human–machine interactions, an interactive and realistic 3D environment is needed. LoopAR aims at a fully immersive experience of a highly automated car encountering critical traffic events. To be able to investigate different driving conditions and scenarios, we created four independent scenes. In the following section, the environment design decisions are presented together with a short description of the experimental scenes.

The LoopAR environment is based on real geographical information of the city of Baulmes in the Swiss Alps. We selected this region due to its variety of terrain, including a small village, a country road, a mountain pass, and a region suitable for adding a highway section, totaling around 25 km2 of environment and an 11 km continuous drive through different roads. To reduce the computational demands, we sliced the terrain into four areas. Due to the road network design, these separate environments can be merged (see Fig. 2.7). These areas demand different driving skills from an automated driving vehicle and a human driver, reacting in different situations with different conditions according to the landscape and traffic rules. To make the region accessible in Unity, we used the collaborative project OpenStreetMap (OSM) (OpenStreetMap, n.d.) and the open-source 3D software Blender (Foundation, n.d.).

OpenStreetMap is a project with the aim of creating a free map of the world. It collects the data of all commonly used terrains on maps. The project itself collects information, so the data are free of charge. The virtual environment contains a mountain road scene (see Fig. 2.8), including curvy roads winding through a forest and steep serpentines running down a mountain. These curvy roads require various driving speeds (from 30 km/h or slower, up to 100 km/h on straight stretches). The overall traffic density is low.

The second area of the environment is the village “Westbrück” (See Fig. 2.8). Here, it is possible to test events in a more inhabited environment. This environment is characterized by narrow streets and dense traffic in low-speed environments.The third scenario is the country road scene (see Fig. 2.8), designed for medium to high speed (70 km/h), medium traffic density, and a long view distance. The last scenario for the participants is the highway scene (see Fig. 2.9), enabling critical

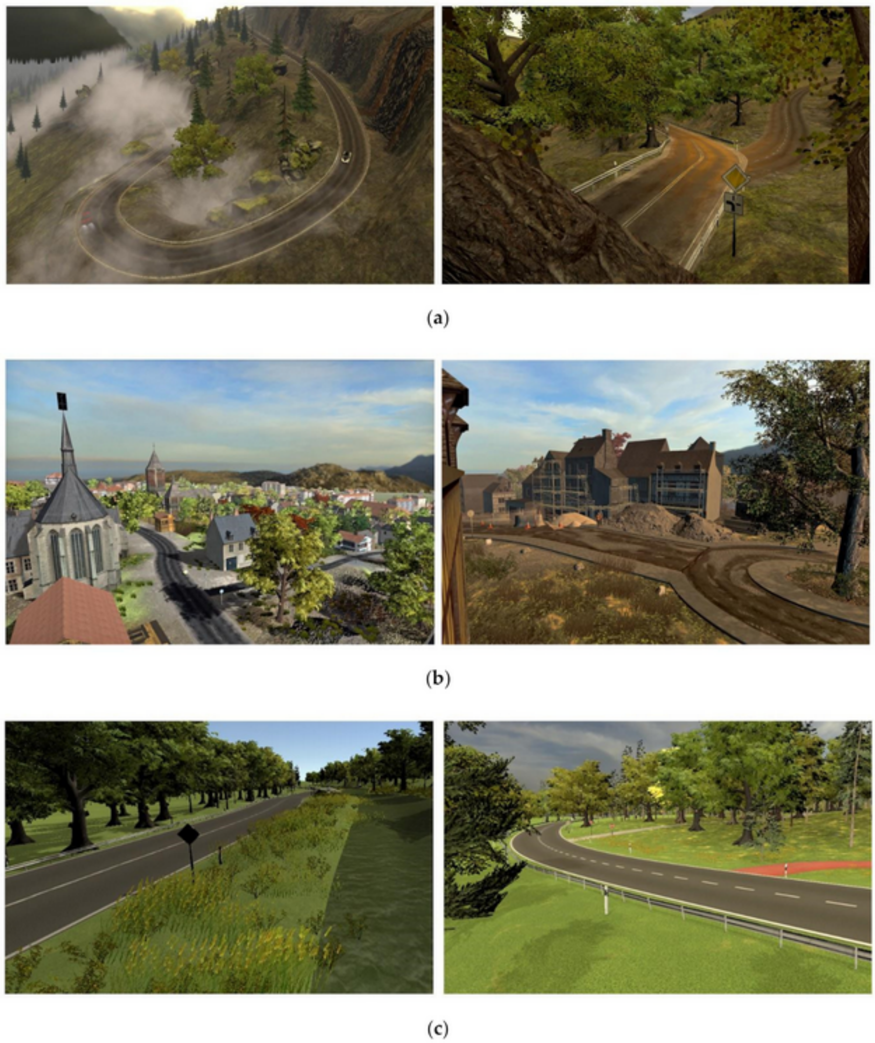

Figure 2.8: (a) Pictures of the different scenes from the mountain road. (b) Pictures of the different scenes from the village “Westbrück”. (c) Pictures of the different scenes from the country road.

traffic events with a higher speed and a low to medium traffic density.

(d)

Figure 2.9: (d) Pictures of the different scenes from the highway.

Critical Traffic Events

To test the participant's behavior in critical traffic events, we created limited event zones, where the monitoring of a participant can be achieved in a well-controlled environment. In Figure 2.10, one example of a traffic event is displayed. Each environment (mountain road, city, country road, and autobahn) has three critical traffic events. These zones are the core of the possible measurements in the presented toolbox. Simply put, the event system is realized by a combination of several trigger components. These independent triggers are activated when the participant enters the start trigger (Figure 4: green gate). The event zone is restricted within "boundary" triggers (Figure 2.10 : yellow boxes). These triggers get activated on contact, which is considered a participant's failure. Contact with the event triggers leads to a black screen followed by a respawn of the car at a point after the event (Figure 2.10: pink box) and giving back the car's control. An event is labeled as "solved" when the participant enters the end trigger (Figure 2.10: red gate) without crashing, i.e., making contact with the "boundary" triggers. All critical events can be adjusted at will, and a prefabricated file is stored in the repo to create new events. The triggers are all visible in editor mode but invisible to the participant.

Cars and Traffic Behavior

Within the event zones, dynamic objects, such as other road users, are needed to create realistic traffic scenarios. The repository presented here contains various animated pedestrians, animals, and cars to create a broad range of critical situations. Additionally, there are some busses and trucks, and some construction site vehicles that can be used. Furthermore, a user's own fbx models, as well as vehicles from the Unity asset store, can be added. For more details, please see the Supplementary Materials. All cars used are based on the Unity wheel collider systems of the Unity3D physics engine. In the Car Core Module, user input is translated into the motor control of the participant's car. The input consists of the motor torque, brake torque, and steering, which are applied to the wheels. This functionality is called AI control. It allows a seamless transition from automated to manual driving when activated. To facilitate realistic traffic behavior, an additional AI module enables cars to follow predefined paths. Paths followed by AI Cars and walking pedestrians were defined by mathematical Bézier curve paths (Prautzsch et al., 2002), which were realized by the Path-creator tool (Lague, 2021). Speed limit triggers inside the scene manipulate the AI's aimed speed, handling the input propagated to the Car Core Module. Another module of the car AI allows the AI cars to keep a distance from each other. The goal is to create an easily configurable and interchangeable traffic AI for multiple study designs. With these measures, we maximized the car physics and traffic simulation realism while ensuring easy adjustments.

Experiment Management

Data sampling, dynamic objects, and driving functionalities within the event zones are controlled by a system of experiment managers that handle scene-relevant information and settings shortly before and during the real experiment phase. It handles different camera settings, the information given by triggers inside the scene, and the participants' respawn in case of failure. Before an experiment starts, initial adjustments start the experiment. These adjustments configure the experiment to the participant and include the eye calibration, eye validation, seat calibration, and a test scene.

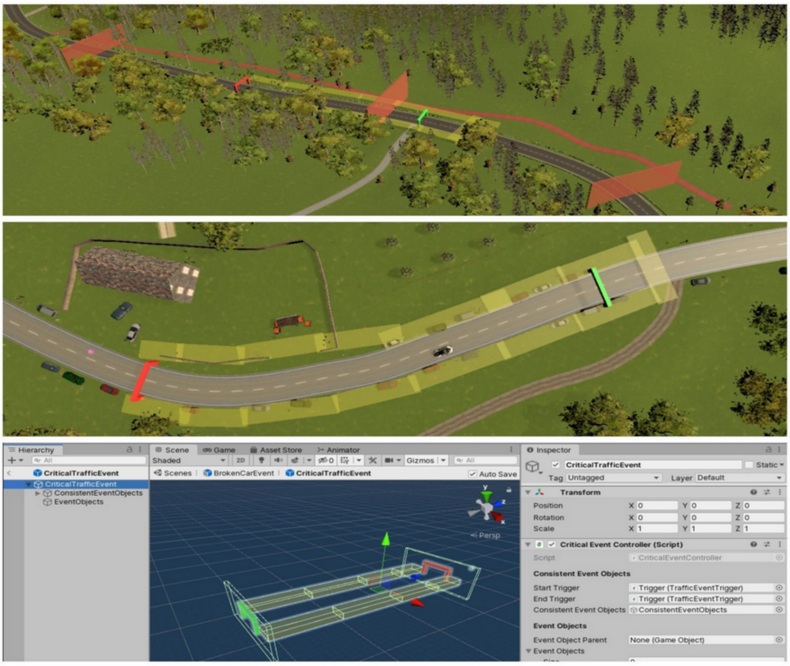

Figure 2.10: Traffic event prefab and its implementation.

The eye-tracking component in this setup comprises an eye-tracking calibration, validation, and online gaze ray-casting, which can record necessary gaze data during the experiment. The component was built for the Tobii HTC Vive Pro Eye device but is intended to keep the VR component interchangeable. It was designed as a simple connector to tap into SRanipal and the Tobii XR SDK (see Fig. 2.11). The eye calibration is performed with the built-in Tobii eye calibration tool. The validation is set in the corresponding validation scene, which provides a simple scenario with a fixation cross. Validation fails if the validation error angles exceed an error angle of 1.5° or the head is moved by 2” from the fixation cross. During the experiment, the eye orientation, position, and collider hits are stored with a calculated gaze ray of both eyes. Currently, it is set to receive information about any object inside these rays to prevent the loss of viable information by objects covering each other.

In addition to the eye-tracking data, input data of the participant as well as scene-relevant information, such as the number of failed critical traffic events, are saved using generic data structures and Microsoft Linq, serialized into JavaScript object notation (JSON), and saved with a unique ID at the end of each scene. The generic data structure used in the project ensures flexibility, as different data types can be added or removed from the serialization component. This approach guarantees the highest compatibility with varying analysis platforms such as R or Python for the data gathered with LoopAR.

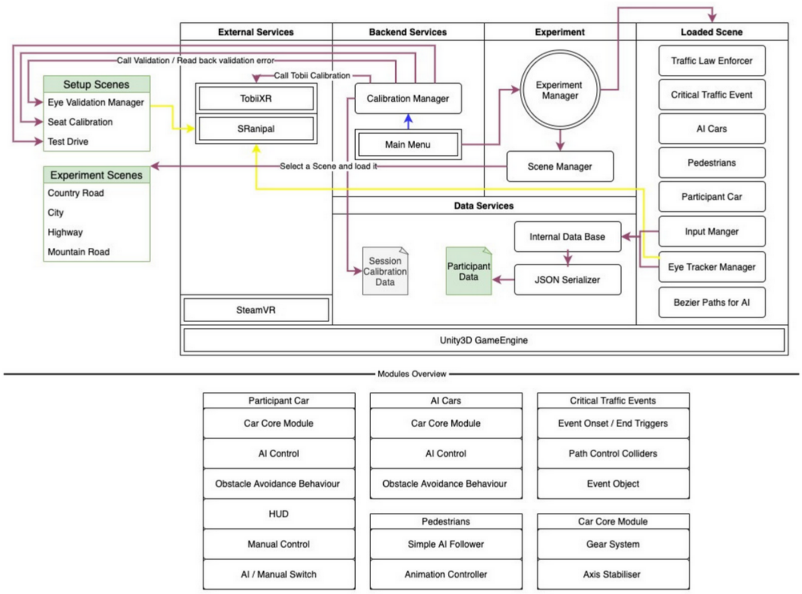

Figure 2.11: Scheme of the LoopAR functionalities and components illustrating the interaction of the different services and manager scripts within the Unity environment.

By conducting data saving, and given the nature of the experimental setup, we aim for a stable and high frame rate to provide a less sickness-inducing experience. A stable visual experience can be seen as a prerequisite to avoid potential sickness (LaViola, 2000). The desired optimum for the experiments is a stable frame rate matching the fixed rate of 90 Hz used by the manufacturers HTC and Oculus. Our current frame rate in the different scenes yields an average of 88 samples per second in our test setup, matching the maximum sampling rate of the HTC Vive with 90 fps.

Requirements

The setup used and presented here is thought to be a cost-efficient and very mobile replacement for maintenance-intensive, rigid, and expensive driving simulators for studies on human behavior in the context of self-driving cars. A key advantage is freedom regarding the selected components. The only requirement for operation is granting the computing power for the entire system, which consists of a core setup only of a computer, a head-mounted display, and a steering wheel (see Table 1).

As a virtual reality device, we used the HTC Vive Pro Eye with an integrated Tobii Eye Tracker. It is a cable-bound head-mounted display that enables the participant to transfer movements into virtual reality. Although we are using the Vive Pro exclusively at our department, the LoopAR experiment is not dependent on this specific VR device. We used the components of the setup with 90 fps sampling and display.

2.3.4 Discussion

In the presented paper, we describe LoopAR as a modular toolkit to test a takeover of control in critical traffic situations from automated cars to human drivers by combining VR and eye-tracking in an interactive and immersive scenario. Its current state and design provide a promising, new, low-cost, and mobile setup to conduct studies that were traditionally only done in stationary simulators. The current code, as well as the 3D environments, can be adjusted at will. With newly implemented code, it is not only possible to simulate a large and highly realistic VR environment, but it is also possible to create a broad range of applications in VR research that is not only bound to HMI investigations. A large part of the assets used are from Unity's asset store and the 3D platforms Sketchfab and Turbosquid. Therefore, it is possible to change the number, size, and shape of all objects in each scene.

All of the functionalities above, and assets presented here, are under constant improvement. By writing, five new projects, ranging from ethical decision-making over EEG implementation to human spatial navigation, arise from the presented toolkit, which will also develop new assets and features implemented into the toolkit later on. The authors want to emphasize the modularity and adaptability of this VR toolkit.

User Reports

To check for the user friendliness of the presented toolkit, a System usability score (SUS)-based report was performed (Lewis, 2018). Here, we asked 11 of the current users between the age of 23 and 34 (5 female) to evaluate the usage of the main features in the toolbox starting from cloning the repository, adjusting the environment, and manipulating dynamic objects in an example scene. While doing so, we asked the participants to evaluate the feasibility of the tasks. User experience in Unity and C# programming varied from no experience to expert levels with more than 3 years of experience. Our top findings, depicted in indicate that the toolbox is perceived as well documented, and advanced Unity users faced no major problems building and altering their project created with this toolbox (see Fig. 2.12). While some steps in the procedures might be challenging to new users, the Westdrive X LoopAR toolbox seems to be a useful foundation for all users.

2.3.5 Conclusion

This article describes a new virtual reality toolkit for Unity applications investigating human–machine interaction in highly automated driving developed by us. The presented setup is thought to be a mobile, cost-efficient, and highly adapt-

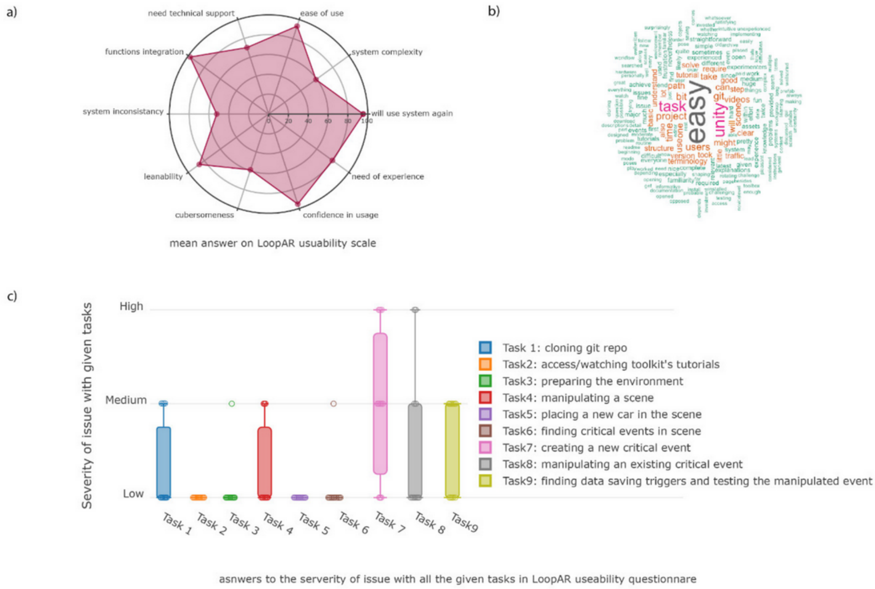

Figure 2.12: Visualization of the usability report items: (a) a radar plot of the system usability scale data; (b) a word cloud showing most frequently used words in the comments; and (c) a severity of issue bar plot, related to the tasks in the usability report. Low equals no delay in time or perceived obstacles, medium refers to a completed task with added effort. High indicates noticeable delay or frustration and that the participant may not be able to complete the task.

able alternative to chassis simulators that closely monitor the participants. It is particularly noteworthy that there is not only a drastic reduction in costs but also an improvement to the adaptability of the software as well as the used hardware. All components are fully upgradable, in case there are better products in terms of image quality or computing power. LoopAR allows interested researchers to conduct various virtual reality experiments without creating the needed environment or functionalities themselves. For this, we have provided an area of almost 25 km2 based on OSM data. The toolkit presented here also includes all the necessary assets and basic prefabs to quickly and precisely create a wide variety of virtual environments. Additionally, the LoopAR toolkit contains components of the experimental procedure and data storage.

Supplementary Materials

The following are available online at https://www.mdpi.com/1424-8220/21/5/1879/s1, Unity 3D: www.Unity3d.com; Online Character animation: www.mixamo.com; Adobe Fuse CC: www.adobe.com/products/fuse.html; Blender 2.81: www.blender.org.

Author Contributions

F.N.N. and M.A.W. wrote this paper. Both authors designed the project. S.U.K., P.K., and G.P. supervised the LoopAR project. N.M. developed major parts of the AI, functional modules, and User Interfaces. J.M.P. realized scene building and the HUD functionalities. L.K. designed the mountain road scene and provided performant assets. L.T. developed large parts of the software architecture and acted as a software engineer for the project's functional compartments. A.H. designed the highway scene and provided additional assets. L.M.K. designed the country road scene, provided assets, and contributed to HUD-related literature background. P.S. was involved in designing the city scene, as well as managing and creating assets. F.N. developed and designed the test drive scene. Additionally, we would like to thank Debora Nolte, Shadi Derakhshan, and Vincent Schmidt for their valuable user feedback. All authors have read and agreed to the published version of the manuscript.

2.4 Stress testing VR Eye-tracking System Performance

This section was submitted as a peer reviewed conference talk in the 3rd International Neuroergonomics Conference together with Ashima Keshava, Nora Maleki, Linus Tiemann, and Peter König. See Publication List for details.

2.4.1 Extended Abstract

Eye-tracking experiments in virtual reality (VR) have become progressively popular in the last decade. These experiments measure human eye movement behavior in naturalistic settings that afford complex, natural head and body movements. Given the complexity, eye-tracking systems require high spatial accuracy and precision of the measured gaze in the face of natural movements, differing illumination, depth of field, and calibration decay. (Holmqvist et al., 2012) have stressed the need for assessing eye-tracking data quality in general. Furthermore, there is a lack of data quality standards when it comes to VR head-mounted displays specifically. The present study aims to introduce a standardized way of benchmarking VR eye-tracking systems to assess their feasibility for vision research in mobile settings.

We adapted a 2D screen-based eye-tracker test battery Ehinger et al., 2019 to VR-based head-mounted displays. The test battery includes ten spatial accuracy and precision tests for standard gaze parameters like gaze position, pupil dilation, blink detection, and smooth pursuit. We then used the test battery to compare the performance of two commercially available VR head-mounted displays (HTC Vive Pro Eye and Varjo VR-2 Pro) with a built-in eye-tracker for 13 participants (Figure 2.13A).

Here, we report our results based on the most critical metrics, namely: 1. Spatial Accuracy (Figure 2.13B): we calculated the calibration error across a 5x5 grid of fixation locations. Our results show that both HTC Vive Pro Eye and Varjo VR-2 Pro have a mean calibration error greater than 1 degree without an explicit validation of the calibration accuracy. In the horizontal direction, HTC Vive had a mean error of 1.28°, IQR=[0.60, 1.17], and Varjo had a mean error of 3.29°, IQR=[1.55, 3.45]. In the vertical direction, HTC Vive had a mean error of 0.89°, IQR=[0.50, 1.19] and Varjo had a mean error of 4.93°, IQR=[2.24, 6.93]; 2. Spatial Precision (Figure 2.13C): we used the median absolute deviation of the calibration error across the 5x5 grid to measure the eye trackers' spatial precisions. We found that the mean precision in the horizontal axis of the HTC Vive Pro Eye was 3.22° (SD: ±0.75) and 4.76° ± 1.48 for Varjo. In the vertical direction, HTC VIVE Pro Eye had a precision of 1.86°± 0.76, and Varjo had 4.02°±2.33; 3. Calibration Decay: to assess the decay of calibration during the experiment, we calculated the mean difference in calibration error just after eye-tracker calibration and at the end of each test block. HTC Vive showed a mean calibration decay of 4.09° ± 1.06 in the horizontal direction and 3.21° ± 1.09 in the vertical direction. In contrast, the Varjo system showed a calibration decay of 5.86° ± 2.46 in the horizontal direction and 5.11° ±1.95 in the vertical direction; 4. Effect of Illumination on Pupil Dilation (Figure 2.13D): we further investigated the pupil size detection differences between the eye trackers for different illumination levels. Our results show that the Varjo VR-2 eye tracker estimated larger normalized pupil sizes than the HTC VIVE (mean difference = 2.55 % ± 4.47); 5. Blink Detection (Figure 2.13E): we investigated how well the two eye trackers detected blinks by asking subjects to voluntarily blink 10 times during a test block. We found the HTC Vive Pro Eye detected 10.49 ± 3.14 blinks, and the Varjo VR-2 Pro system detected 1.40 ± 1.82 blinks; 6. Smooth Pursuit (Figure 2.13F): In the smooth pursuit task, we found that the HTC Vive system tracks the eyes at -0.18°/s ± 4.11 slower than the stimulus velocity, whereas the Varjo system tracks the eyes at -3.84°/s ± 3.27 slower than the stimulus velocity. Our results show that both VR eye-tracking systems are somewhat error-prone and can have high variance across different subjects. Hence, vision researchers should not take the quality of the data measured by these systems as a given. In studies that rely on high spatial accuracy or measurement of specific gaze features like blinks or pupil dilation, the eye-tracking equipment alone can make an immense difference. Our study offers an implemented test battery to evaluate and benchmark VR eye-tracking systems based on several gaze features useful for naturalistic experiments. The tests can comprehensively assess the quality of commercially available VR eye trackers beyond the values provided by the manufacturers. Furthermore, we have made the VR setup, the collected data, and the analysis pipeline available publicly to help researchers adapt this study for any VR-based eye tracker.

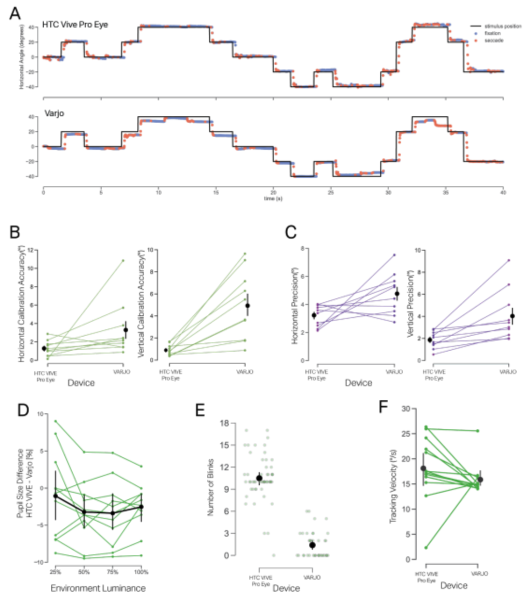

Figure 2.13: A) Exemplar raw data showing horizontal gaze angle for the fixation probe shown in VR for the two head-mounted displays (HMDs). The blue samples correspond to fixations and the orange to the saccades. B) The calibration error for the two HMDs across 10 subjects. We computed the calibration error (in visual degrees) as the 20% winsorized mean of the difference between the fixation probe position and the actual fixation position. Thus, each dot represents one subject and the calibration error for the two devices. C) The precision of the HMDs across subjects. Here, we used median absolute deviation as a metric of precision, where lower values correspond to high precision and vice versa. D) % Difference in the normalized pupil sizes measured by the two devices for the different environment luminance. Green dots indicate each subject, and the black dots represent mean difference, and the error bars represent the standard error of mean. E) Number of Blinks detected by the eye tracker. Each green dot represents the number of blinks per subject and test block. The filled black dots represent the mean number of blinks, and the error bars show the standard error of the mean. F) Velocity of the tracked gaze for a moving stimulus during smooth pursuit. Green dots indicate each subject, and the black dots represent mean difference, and the error bars represent the standard error of mean.

2.5 A framework for low-level joint action in VR

This section was submitted as a peer reviewed conference talk in the 3rd International Neuroergonomics Conference together with Nora Maleki1, Marten Mildt, Florian Pätzold, Vincent Schmidt1, Linus Tiemann1, Jasmin L. Walter1, Josefine A. Zerbe1, Dirk C. Gütlin, Anke Haas Anne Lang, Peter König and Artur Czeszumski. See Publication List for details.

2.5.1 Extended Abstract

Social interactions, including joint actions, are a central aspects of human life (De Jaegher et al., 2010; Frith, 2007; McCabe et al., 2001). Joint action can be described as any social interaction whereby two or more people temporally and spatially align their actions (Sebanz et al., 2006). Due to their interactive nature, however, joint action studies are usually conducted under strictly supervised laboratory conditions with simplistic stimuli to obtain maximum control over all variables (Redcay and Schilbach, 2019). As a consequence, traditional paradigms often struggle to achieve an adequate level of ecological validity (Parsons et al., 2017). A potential solution to studying joint action in a more realistic setting without incriminating experimental control is the use of Virtual Reality (VR), in particular head-mounted displays (Chicchi Giglioli et al., 2017; Marín-Morales et al., 2018). Moreover, aspects of participants' behavior can be measured and controlled in real-time, including subtle factors like non-verbal communication or interpersonal distance. Furthermore, VR technologies enable researchers to conduct experiments that are dangerous or unethical in real life (Niforatos et al., 2020; Skulmowski et al., 2014a). Consequently, implementing joint action paradigms in VR could significantly reduce their variability while substantially increasing an experiment's reliability, replicability, and transparency (Pan and Hamilton, 2018).

Given its advantages, it is curious why there is a lack of low-level joint action research conducted in VR and a possible reason might be technical limitations due to multiplayer networking. Joint action studies focus on subtle behavioral factors and often rely on eye-tracking or reaction time measurements. Networking these variables becomes crucial for real-time interaction, and hence, an authentic simulation. However, since the majority of networking solutions are designed for consumer applications such as online gaming, they often lack low-level control of the networking variables and other essential modification options. To solve this need, we propose the new networking framework “LightNet” that is specifically designed for multiplayer experiments in VR. LightNet is a C# library created for - but not limited to - the usage with the game engine Unity and it allows for customizable real-time interaction between participants. LightNet provides complete control over sent and received data, and allows precise assignment of transferred variables, therefore improving data management options, performance, and frame rate of the virtual experiment. Like this, redundant or irrelevant information will not be transferred which makes data propagation more efficient. Further, LightNet is utilizing a reliable but slow TCP (Transmission Control Protocol) channel for transferring sensitive information like the experiment state, and an unreliable but fast UDP (User Datagram Protocol) channel for data that requires a quick response, like precise synchronization of position data between participants. Typical network solutions also emphasize a symmetric design of the contents and roles of agents inside the virtual environment, thereby restricting the design of dyadic experiments which is not the case with LightNet. Additionally, it benefits from a lightweight architecture which can facilitate the usage for experimenters. In short, due to its customizable structure, LightNet allows individual modifications and customization of data transmission between participants, providing the necessary control that is crucial for low-level joint action experiments. To test its practicality we implemented two networking examples, each based on a well-known joint action study. The first example is based on a shared gaze study during a visual search task (Brennan et al., 2008). Similar to the original design, participants complete an O-in-Q search task but we adjusted the stimuli to trophies on a “Wall of Fame” (Figure 2.14b). To implement the networking functionality, we needed to transfer the continuous shared gaze data of both participants while also transmitting the complex stimuli information due to changing, randomized number and rotation of distractors and target. The second experimental design originally examined mechanisms of anticipatory control during joint action (Knoblich and Jordan, 2003). While the task stayed similar to the original study, we adjusted the design so that participants control the beam of a laser cannon (tracker) on a spaceship to stay as close as possible to a moving target (Figure 2.14c,d). The challenges of this example centered around networking and controlling the tracker and auditory feedback as they depend on both participants’ input. In both networking examples, we explore different levels of modifying the 3D environment to increase immersion, encourage participant engagement, and add storytelling elements to the task. As our focus was to design a general framework, we refrained from recording data after successfully piloting the examples. However, LightNet (github.com/Ben1138/lightnet-unity) and the second networking example (github.com/Westdrive-Workgroup/Dyadic-interactions-2) are freely available for implementation or conduction. Overall, the networking examples demonstrate the usability of our networking solution LightNet and provide a framework for low-level joint action research in VR.

In conclusion, VR could be a promising solution, allowing real-time interaction in a controlled and ecologically valid setting. Applied to joint action, VR potentially increases a study's reproducibility, replicability, and transparency. Since appropriate networking is crucial for multiplayer experiments, we propose the new networking solution LightNet. Further, by implementing two networking examples, we provide a proof of concept for our framework. Thus, the presented networking solution LightNet and the networking examples can make VR more accessible for the scientific landscape of joint action research.

Figure 2.14: Networking examples. Example 1, (a) main menu (left) and LightNet graphical interface (right), (b) experimental setup with visible gaze spheres. Example 2, (c) handle to control the laser beam (tracker), and target, (d) dyadic experimental setup.

3 From Lab to Virtual Lab

3.1 Layman's summary

This thesis tries to brand virtual reality as a middle ground between traditional laboratory-based experiments and real work experimentation. As such, it advocates the use of virtual reality as a valid experimentation method. Therefore, we try to validate this claim by replicating well known experiments using virtual reality. If we observer the same result as the original experiments, we can confidently state that virtual reality experiments are at least on the same level as the traditional lab based experiments. Additionally, we tried to recreate or design experiments that benefit from natural interaction with the environment or environmental objects.

The experiments presented in this chapter involve natural interactions with environmental objects. Therefore, this requirement needs planning for the action and will affect where and in what order participants will look at different objects. In the first study, the goal is to lift or use a tool. The tool might be familiar or unfamiliar in shape. Nonetheless depending on whether one needs to use or simply lift a tool, it requires the participants to study and analyze the presented tool with their eyes. Moreover, we have asked the participants to use natural hand gestures and movements to grasp and utilize the tools. In the second experiment, participants had to sort objects on a shelf based on given instruction. The task resembled a game of Sudoku[I] but on color and shape of objects instead of number and in a more simplified manner. However, these experiments involve not only planning but also the full-body movement of the participant. Both studies shed light on the importance of naturalistic interaction and realism on the data we can gather in an experiment and, therefore, their contribution to understanding complex cognitive processes such as action planning.

[I] a Japanese logic-based number-placement puzzle

3.2 Action Affordance Affects Proximal and Distal Goal-oriented Planning

This section was submitted as a peer reviewed article in the European Journal of Neuroscience together with Ashima Keshava, Nina Gottschewsky, Stefan Balle, Thomas Schüler, and Peter König. See Publication List for details.

3.2.1 Abstract

Seminal studies on human cognitive behavior have been conducted in controlled laboratory settings, demonstrating that visual attention is mainly goal-directed and allocated based on the action performed. However, it is unclear how far these results generalize to cognition in more naturalistic settings. The present study investigates active inference processes revealed by eye movements during interaction with familiar and novel tools with two levels of realism of the action affordance. We presented participants with 3D tool models that were either familiar or unfamiliar, oriented congruent or incongruent to their handedness, and asked participants to interact with them by lifting or using. Importantly, we used the same experimental design in two setups. In the first experiment, participants interacted with a VR controller in a low realism environment; in the second, they performed the task with an interaction setup that allowed differentiated hand and finger movements in a high realism environment. We investigated the differences in odds of fixations and their eccentricity towards the tool parts before action initiation. The results show that participants fixate more on the tool's effector part before action initiation for the use task for unfamiliar tools. Furthermore, with more realistic action affordances, subjects fixate more on the tool's handle as a function of the tool's orientation, well before the action was executed. Secondly, the spatial viewing bias on the tool reveals early fixations are influenced by the task and the familiarity of the tools. In contrast, later fixations are associated with the manual planning of the interaction. In sum, the findings from the experiments suggest that fixations are made in a task-oriented way to plan the intended action well before action initiation. Further, with more realistic action affordances, fixations are made towards the proximal goal of optimally planning the grasp even though the perceived action on the tools is identical for both experimental setups. Taken together, proximal and distal goal-oriented planning is contextualized to the realism of action/interaction afforded by an environment.

3.2.2 Introduction

A longstanding goal of the cognitive sciences is to understand cognition, behavior, and experience as it unfolds in the natural world (Parada and Rossi, 2020). Given the technological advancements made in the last decade, there are few methodological roadblocks to understanding natural cognition where laboratory studies can be extended to naturalistic settings and hopefully lead towards new insights (Ladouce et al., 2017; Parada, 2018). More recently, a pragmatic turn has emerged in the field where there is a greater push towards incorporating the body and bodily actions to infer cognitive function (Engel et al., 2013).

Human tool use is an explicitly natural cognitive function that involves the transfer of proximal goals (e.g., placement of grasp) to distal goals for the tool (Arbib et al., 2009). Moreover, simple tools fundamentally expand the body representations to include representations of the tool in the peripersonal space (Berti and Frassinetti, 2000; Farnè et al., 2005; Maravita et al., 2002). Furthermore, tool use is differentiated from other object-based actions where the tool is “acted with” (S. H. Johnson and Grafton, 2003 and requires semantic knowledge of the tool as well as the necessary skill to perform actions with it Johnson-Frey, 2004). Hence, tool use involves complex behaviors ranging from cognitive and semantic reasoning to perceptual and motor processing.

When using tools, a wealth of information is parsed to produce the relevant action. The semantic knowledge associated with the tool helps understand how it is used, the mechanical knowledge maps the physical properties of the tool for potential usage, and finally, sensorimotor knowledge helps decipher possible movements required to use the tool (Baumard et al., 2014). The amalgamation of these knowledge sources (which can be unique to a tool) necessitates planning any action associated with the tool. When this knowledge is not readily available, inferential processes must be deployed to deduce the relevant action.

In naturalistic settings, studies have shown that eye movements are made to locations in the scene in anticipation of the following action (Hayhoe, 2004; M. F. Land and Hayhoe, 2001; Pelz and Canosa, 2001). (Belardinelli, Stepper, et al., 2016) showed that eye movements are goal-oriented and are modulated in anticipation of the object interaction task. There is strong evidence that task plays a vital role in how the eyes scan the scene and are differentiated between passive viewing and pantomimed interaction (Belardinelli et al., 2015). Similarly, Keshava et al., 2020 showed that rudimentary object interactions can be decoded using eye-movement data alone. rudimentary object interactions can be decoded using eye-movement data alone. Even in the absence of an interaction, task relevance plays an important role (Castelhano et al., 2009; Henderson and Hayes, 2017). These studies point towards gaze control being the consequence of knowledge and task-driven predictions (Henderson, 2017).

Moreover, (Belardinelli, Barabas, et al., 2016) investigated the role of anticipatory eye movements when interacting with familiar and unfamiliar tools in a controlled lab setting. These tools had differentiable parts: tool handle and effector. The results showed that in the case of unfamiliar tools, preparatory eye movements are made to the tool-effector to extract the mechanical properties of the tool as the semantic information was not readily available. This effect was enhanced when subjects were asked to perform tool-specific movements instead of a generic action of lifting the tool by the handle. The authors, hence, concluded that eye movements are used to actively infer the appropriate usage of the tools from their mechanical properties. In the study, the tools were presented as images on a screen, and participants pantomimed lifting or using the tool. While the study revealed valuable insights into anticipatory gaze control, a question remains if these results are part of natural cognition and can be reproduced in more realistic environments.

Herbort and Butz, 2011 further investigated the interaction of habitual and goal-directed processes that affect grasp selection while interacting with everyday objects. They presented objects in different orientations and showed that grasp selection depended on the overarching goal of the movement sequence dependent on the object's orientation. Belardinelli, Stepper, et al., 2016 further showed that fixations have an anticipatory preference for the region where the index finger is placed. Consequently, the location of fixations is predictive of both proximal goals of manual planning and task-related distal goals.

When studying anticipatory behaviors corresponding to an action, one must also ask whether symbolized action is enough and how real the action should be. Króliczak et al., 2007 showed brain areas typically involved in real actions are not driven by pantomimed actions and that pantomimed grasps do not activate the object-related regions within the ventral stream. Similarly, Hermsdörfer et al., 2012 showed a weak correlation between the hand trajectories for pantomimed and actual tool interaction. These studies indicate that the realism of sensory and tactile feedback while acting (e.g., a grasp) can be an essential factor when studying anticipatory behavioral control.

In virtual reality (VR), realistic actions can be studied by simulating an interaction within an environment. Using interfaces such as VR controllers, ego-centric visual feedback of a hand can be simulated. These interfaces usually consist of hand-held devices that are tracked in space and through which different actions are controlled by pressing buttons. One advantage of controller-based VR interaction is the possibility of haptic feedback. One disadvantage is that the hand posture while holding the controller does not always correspond to the user's virtual visual feedback when the simulated hand performs the action. Conversely, camera-based interaction interfaces such as LeapMotion, capture the real-time movements of the user's hand and use finger gestures, like wrap grasp or pinch grasp, to control different actions in the environment. These interfaces give the user realistic visual feedback of their finer hand and finger movements, while they can not give haptic feedback. Consequently, the chosen method of interaction in VR can afford different levels of realism and could elicit different behavioral responses.

In the present study, we investigated anticipatory gaze control in two different experiments. We were interested in the extent to which the realism of the action affordance and the environment modified the results shown by Belardinelli, Barabas, et al., 2016. We asked participants to lift or use 3D models of tools in VR that were categorized as familiar or unfamiliar. Additionally, we extended the experimental design to include the tool handle's spatial orientation, congruent or incongruent to the subjects' handedness.

In experiment-I, subjects performed the experiment in a low realism environemnt and action affordance and interacted with the tool models using a VR controller, which mimicked grasp in the virtual environment by pulling the index finger. In experiment-II, subjects were immersed in a high realism setting where they interacted with the tools using LeapMotion, which required natural hand and finger movements. Thus, the action affordance appeared closer to the real world. With this experimental design, we investigate the influence of task, tool familiarity, the spatial orientation of the tool, and, notably, the impact of the realism of the action affordance.

3.2.3 Methods

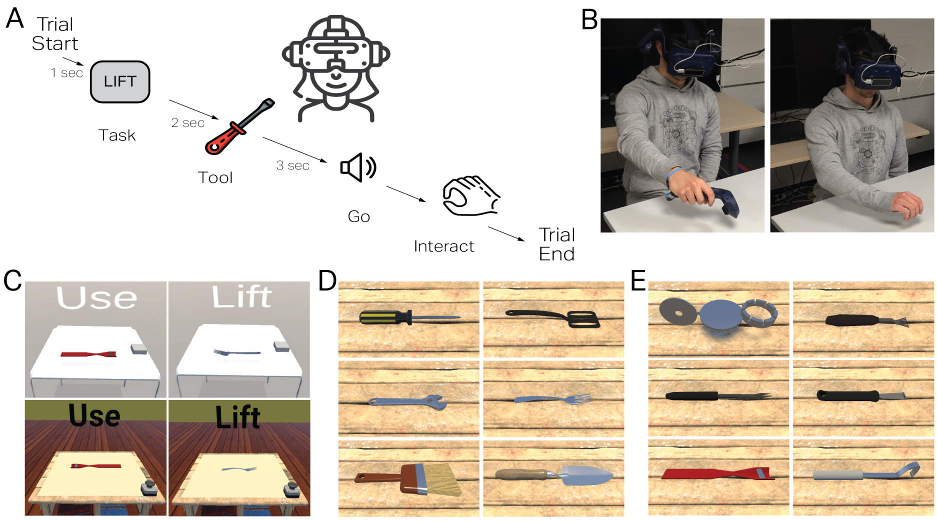

Figure 3.1: Experimental Task. In two virtual environments participants interacted with tools in two ways (LIFT, USE). The tools were categorized based on familiarity (FAMILIAR, UNFAMILIAR) and presented to the participants in two orientations (HANDLE LEFT, HANDLE RIGHT). The two virtual environments differed based on the mode of interaction and perceived realism, wherein in one experiment, subjects' hand movements were rendered virtually using the HTC-VIVE controllers. In the other experiment, the hands were rendered using LeapMotion, allowing finer hand and finger movements. **Panel A** shows the timeline of a trial. **Panel B** shows a subject in real-life performing the task in the two experiments. **Panel C** shows the differences in realism in the two experiments; TOP panels correspond to experiment with the controllers, the USE and LIFT conditions for an UNFAMILIAR and FAMILIAR tool, respectively with the tool handles presented in two different orientations. BOTTOM panels illustrate the three different conditions in a more realistic environment with LeapMotion as the interaction method. **Panel D** Familiar tools, from top-left: screwdriver, spatula, wrench, fork, paintbrush, trowel. **Panel E** Unfamiliar tools, from top-left: spoke-wrench, palette knife, daisy grubber, lemon zester, flower cutter, fish scaler.

Experimental Task

Subjects were seated in a virtual environment where they had to interact with the presented tool by either lifting or pretending its use. The time course of the trials is illustrated in Figure 3.1A. At the start of a trial, subjects saw the cued task for 2 sec after which the cue disappeared, and a tool appeared on the virtual table. Subjects were given 3 sec to view the tool, after which there was a beep (go cue) which indicated that they could start manipulating the tool based on the cued task. Subjects were seated in a virtual environment where they had to interact with the presented tool by either lifting or pretending its use. After interacting with the tool, subjects pressed a button on the table to start the next trial.

Participants

For experiment-I with the HTC Vive controller's interaction method, we recruited 18 participants (14 females, mean age=23.68, SD=4.05 years). For experiment-II with the interaction method of LeapMotion, we recruited 30 participants (14 female, mean age=22.7, SD=2.42 years). All participants were recruited from the University of Osnabrück and the University of Applied Sciences Osnabrück. Participants had a normal or corrected-to-normal vision and no history of neurological or psychological impairments. All of the participants were right-handed. They either received a monetary reward of C10 or one participation credit per hour. Before each experimental session, subjects gave their informed consent in writing. They also filled out a questionnaire regarding their medical history to ascertain they did not suffer from any disorder/impairments which could affect them in the virtual environment. Once we obtained their informed consent, we briefed them on the experimental setup and task.

Experimental Design and Procedure

The two experiments differed based on the realism of the action affordance and the environment. Figure 3.1B illustrates the physical setup of the participants for the two experiments. In experiment-I, subjects interacted with the tool models using the HTC Vive VR controllers. While in experiment II, subjects' hand movements were captured by LeapMotion.

Figure 3.1C illustrates two exemplar trials from the experiments. We used a 2x2x2 experimental design for both experiments, with factors task, tool familiarity, and handle orientation. Factor task had two levels: LIFT and USE. In the LIFT conditions, we instructed subjects to lift the tool to their eye level and place it back on the table. In the USE task, they had to pantomime using the tool to the best of their knowledge. Factor familiarity had two levels, FAMILIAR and UNFAMILIAR, which corresponded to tools either being everyday familiar tools or tools that are not seen in everyday contexts and are unfamiliar. The factor handle orientation corresponded to the tool handle, which was presented to the participants either on the LEFT or the RIGHT. Both experiments had 144 trials per participant, with an equal number of trials corresponding to the three factors. Subjects performed the trials over six blocks of 24 trials each. We measured the eye movements and hand movements simultaneously while subjects performed the experiment. We calibrated the eye-trackers at the beginning of each block and ensured that the calibration error was less than 1 degree of the visual angle. At the beginning of the experiment, subjects performed three practice trials with a hammer to familiarize themselves with the experimental setup and the interaction method. Each experiment session lasted for approximately an hour. After that, subjects filled out a questionnaire to indicate their familiarity with the 12 tools used in the experiment. They responded to the questionnaire based on a scale of 5-point Likert-like scale where 1 corresponded to "I have never used it or heard about it," and 5 referred to "I see it every day or every week."

Experimental Stimuli

The experimental setup consisted of a virtual table that mimicked the table in the real world. The table's height, width, and length were 86cm, 80cm, and 80cm, respectively. In experiment-I, subjects were present in a bare room with grey walls and constant illumination. They sat before a light grey table, with a dark grey button on their right side to indicate the end of the trial. Similarly, in experiment-II, subjects were present in a more immersive, realistic room. They sat in front of a wooden workbench with the exact dimensions of the real-world table and a buzzer on the right to indicate the end of a trial. We displayed the task (USE or LIFT) over the desk 2m away from the participants for both experiments.

For both experiments, we used the tool models as presented in Belardinelli, Barabas, et al., 2016. Six of the tools were categorized as familiar (Figure 3.1D) and the other six as unfamiliar (Figure 3.1E). We further created bounding box colliders that encapsulated the tools to capture the gaze position on the tool models. The mean length of the bounding box was 34.04cm (SD=5.73), mean breadth=7.60cm (SD=3.68) and mean height= 4.17cm (SD=2.13). To determine the tool effector and tool handle regions of interest, we halved the length bounding box colliders from the center of the tool and took one half as the effector and the other half as the handle. This way we refrained from making arbitrary-sized regions-of-interest for the different tool models.

Apparatus

For both experiments, we used an HTC Vive head-mounted display (HMD)($110°$ field of view, 90Hz, resolution 1080 x 1200 px per eye) with a built-in Tobii eye-tracker [II]. The HTC Vive Lighthouse tracking system provided positional and rotational tracking and was calibrated for a 4m x 4m space. For calibration of the gaze parameters, we used 5-point calibration function provided by the manufacturer. To make sure the calibration error was less than $1°$, we performed a 5-point validation after each calibration. Due to the nature of the experiments, which allowed a lot of natural head movements, the eye tracker was calibrated repeatedly during the experiment after each block of 36 trials. We designed the experiment using the Unity3D game engine [III] (v2019.2.14f1) and controlled the eye-tracking data recording using HTC VIVE Eye Tracking SDK SRanipal[IV] (v1.1.0.1).

For experiment-I, we used HTC Vive controller[V] (version 2.5) to interact with the tool. The controller in the virtual environment was rendered as a gloved hand. When participants pulled the trigger button of the controller with their right index finger, their right virtual hand made a power grasp action. To interact with the tools, subjects pulled the trigger button of the controller over the virtual tools and the rendered hand grasped the tool handle.

Similarly, in experiment-II, we used LeapMotion[VI] (version 4.4.0) to render the

[II] https://enterprise.vive.com/us/product/vive-pro-eye-office/
[III] Unity, www.unity.com
[IV] SRanipal, developer.vive.com/resources/vive-sense/sdk/vive-eye-tracking-sdk-sranipal/
[V] SteamVR, https://valvesoftware.github.io/steamvr_unity_plugin/articles/Quickstart.html
[VI] LeapMotion Unity modules, https://developer.leapmotion.com/unity

hand in the virtual environment. Here, subjects could see the finer hand and finger movements of their real-world movements rendered in the virtual environment. When participants made a grasping action with their hand over the virtual tool handle, the rendered hand grasped the tool handle in the virtual environment.

Data pre-processing

Gaze Data

As a first step, using eye-in-head 3D gaze direction vector for the cyclopean eye we calculated the gaze angles in degrees for the horizontal θ_h and vertical θ_v directions. All of the gaze data was sorted by the timestamps of the collected gaze samples. The 3D gaze normals are represented as (x, y, z) a unit vector that defines the direction of the gaze in VR world coordinates. In our setup, the x coordinate corresponds to the left-right direction, y in the up-down direction, z in the forward-backward direction. The formulas used for computing the gaze angles are as follows:

$$\theta_\mathrm{h} = \frac{180}{\pi} \arctan \frac{x}{z}$$

$$\theta_\mathrm{v} = \frac{180}{\pi} \arctan \frac{y}{z}$$

Next, we calculated the angular velocity of the eye in both the horizontal and vertical coordinates by taking a first difference of the angular velocity and dividing by the difference between the timestamp of the samples using the formula below:

$$\omega_\mathrm{h} = \Delta\theta_\mathrm{h} / \Delta t$$

$$\omega_\mathrm{v} = \Delta\theta_\mathrm{v} / \Delta t$$

Finally, we calculated the magnitude of the angular velocity (ω) at every timestamp from the horizontal and vertical components using:

$$\omega = \sqrt{\omega_h^2 + \omega_v^2}$$

To classify the fixation and saccade-based samples, we used an adaptive threshold method for saccade detection described by Voloh et al., 2019. We selected an initial saccade velocity threshold θ_0 of 200 $\circ$/sec. All eye movement samples with an angular velocity of less than θ_0 were used to compute a new threshold θ_1. θ_1 was three times the median absolute deviation of the selected samples. If the difference between θ_1 and θ_0 was less than 1 $\circ$/sec θ_1 was selected as the saccade threshold else, θ_1 was used as the new saccade threshold and the above process was repeated. This was done until the difference between θ_n and θ_{n+1} was less than or equal to 1 $\circ$/sec. This way we arrived at the cluster of samples that belonged to fixations and the rest were classified as saccades.

After this, we calculated the duration of the fixations and removed those fixations that had a duration less than 50 ms or were larger than 3.5 times the median absolute deviation of the fixation duration. For further data analysis, we only considered those fixations that were positioned on the 3D tool models. We further categorized the fixations based on their position on the tool, i.e., whether they were located on the effector or handle of the tool.

Data Analysis

Odds of Fixations in favor of tool effector

After cleaning the dataset, we were left with 2174 trials from 18 subjects in experiment-I and 3633 trials from 30 subjects in experiment-II. For both experiments, we analysed the fixations in the 3 second period from the tool presentation till the go cue. For the two experiments, we modeled the linear relationship of the log of odds of fixations on the effector of the tools and the task cue (LIFT, USE), the familiarity of the tool (FAMILIAR, UNFAMILIAR), and orientation of the handle (LEFT, RIGHT). All within-subject effects were also modeled with random intercepts and slopes based on the subjects. We were also interested in modeling the random effects based on the tool to assess the differential effects on the individual tools. We did not have enough data to estimate random item effects, so we fitted a random intercept for the 12 tools.

We used effect coding (Schad et al., 2018) to construct the design matrix for the linear model, where we coded the categorical variables LIFT, FAMILIAR, RIGHT to -0.5 and USE, UNFAMILIAR, LEFT to 0.5. This way, we could directly interpret the regression coefficients as main effects. The model fit was performed using restricted maximum likelihood (REML) estimation (Corbeil and Searle, 1976) using the lme4 package (v1.1-26) in R 3.6.1. We used the L-BFGS-B optimizer to find the best fit using 10000 iterations. Using the Satterthwaite method (Luke, 2017), we approximated degrees of freedom of the fixed effects. For both experiments, the Wilkinson notation (Wilkinson and Rogers, 1973) of the model was:

$$
\begin{aligned}
&log \frac{p(fixations \quad on \quad effector)}{p(fixations \quad on \quad handle)} \sim \\
&1 + task * familiarity * handle\_orientation \\
&\quad + (1 + task * familiarity * handle\_orientation|Subject) + (1|tool)
\end{aligned} \tag{3.1}
$$

As we used effects coding, we can directly compare the regression coefficients of the two models. The fixed-effect regression coefficients of the two models would describe the differences in log-odds of fixations in favor of tool effector for the categorical variables task, familiarity, and handle orientation.

Spatial bias of fixations on the tools

In this analysis, we wanted to assess the effects of task, tool familiarity, and handle orientation on the eccentricity of fixations on the tools. To do this, we studied the fixations from the time when the tool was visible on the table (3s from the start of trial) till the go cue indicated when subjects could start manipulating the tool. We divided this 3s period into 20 equal bins of 150ms each. For each trial and time bin, we calculated the median distance of the fixations from the tool center. Next, we normalized the distance with the length of the tool so that we could compare the fixation eccentricity across different tools.

To find the time-points where there were significant differences for the 3 conditions and their interactions, we used the cluster permutation method. Here, we use the t-statistic as a test statistic for each time-bin, where t is defined as:

$$t = \sqrt{N} * \frac{x}{\sigma}$$

and, x is the mean difference between conditions, and σ is the standard deviation of the mean and N is the number of subjects. We used a threshold for t at 2.14 which corresponds to the t-value at which the p-value is 0.05 in the t-distribution. We first found the time-bins where the t-value was greater than the threshold. Then, we computed the sum of the t-values for these clustered time-bins which gave a single value that represented the mass of the cluster. Next, to assess the significance of the cluster, we permuted all the time-bins across trials and subjects and computed the t-values and cluster mass for 1000 different permutations. This gave us the null distribution over which we compared the cluster mass shown by the real data. We considered the significant clusters to have a p-value less than 0.05. In the results, we report the range of the significant time-bins for the 3 different conditions and their interactions and the corresponding p-values.

3.2.4 Results

The present study investigated the differences in gaze-based strategies dependent on task, tool familiarity, and handle orientation. Here, we investigated two anticipatory gaze-based strategies 3 seconds before action initiation; the odds of fixations in favor of the tool effector and the eccentricity of the fixations through time towards the tool effector. We further compared the differences in two experiments that had the same experimental design but differed in the realism of the action affordance and environment.

First, we were interested in how the participants subjectively assessed the familiarity of the 12 tools. 3.2A shows the subjective familiarity ratings for each of the familiar and unfamiliar tools used in the study. The mean familiarity rating for familiar tools in experiment-I was 4.55 (SD=0.60) and for unfamiliar tools 1.81 (SD=1.17). In experiment-II, the mean familiarity rating for familiar tools was 4.48 (SD=0.52) and for unfamiliar tools 1.56 (SD=1.04). To determine the differences in the subjective familiarity ratings for the two experiments and our categorization of familiarity, we performed a mixed-ANOVA with familiarity as a within-subject factor and the experiment group as the between-subject factor. We found no

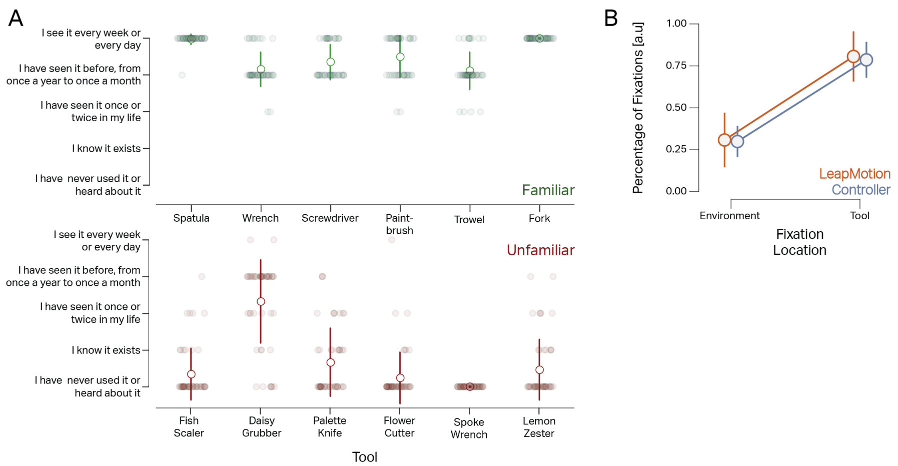

Figure 3.2: **A)** Participants' familiarity rating of the tools. Participants provided their subjective rating of familiarity with the 12 tool stimuli on a 5-point Likert scale. The small circles correspond to ratings from individual subjects. The larger circles correspond to the mean rating for each tool, and error bars represent the standard deviation across subjects. **B)** Percentage of fixations allocated to the environment vs. the tools for the two different experiments. The circles correspond to the mean percentage of fixations across subjects, and the error bars represent the standard deviation. As seen here, the realism of the environment did not affect how participants allocated their attention in the experiments.

differences in the familiarity ratings between the two experiments ($F(44)=3.08$, p-value=0.08). Furthermore, there were significant differences in the subjective rating of the tools ($F(44)=3094.05$, p-value<0.001). There were also no significant interactions between the two factors ($F(44)=2.52$, p-value=0.11). Figure In sum, our experimental condition of familiarity was consistent with the participants' subjective rating as well.

Next, we wanted to make sure that the differences in the virtual environments did not affect the way subjects allocated their attention to the experimental task. We calculated the mean percentage of fixations positioned on the tool vs. anywhere else in the environment for each subject across trials. Figure 3.2B shows the percentage of fixations allocated to the tools vs. the environment for the two experiments. For experiment-I with the interaction method of VR controller and a less realistic environment, the mean percentage of fixations on the environment was 0.29 (SD=0.08) and on the tools 0.78 (SD=0.10). Conversely, in experiment-II with LeapMotion as the interaction method and a more realistic environment, the percentage of fixations allocated to the environment was 0.30 (SD=0.15) and on the tools 0.80 (SD=0.14). To test if these differences were significant, we performed a mixed-ANOVA with fixation location as a within-subject factor, the two experiments as a between-subject factor, and the percentage of fixations as the dependent variable. We found no differences in the percentage of fixations between the two experiments (F(47)=2.86, p-value=0.09). There were significant differences in the percentage of fixations located on the tool vs. the environment (F(47)=217.47, p-value<0.001). We did not find any interactions between the two factors (F(47)=0.02, p-value=0.87). These results show that the allocation of attention was primarily task-oriented and was not affected by the differences in the virtual environment of the two experiments.

Next, we compared the log-odds of fixations in favor of the tool effector across the three conditions: task, tool familiarity, and handle orientation in the 3s period when the subjects studied the tool. Figure 3A shows the log-odds of the fixations on the tool effector for experiment-I (with HTC VIVE Controllers) and experiment-II (with LeapMotion). In experiment-I (Figure 3.3A, left panel), subjects showed a mean log odds of 0.01 (95%CI = [-0.04, 0.08]) for the LIFT task and for the USE task the mean log-odds were 0.19 (95%CI = [0.11, 0.28]). For the FAMILIAR tools, the mean log-odds in favor of the tool effector were -0.16 (95%CI = [-0.23, -0.09]) and for UNFAMILIAR 0.35 (95%CI = [0.25, 0.45]). For the RIGHT oriented tool handle, the mean log-odds were 0.14 (95%CI = [0.06, 0.21]) and for the LEFT oriented tool handle, the mean log-odds were 0.08 (95%CI = [-0.09, 0.26]). To assess the significance of the factors, we used linear mixed models. For the linear model, we used effect coding so the regression coefficients can be directly interpreted as main effects. There was a significant main effect of factor task (USE - LIFT) β = 0.18 (95%CI = [0.08, 0.27], t(70.09)=3.7), with a p-value < 0.001. There was a significant main effect of familiarity (UNFAMILIAR - FAMILIAR) β = 0.58 (95%CI = [0.09, 1.08], t(10.79)=2.33), with p-value = 0.04. The main effect of handle orientation was not significant (LEFT - RIGHT) β = -0.04, (95%CI = [-0.29, 0.21], t(16.94)=-0.32), p-value = 0.75. We found a significant interaction of task and familiarity with β = 0.24 (95%CI = [0.03, 0.45], t(25.88)=2.21), p-value = 0.036. The interaction of task and handle orientation was not significant, β = 0.19 (95%CI = [-0.05, 0.43], t(21.26)=1.55), p-value = 0.13. The interaction of familiarity and orientation was not significant , β = -0.28 (95%CI = [-0.56, -0.004], t(17.07)=-1.99), p-value = 0.06. The 3-way interaction was also not significant, β = -0.005 (95%CI = [-0.37, 0.36], t(1695)=-0.02), p-value=0.97.

In experiment-II (Figure 3.3A, right panel), subjects showed a mean log odds of 0.08 (95%CI = [-0.04, 0.22]) of fixations on the tool effector for the LIFT task and for the USE task the mean log-odds were 0.22 (95%CI = [0.09, 0.36]). For the FAMILIAR tools, the mean log-odds in favor of the tool effector were 0.04 (95%CI = [-0.08, 0.16]) and for UNFAMILIAR 0.25 (95%CI = [0.12, 0.39]). For the RIGHT oriented tool handle, the mean log-odds were 1.18 (95%CI = [1.06, 1.29]) and for the LEFT oriented tool handle, the mean log-odds were -0.32 (95%CI = [-0.45, -0.20]). Using the linear mixed model, we assessed the significance of the three factors. There was a significant main effect of factor task (USE - LIFT) β = 0.13 (95%CI = [0.01, 0.25], t(27.28)=2.13) with a p-value = 0.04. The main effect of familiarity (UNFAMILIAR - FAMILIAR) was not significant β = 0.35 (95%CI = [-0.06, 0.77], t(10.02)=1.67) with p-value = 0.12. The main effect of handle orientation (LEFT - RIGHT) was significant, β = -1.74 (95%CI = [-1.99, -1.48], t(28)=-13.65), p-value <0.001. We found a significant interaction of task and familiarity with β = 0.25 (95%CI = [0.09, 0.41], t(43.64)=3.13), p-value = 0.003. The interaction of task and handle orientation was not significant, β = 0.11 (95%CI = [-0.06, 0.28], t(31.90)=1.29), p-value = 0.20. Similarly, the interaction of familiarity and orientation was significant, β = 0.33 (95%CI = [0.15, 0.50]), p-value = 0.001. The 3-way interaction was also not significant, β = -0.09 (95%CI = [-0.39, 0.19], t(2201)=-0.65), p-value=0.51.

Figure 3.3B summarizes the regression coefficients of the linear model from both experiments. Importantly, we see that the main effect of the task is significant for both experiments. Similarly, the interaction of task and familiarity is significant for both experiments. However, the effect of handle orientation is only significant in experiment-II with the LeapMotion interaction method.

Next, we were interested in the effect of task, tool familiarity, and handle orientation on the eccentricity of the fixations on the tool before action initiation. We calculated the relative distance of fixations from the center of the tool in the 3s period when the subjects studied the tool. We used cluster permutation tests to evaluate the time periods when the effects of the different conditions were significant. As shown in Figure 3.4A, in experiment-I, the differences in task (USE - LIFT) were significant from 1.05s to 1.95s period, p-value<0.001. Differences in tool familiarity (FAMILIAR - UNFAMILIAR) were significant from 0.15s to 3s with a p-value <0.001. Moreover, the differences in the two orientations (LEFT - RIGHT) were not significant. The interaction of task and familiarity were significant from 0.3s to 2.55s, p-value=0.006. The interactions of task and handle orientation, and the interaction of handle orientation and tool familiarity were not significant in any time period.

Similarly, figure 3.4B shows the eccentricity of fixations from the center of the tool during the 3s period when the subjects studied the tool in experiment-II. The differences in task were significant in two time periods 0.45s to 2.4s, p-value< 0.001 and from 2.7s to 3s, p-value=0.03. The differences in familiarity were significant from 0.15s to 3s, p-value < 0.001. Furthermore, the differences in handle orientation were significant from 0.75s to 1.8s, p-value < 0.001. A significant interaction of task and familiarity from 0.45s to 1.5s with p-value <0.001. There were also significant clusters in the interaction of task and handle orientation, from 0.3s to 3s with p-value<0.001 and for tool familiarity and handle orientation from 0.15s to 3s with p-value < 0.001.

3.2.5 Discussion

The primary aim of this study is to investigate how gaze-based strategies vary for tasks, tool familiarity, and manual planning in naturalistic settings. With our study, we successfully added to the current body of research in two important ways. Firstly, irrespective of the realism of the action affordance in virtual environments, the number, and location of anticipatory fixations were modulated by goal-oriented factors of task and tool familiarity. Secondly, anticipatory fixations related to proximal manual planning were only seen when the setup allowed for more realistic action affordances with the virtual hand mimicking finer hand and finger movements. In sum, proximal and distal goal-oriented planning is highly contextualized to the realism of action/interaction afforded by the environment.

We conducted two experiments to disentangle the role of action affordance for goal-oriented planning. Participants interacted with 3D tool models using VR controllers in a low realism setup, which produced a virtual grasp by pulling their index fingers. Here, we showed that the odds of sampling visual information from the mechanical properties of a tool are different based on the specificity of the task. Moreover, given tool familiarity, the odds of fixating on the effector increased for unfamiliar tools. Tool-specific knowledge also played a major role when subjects were instructed to produce tool-specific movements. Moreover, the spatial orientation of the tool did not affect the odds of fixations for the tool effector. In sum, with the preparation of a symbolic grasping action, fixations were affected by distal goal-oriented factors of task and tool familiarity.

In a high realism setup, participants interacted with tool models by producing an actual grasp over the tools. The results were similar to the first experiment. However, we additionally found a significant effect of spatial orientation of the tool where the odds of fixations in favor of the tool effector decreased when the tool was presented incongruent to the subjects' handedness. These results suggest that fixations are directed towards the handle of the tool in anticipation of planning the proximal goal of an optimal grasp. Interestingly, the optimal grasp planning is initiated from the beginning until the end of the viewing time window and might be more critical than inspecting the tool effector to produce the correct action. Taken together, the preparation of a realistic grasping action modulated anticipatory fixations related to both proximal and distal goal planning.

These results are in line with the findings reported by Belardinelli, Barabas, et al., 2016. They investigated behavioral responses to task and tool-based affordances in a lab where subjects responded to stimuli images on a computer screen and pantomimed their manual actions. Moreover, they presented the tools with the handle always oriented on the right and congruent to the subject's handedness. Our results suggest that well before action initiation, subjects had to substantially plan their hand movement on the tool to interact with it. This effect is indicative of an end-state comfort planning (Herbort and Butz, 2012) where both proximal and distal goal-oriented planning interacts to modulate anticipatory fixations. From the perspective of ecological validity, our findings give a fuller view of how different planning strategies are needed to produce relevant action. Our study shows that within a naturalistic setting, task, tool familiarity, and the spatial orientation of the tool affect the planning and production of relevant actions. Hence, our study offers a veridical and ecological valid context to aspects of anticipatory behavior control.

Studies in eye-hand coordination (Belardinelli et al., 2018; Johansson et al., 2001; Lohmann et al., 2019) have shown that eye movements are predictively made towards the grasp contact points. Furthermore, Flanagan et al., 2006 proposed that predictions are made in an event-oriented manner and are at the heart of successful control strategies for object manipulations. They posit that predicted sensory events are compared with actual events like grasping, lifting, moving the object to monitor task progression. In contrast, Iacoboni et al., 2005 and Wohlschläger et al., 2003 showed that goal-oriented planning is specified at an abstract level rather than at the movement level. Our results suggest that the anticipatory gaze behavior specific to task and tool familiarity is seen only when additional grasp control planning is not needed. Inversely, optimal motor control might supersede planning based on other distal goals. Here, we make the case that predictions are made for action outcomes at various scales, and that eye movements are used to plan both optimal grasp control and task-specific requirements well before action initiation.

Our study adds to the growing body of evidence that anticipation and prediction are at the core of cognition (Pezzulo et al., 2007). Motor theories of cognition have proposed that simulations of actions reuse internal models of motor commands to effect multiple predictions (Jeannerod, 2006). The simulation of action theory has been used to explain numerous phenomena of planning, prediction of external events, visual perception, and imitation. Hoffmann, 2003 introduced anticipatory behavior control as the mechanism by which action-effect representations are activated by the need for an effect-related goal and contingent stimuli. Furthermore, Pezzulo et al., 2021 recently proposed that generative models provide top-down predictive signals for perception, cognition, and action during active tasks and these signals are otherwise weak and/or absent when the brain is at rest or the stimuli are weak. Our study shows that anticipatory behavior is tightly linked to the production of task-relevant actions and contextualized to the realism of the action affordance.

Notably, our study shows that different constraints on the method of interaction can also result in different anticipatory behavioral responses. From the perspective of Gibson, 1977, the affordances of the environment are tightly linked to the actions that one can perform in it. Similarly, O'Regan and Noë, 2001 posited that actions constitute the cognitive processes that govern relevant sensorimotor contingencies. In our study, the production of relevant actions significantly modulated the visual sampling of the tool parts in accordance to goal-oriented factors such as task and tool familiarity irrespective of the action affordance. Taken together, our study shows that some aspects of anticipatory gaze are dependent on the realism of the action afforded by the environment.

We conducted the present study in virtual reality, which is still a burgeoning technology for vision research. While VR environments pose an exciting avenue of research, there are still limitations that practitioners must face while conducting experiments in these scenarios. First, the naturalistic setting of both experiments I and II afforded natural head movements. To maintain optimal quality over the data, we asked the participants in the study to make limited head movements. Additionally, we presented the tools and the task cues not to cause extreme pitch head movements. Secondly, mobile eye-trackers can be error-prone and might suffer from variable sampling rates (Ehinger et al., 2019) or calibration errors due to slippage (Niehorster et al., 2020). To mitigate any calibration errors, we also made sure that we calibrated the eye-trackers at regular intervals. Thirdly, both controller-based and camera-based VR interaction methods are still new technology. It could have been challenging for participants to get used to, even though we made sure they practiced the interaction method before the experiment. While we simulated grasping the tool using LeapMotion's gesture recognition and were able to produce a more realistic action affordance through mimicking finer hand and finger movements, it is still an inadequate substitute for a real grasp where the tactile feedback of the tool in hand might elicit more accurate responses. For example,Ozana et al., 2018 showed that grasping movements within a virtual environment differ both quantitatively and qualitatively from typical grasping. Lastly, there are obvious differences in the realism of the two virtual environments used in the study in terms of the visual scene. While there are visible differences between the environments, we see that there are no significant differences between the percentage of fixations allocated to the background vs. tool for both experimental settings. Hence, we contend that the differences in the eye movement behavior reported in the study are largely a consequence of the differences in the action affordance and much less because of mere visual differences. In light of these limitations, we know that our study must be considered from a nuanced perspective. Furthermore, there is still room for replicating our study with novel and more realistic interaction methods.

There are still some open questions pertaining to anticipatory behavior elicited by tool interactions. Firstly, while our study distinguishes between levels of action affordances, future work can look at goal-oriented planning for passive observers at both proximal and distal levels. Secondly, it would be interesting to dive deeper into the predictive brain signals that give rise to the present oculomotor behaviors. Our study provides a first step towards distinctly investigating proximal and distal goal-oriented planning.

3.2.6 Conclusion

The present study gives a veridical and ecologically valid context to planning and anticipatory behavior. Our results support the hypothesis that eye movements serve the cognitive function of actively sampling information from the environment to produce relevant actions. When semantic information about the object is not readily available, eye movements are used to seek information from its mechanical properties from specific locations. Furthermore, we show that fixations are made in a goal-oriented way in anticipation of the relevant action. When considering the realism of the action affordance, our results show that eye movements prioritize proximal goals of optimal grasp over task-based demands. Lastly, our study is at the frontiers of naturalistic vision research, where novel technologies can be harnessed to answer questions that were previously far-fetched.

Author Contributions

AK, PK: conceived and designed the study. TS, PK: Procurement of funding. NG, AK, FNN: programmed the controller study. SB, AK, FNN: programmed the LeapMotion study. NG, SB: data collection. AK: data analysis. AK: initial draft of the manuscript. AK, SB, NG, FNN, TS, PK: revision and finalizing the manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgement

We are grateful for the financial support by: the German Federal Ministry of Education and Research for the project ErgoVR (Entwicklung eines Ergonomie-Analyse-Tools in der virtuellen Realität zur Planung von Arbeitsplätzen in der industriellen Fertigung)-16SV8052; German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) project number GRK2340/1 (DFG Research Training Group Computational Cognition).

Data Accessibility

The authors have made the data associated with this study available on request.

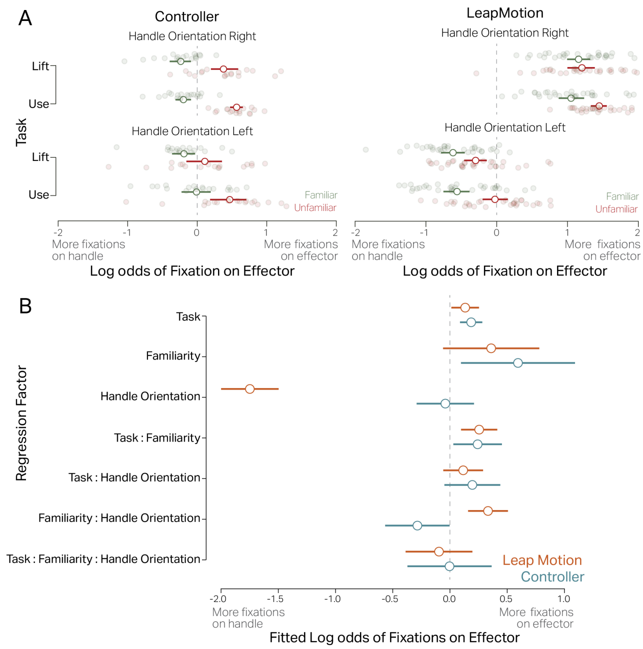

Figure 3.3: Experiment results. **A) top-left:** shows the log-odds of fixation on effector vs. handle in the controller study when the tool handle is oriented to the right. The log odds on fixations are higher on the effector for unfamiliar tools (red) than the familiar tools (green) for both the LIFT and the USE tasks. **Bottom-left:** log odds of fixation on effector when the tool handle is oriented to the left and is incongruent to the subjects' handedness. The plot shows that the orientation of the tool does not significantly affect the log-odds fixation on the effector. **Top-right:** the log-odds of fixation on effector in the LeapMotion study when the tool handle is oriented to the right. The log odds of fixations on the effector are higher for unfamiliar tools (red) than the familiar tools (green) and the USE task. **Bottom-right:** log odds of fixation on effector when the tool handle is oriented to the left and is incongruent to the subjects' handedness. The plot shows that the orientation of the tool results in significant log-odds of fixations over the handle in the LIFT task, while in the USE task and with unfamiliar tools (red) significantly more fixations were on the effector. **B)** The linear regression coefficients for the two experiments. The effect of the task is significant for both experiments with higher log-odds of fixations on the effector. For the factor orientation, the log-odds are significant in the LeapMotion experiment and not for the controller experiment. Similarly, the interaction between task and familiarity is significant for both experiments.

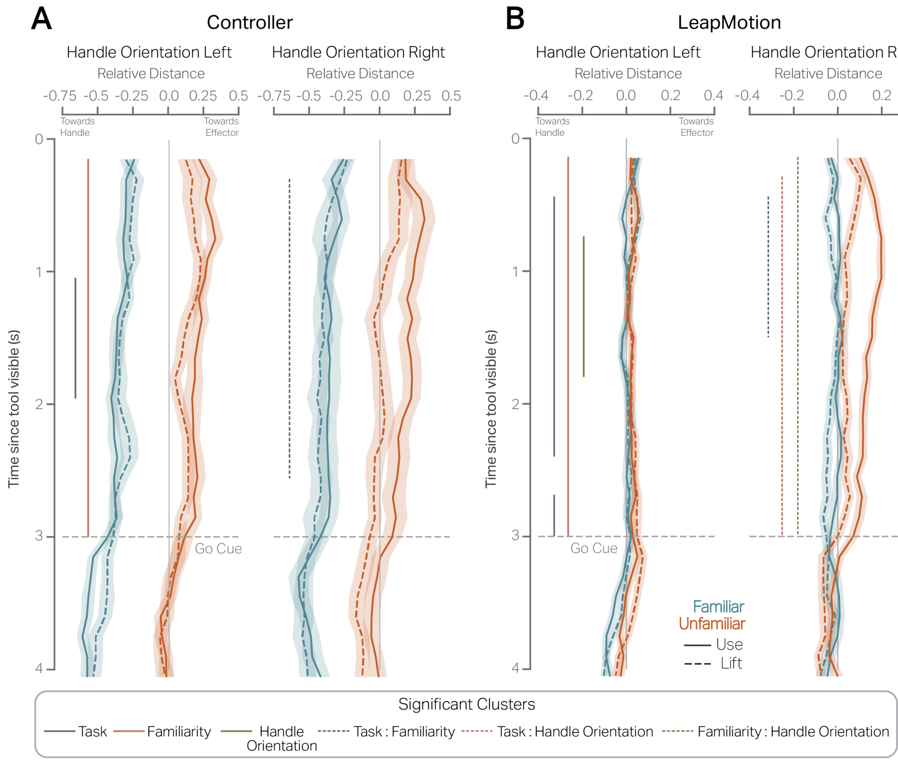

Figure 3.4: Eccentricity of fixations on the tool models. The negative values of the abscissa correspond to fixations towards the handle, the positive values refer to fixations towards the tool effector, and zero represents the center of the tool. The ordinate axis refers to the time elapsed since the tool is visible on the virtual table. The go cue is given to participants at 3s after which they can start interacting with the tool. The blue lines correspond to the FAMILIAR tool and red to the UNFAMILIAR tools. The error bars represent the standard error of the mean across subjects. The vertical solid lines correspond to the significant time clusters for main effects and the vertical dashed lines to the interactions. **Panel A** shows the findings from experiment-**I** and the two handle orientations. **Panel B** shows the findings from experiment-**II** and the two handle orientations.

3.3 Task Complexity Affects Gaze Guidance Behavior while Action Planning and Execution in Naturalistic VR

This section was submitted as a peer reviewed article together with Ashima Keshava, Henri Neumann, Krzysztof Izdebski, Thomas Schüler, and Peter König. See Publication List for details.

3.3.1 Abstract

Eye movements in the natural environment have primarily been studied for over-learned everyday activities such as tea-making, sandwich making, driving that have a fixed sequence of actions associated with them. These studies indicate an interplay of low-level action schemas that facilitate task completion. However, it is unclear if this strategy is also in play when the task is novel and a sequence of actions must be planned in the moment. To study attention mechanisms in a novel task in a natural environment, we recorded gaze and body movement data in a virtual environment while subjects performed a sorting task where they sorted objects on a life-size shelf based on some object features. To study the action planning and execution related gaze guidance behavior we also controlled the complexity of the sorting task by introducing EASY and HARD tasks. We show that subjects are close to optimal while performing EASY trials and are more sub-optimal while performing HARD tasks. Fixations aligned with action onset show task complexity elicits greater proportion of look-ahead, and monitoring fixations but not directing and guiding fixations. Task complexity affected the scan-paths on the task-relevant regions of interest during action planning and execution where subjects exhibit a greater search and action monitoring behaviors in HARD tasks and less so in EASY tasks. Task complexity also affected the temporal sequence of first fixations on the task-relevant regions of interest systematically for action planning but not for action execution. Our findings show that task complexity modulates the competition of low-level cognitive schemas during planning and execution even when sub-optimal decisions are made by the actor.

3.3.2 Introduction

In a pragmatic turn in cognitive science, there is a more significant push towards incorporating the study of cognitive processes while interacting with the external world (Parada & Rossi, 2020). Moreover, Engel et al. (2013) proposed that cognition encompasses the body, and in turn, bodily action can be used to infer cognition. To this effect, understanding the control of eye movements in natural, real-life situations requires a mobile setup that allows for a subject to be recorded in tandem with voluntary actions in a controlled yet unconstrained environment. Studying eye movements in mobile subjects might give us a richer picture of cognitive processing in more naturalistic settings.

Humans actively use vision during everyday activities to gather and refine information about the environment. Since the seminal works of Yarbus (2013) and Buswell (1935) there has been consistent evidence that eye movements depend on the viewing task the observer is performing. Kollmorgen et al. (2010) demonstrated stimulus-dependent features, spatial viewing biases, and task-dependent features all influence the exploration of a visual scene. This is further supported by studies that emphasize the relevance of semantics in the guidance of eye movements (Einhäuser et al., 2008; Henderson & Hayes, 2017). Thus, we have growing evidence that task demands can affect eye movements behavior.

Seminal studies have already investigated eye movement behavior in natural environment with fully mobile participants. In the pioneering studies of M. Land et al. (1999) and Hayhoe et al. (2003), subjects performed everyday activities of tea-making and sandwich-making, respectively. These studies required a sequence of actions that involved manipulating objects one at a time to achieve the goal. Both studies showed that nearly all the fixations were task-related. Further studies investigated eye movements under a plethora of natural conditions while walking (Matthis et al., 2018), driving (Mars & Navarro, 2012; Navarro et al., 2020; B. T. Sullivan et al., 2012), hand-washing (Pelz & Canosa, 2001), hitting a ball (M. F. Land & McLeod, 2000), and free exploration (Schumann et al., 2008). These experiments in naturalistic settings have revealed several distinct functions of the eye movements during habitual everyday tasks.

Studies investigating habitual tasks uncovered a systematic timing of visual fixations and object manipulation. Specifically, fixations are made to target objects about 600ms before manipulation. More importantly, Ballard et al. (1995) proposed a "just-in-time" strategy that universally applies to this relative timing of fixations and actions. In other words, fixations that provide information for a particular action immediately precede that action and are crucial for the fast and economical execution of the task.

While performing habitual tasks, fixations have been broadly categorized into four functional groups (M. F. Land & Hayhoe, 2001). 'Locating' fixations retrieve visual information. 'Directing' fixations acquire the position information of an object and accompany a manipulation action and facilitate reaching movements. 'Guiding' fixations alternate between two objects being manipulated e.g., knife, bread, and butter. 'Checking' fixations monitor where the task constraints in the scene have been met. These findings have also been corroborated by Pelz and Canosa (2001) and Mennie et al. (2007). Pelz and Canosa (2001) showed similar just-in-time strategy of gaze allocation while performing a hand-washing task. They also reported a small number of fixations of about 5% that did not serve the immediate sub-task but rather provided information that would be needed for a future action. The authors hypothesize these ´Look-ahead´ fixations provide a mechanism to stabilize the visual input stream that result from a sequence of actions, facilitate task-switching, and reduce conscious effort required to complete the actions in a sequence. Hence, look-ahead fixations can be explained as a perceptual strategy to ease the cognitive load attending to complex tasks in the real world. In sum, the wide-ranging functions of eye movements are well documented in natural routine tasks.

Based on these observations M. F. Land and Hayhoe (2001) proposed a framework that outlines that flow of visual and motor control during task execution Figure 3.5A. The process summarizes the various operations that *must* occur during an 'object-related action' i.e., individual actions performed on separate objects to achieve the desired goal. Each schema "specifies the object to be dealt with, the action to be performed on it, and the monitoring required to establish that the action has been satisfactorily completed." (M. F. Land, 2006). Further, the gaze control samples the information about the location and identity of the object and directs the hands to it. Subsequently, the motor control system of the arms and hands implement the desired actions. Here, vision provides the information of where to direct the body, which state to monitor, and determine when the action must be terminated. Taken together, a 'script' of instructions is sequentially implemented where the eye movements earmark the task-relevant locations in the environment that demand attentional resources for that action.

This matches the common theme in the above studies investigating natural tasks (e.g., tea-making, sandwich-making, hand-washing) with an organized and well-known structure. These tasks involve specific object-related actions such as picking up the knife, picking up the teapot, etc. and have a predefined 'script' for the execution of the tasks. The studies, therefore, study eye movements that are under strict control of a task sequence. Moreover, these tasks are over-learned and over-generalized as they are part of a habitual action repertoire for an adult human being. As discussed by M. F. Land (2006) the low-level schemas (locating, directing, guiding, monitoring) defined above are likely not executed under deliberate conscious control. This distinction corresponds to James (2007) distinction between "ideo-motor" and "willed" acts. As James described, ideo-motor actions correspond to movements where we are "aware of nothing between the conception and execution" of the said action. In contrast, the willed actions require "an additional conscious element in the shape of a fiat, mandate, or expressed consent." Hence, it is unclear whether these low-level schemas of gaze control operate similarly for deliberate actions where an internal task script is not already known.

Norman and Shallice (1986) proposed a theoretical framework for the components of attentional mechanisms that govern deliberate/planned actions. In comparison to the low-level schema proposed by M. F. Land (2006), which can account for routine, well-learned tasks, the Norman and Shallice (1986) model suggests another supervisory module that selects a schema to implement. In well-learned tasks, a schema is triggered automatically without conscious control. However, when a task is fairly complex and requires planning, multiple low-level schemas might compete for resources at the same time and require contention scheduling. For example, contentions can arise on whether to monitor the current action with respect to previous actions or future planned actions to fulfill the task-relevant goals. Such a scheduling mechanism is then required to provide conflict resolution for potentially relevant schemas either by inhibition or activation. Taken together, the model predicts that a failure of the supervisory control can lead to an instability of attention and heightened distraction.

To generalize the above oculomotor behaviors, one could examine the spatial temporal profiles of the fixation in novel task scenarios. First, in cognitively complex tasks an abundance of look-ahead saccades would give evidence of elaborate cognitive planning. That is, high-level planning processes with matching eye movements would also support optimal decision making. Second, the concurrence of cognitive processing and actions would emphasize a strict sequence of fixations with specific purposes, e.g., locating, directing, guiding, checking where cognitive schemas and actions would evolve in parallel. Experiments with complex, variable tasks are needed to differentiate these hypotheses.

To further pursue this line of research, it is desirable to perform such experiments under tightly controlled laboratory conditions. In recent years, virtual reality (VR) and mobile sensing has offered great opportunity to create controlled, natural environments. Here, subjects' eye and body movements can be measured reliably along with their interactions with the environment (Clay et al., 2019; Keshava et al., 2020; Keshava et al., 2021; Mann et al., 2019). Experiments in virtual environments have grown popular in recent years and have shown promise towards studying cognition in naturalistic and controlled environments.

In the present study, we investigate the mechanisms of allocation of attention while performing a novel task in a naturalistic environment. We created two types of tasks that varied in complexity and required performing a sequence of actions to accomplish the cued goal. We asked subjects to sort objects on a life-size shelf based on the object features. The complexity of the task depended on sorting based on one object feature or both. We designed the tasks to be novel in a way that subjects had to plan their action sequences on-the-fly and in absence of a pre-defined action "script". We concurrently measured the eye and body movements while subjects performed the tasks.

3.3.3 Methods

Participants

A total of 60 participants (39 females, mean age = 23.9 ± 4.6 years) were recruited from the University of Osnabrück and the University of Applied Sciences Osnabrück. Participants had a normal or corrected-to-normal vision and no history of neurological or psychological impairments. They either received a monetary reward of €7.50 or one participation credit per hour. Before each experimental session, subjects gave their informed consent in writing. They also filled out a questionnaire regarding their medical history to ascertain they did not suffer from any disorder/impairments which could affect them in the virtual environment. Once we obtained their informed consent, we briefed them on the experimental setup and task. The Ethics Committee of the University of Osnabrück approved the study.

Apparatus & Procedure

For the experiment, we used an HTC Vive Pro Eye head-mounted display (HMD)(110° field of view, 90Hz, resolution 1080 x 1200 px per eye) with a built-in Tobii eye-tracker[VII]. Participants used an HTC Vive controller to manipulate the objects during the experiment with their right hand. The HTC Vive Lighthouse tracking system provided positional and rotational tracking and was calibrated for 4m x 4m space. For calibration of the gaze parameters, we used 5-point calibration function provided by the SRanipal SDK. To make sure the calibration error was less than $1°$, we performed a 5-point validation after each calibration. Due to the study design, which allowed a lot of natural body movements, the eye tracker was calibrated repeatedly during the experiment after every 3 trials. Furthermore, subjects were fitted with HTC Vive trackers on both ankles, both elbows and, one on the midriff. The body trackers were also calibrated subsequently to give a reliable pose estimation using inverse kinematics of the subject in the virtual environment. We designed the experiment using the Unity3D[VIII] 2018.x.x (version) and SteamVR game engine and and controlled the eye-tracking data recording using HTC VIVE Eye Tracking SDK SRanipal[IX] (v1.1.0.1)

The experimental setup consisted of 16 different objects placed on a shelf of 5x5 grid. The objects were differentiated based on two features: color and shape. We used four high contrast colors (red, blue, green and yellow) and four 3D shapes (cube, sphere, pyramid and cylinder). The objects had an average height of 20cm and width of 20cm. The shelf was designed with a height and width of 2m with 5 rows and columns of equal height, width and, depth. Participants were presented with a display board on the right side of the shelf where the trial instructions were displayed. Subjects were also presented with a red buzzer that they could use to end the trial once they finished the task.

[VII] https://enterprise.vive.com/us/product/vive-pro-eye-office/
[VIII] Unity, www.unity.com
[IX] SRanipal, developer.vive.com/resources/vive-sense/sdk/vive-eye-tracking-sdk-sranipal/

Experimental Task

Subjects performed two practice trials where they familiarized themselves with handling the VR controller and the general aspects of the setup. In these practice trials they were free to explore the virtual environment and handle the objects. After the practice trials, subjects were asked to sort object based on the one and/or two features of the object. There were two types of trials: EASY and HARD. Subjects were not limited by time to complete the task. Each subject performed 24 trials with each trial type (as listed below) randomly presented twice throughout the experiment. The experimental setup is illustrated in Figure 3.5B. The EASY trials instructions were as follows:

1. Sort objects so that each row has the same shape or is empty
2. Sort objects so that each row has all unique shapes or is empty
3. Sort objects so that each row has the same color or is empty
4. Sort objects so that each row has all unique colors or is empty
5. Sort objects so that each column has the same shape or is empty
6. Sort objects so that each column has all unique shapes or is empty
7. Sort objects so that each column has the same color or is empty
8. Sort objects so that each column has all unique colors or is empty

The HARD trials instructions were as follows:

1. Sort objects so that each row has all the unique colors and all the unique shapes once
2. Sort objects so that each column has all the unique colors and all the unique shapes once
3. Sort objects so that each row and column has each of the four colors once.
4. Sort objects so that each row and column has each of the four shapes once.

3.3.4 Data pre-processing

Gaze Data

The data preprocessing steps are illustrated in Figure 3.5C. As a first step, using eye-in-head 3d gaze direction vector for the cyclopean eye we calculated the gaze angles for the horizontal θ_{h} and vertical θ_{v} directions. All of the gaze data was sorted by the timestamps of the collected gaze samples. The 3d gaze direction vector of each sample is represented in (x, y, z) coordinates as a unit vector that defines the direction of the gaze in VR world space coordinates. In our setup, the x coordinate corresponds to the left-right direction, y in the up-down direction, z in the forward-backward direction. The formulas used for computing the gaze angles are as follows:

$$\theta_{\mathrm{h}} = \frac{180}{\pi} * \arctan\frac{x}{z}$$

$$\theta_{\mathrm{v}} = \frac{180}{\pi} * \arctan\frac{y}{z}$$

Next, we calculated the angular velocity of the eye in both the horizontal and vertical coordinates by taking a first difference of the angular velocity and dividing by the difference between the timestamp of the samples using the formula below:

$$\omega_{\mathrm{h}} = \Delta\theta_{\mathrm{h}} / \Delta t$$

$$\omega_{\mathrm{v}} = \Delta\theta_{\mathrm{v}} / \Delta t$$

Finally, we calculated the magnitude of the angular velocity (ω) at every timestamp from the horizontal and vertical components using:

$$\omega = \sqrt{\omega_h^2 + \omega_v^2}$$

To filter the samples where gaze was relatively stable, we used an adaptive threshold method for saccade detection described by Voloh et al. (2019). We selected an initial saccade velocity threshold θ_0 of 200 $\circ$/sec. All eye movement samples with an angular velocity of less than θ_0 were used to compute a new threshold

θ_1. θ_1 was three times the median absolute deviation of the selected samples. If the difference between θ_1 and θ_0 was less than 1 ∘/sec θ_1 was selected as the saccade threshold else, θ_1 was used as the new saccade threshold and the above process was repeated. This was done until the difference between θ_n and $\theta_{\mathrm{n}+1}$ was less than or equal to 1 ∘/sec. This way we arrived at the cluster of samples that belonged to fixations and the rest were classified as saccades. After this, we calculated the duration of the fixations and saccades. To handle miniscule fixations and saccades, we labeled all samples with saccade duration less than 0.03 seconds as a fixation. We also labeled all fixation samples with duration of less than 0.05 seconds as saccades. Following this, we recalculated the fixation and saccade durations. Finally, we rejected all fixations with duration greater than 3.5 times the median absolute deviation of the sample fixation duration as well as fixations that were less than 0.1 seconds long.

Grasp data

Subjects used the trigger button of the HTC vive controller to virtually grasp the objects on the shelf and displace them to other locations. In the data, the trigger was recorded as a boolean which was set to TRUE when a grasp was initiated and was reset to FALSE when the grasp ended. Using the position of the controller in the world space, we determined the locations from the shelf where a grasp was initiated and ended. We also removed trials where the controller data was showed implausible locations in the world space. These faulty data can be attributed to loss of tracking during the experiment. Next, we removed grasping periods where the beginning and final locations of the objects on the shelf were the same. We calculated the inter-quartile range (IQR) of the participants object displacement behavior for the two trial types (EASY and HARD). To remove the outlying object displacements in the trials, we removed the trials with 1.5 times the IQR of object displacements. We also removed those trials with fewer than three object displacements.

3.3.5 Data Analysis

In order to study the function of eye movements for both action planning and execution, we divided each trial into 2 types of epochs. The action execution epoch spanned the time from start of object displacement to the end. The action planning epochs started from end of previous object displacement to start of current object displacement. The schematic of this epoch creation is illustrated in Figure 3.5D. This division of time within each trial into separate epochs allows us to parse the role of overt eye movements in planning and execution of object related actions separately.

For the action planning and execution epochs, we examined the spatial and temporal characteristics of eye movements while performing the sorting tasks. We divided the object and shelf locations into 7 regions-of-interest (ROIs) comprising of previous, current, and next target object and target shelf. More specifically, the previous target object refers to the object that was handled in the previous action epoch, and previous target shelf as the shelf where the previous target object was placed. Similarly, the current target object refers to the object that is picked up and placed on the target shelf in the current epoch and the next target object and next target shelf in the immediately following epoch. All other regions which did not conform to the above 6 ROIs are categorized as 'other' and not relevant to the action sequence. As we need at least 3 object related actions within a trial to form the ROIs for the action planning and action execution epochs, we removed trials where subjects made fewer than three object displacements. In this format, we could parse the sequence of eye movements on the seven ROIs that are relevant for planning and execution of the object related actions.

Task-based behavioral differences

In order to assess the planning behavior of the participants, we determined the optimal object displacements required to accomplish the tasks for the two trial type. To determine the optimal object displacements we designed a depth-first search algorithm that computed the minimum number of displacements required to sort the objects for the 5000 random initial configurations of 16 objects in 25 shelf locations for both EASY and HARD trial constraints. We compared the mean number of object displacements made by the participants in the EASY and HARD trials with the model based object displacements using independent t-tests.

Action Locked Gaze Control

We were interested in the average fixation behavior time-locked to action initiation. For each grasp onset in a trial we chose the time period from 2 seconds before grasp onset and 2 seconds after. We divided this 4 second period into bins of 0.15 seconds and calculated the number of fixations on the seven ROIs described above. For each time bin, we calculated the proportion of fixations on the ROIs per trial type (EASY, HARD). To find the time-points where there were significant differences between EASY and HARD trials for a given ROI, we used the cluster permutation method. Here, we use the t-statistic as a test statistic for each time-bin, where t is defined as:

$$t = \sqrt{N} * \frac{x}{\sigma}$$

and, x is the mean difference between the trial types, and σ is the standard deviation of the mean and N is the number of subjects. We used a threshold for t at 2.14 which corresponds to the t-value at which the p-value is 0.05 in the t-distribution. We first found the time-bins where the t-value was greater than the threshold. Then, we computed the sum of the t-values for these clustered time-bins which gave a single value that represented the mass of the cluster. Next, to assess the significance of the cluster, we permuted all the time-bins across trials and subjects and computed the t-values and cluster mass for 1000 different permutations. This gave us the null distribution over which we compared the cluster mass shown by the real data. To account for the multiple independent comparisons for the seven ROIs, we considered the significant clusters to have a Bonferroni corrected p-value less than 0.007 (0.05/7). In the results, we report the range of the significant time-bins for the seven ROIs for the two trial types the corresponding p-values.

Spatio-temporal Gaze Control in Action Planning and Execution

To compute the scan paths within the action planning and execution epochs we created transition matrices that show the origin and destination locations of the fixations on the 7 ROIs. We used the steps described by Hooge and Camps (2013) to first create the scan paths and then the transition matrices. We calculated the transition matrices summarizing gaze transitions from and to the 7 ROIs from the action planning and execution epochs for each object displacement. Using the transition matrices, we calculated the net and total transitions from and to each ROI. For every transition matrix 'A' per trial, net and total transition are defined as follows:

$$A_{net} = A - A^T \tag{3.2}$$

$$A_{total} = A + A^T \tag{3.3}$$

As discussed in Hooge and Camps (2013), if subjects make equal number of transitions between all ROIs, we can expect no transitions in the net transition matrix and can surmise that the gaze was allocated more randomly. Conversely, with strong gaze guidance we would expect more net transitions. Hence, using the net and total transitions per trial, we then calculated the relative net transitions as:

$$Relative_{Transitions} = \frac{\sum A_{net}}{\sum A_{total}} \tag{3.4}$$

We then took the mean of the relative transitions per trial as a measure of gaze guidance in that trial. Higher mean relative transitions would indicate gaze allocated to ROIs in a systematic manner whereas, relative transitions would represent a random gaze allocation towards the ROIs.

Further, we also calculated the time required to first fixation on the 7 ROIs in a given planning or execution epoch. We then took the median time to first fixation per trial that would indicate time to first fixate on the ROIs for 50% of the action planning and action execution epochs. This method was used by Montfoort et al. (2007) and further applied by Hooge and Camps (2013) to capture the gaze attraction power of ROIs. As the action planning and execution epochs varied in duration, we normalized the time points by dividing them by the duration of the epoch. This way, time elapsed since start of an epoch is comparable to all epochs across trials and subjects.

Linear Mixed Effects Models

We modelled the linear relationship of the relative net transitions dependent on trial type (EASY, HARD), epoch type (planning, execution) and number of object displacements and their interactions. All within-subject effects were modeled with random intercept and slopes grouped by subject. The categorical variables trial type and epoch type were effect coded (Schad et al., 2018), so that the model coefficients could be interpreted as main effects. The object displacement variable which pertained to the number of object displacements in the trial were coded as a continuous numeric variable and centered on zero mean. The model fit was performed using restricted maximum likelihood (REML) estimation (Corbeil & Searle, 1976) using the lme4 package (v1.1-26) in R 3.6.1. We used the L-BFGS-B optimizer to find the best fit using 20000 iterations. Using the Satterthwaite method (Luke, 2017), we approximated degrees of freedom of the fixed effects. The full model in Wilkinson notation (Wilkinson & Rogers, 1973) is defined as:

$$Relative_{Transitions} \sim 1 + trial\_type * epoch\_type * object\_displacements \tag{3.5}$$

$$+(1 + trial\_type * epoch\_type * object\_displacements|Subject) \tag{3.6}$$

We modelled the linear relationship of the median time to first fixation dependent on trial type (EASY, HARD) and the 7 ROIs and their interactions. We computed two models for the action planning and execution epochs as the. All within-subject effects were modeled with random intercept and slopes grouped by subject. The categorical variables trial_type and ROI were effect coded, so that the model coefficients could be interpreted as main effects. For both models, we chose the latency of the first fixation on current target object as the reference factor so that the latency of the first fixation of all other ROIs could be compared to it. The model fit was performed using restricted maximum likelihood (REML) estimation (Corbeil & Searle, 1976) using the lme4 package (v1.1-26) in R 3.6.1. We used the L-BFGS-B optimizer to find the best fit using 20000 iterations. Using the Satterthwaite method (Luke, 2017), we approximated degrees of freedom of the fixed effects. The full model in Wilkinson notation (Wilkinson & Rogers, 1973) is defined as:

$$Fixation_{time} \sim 1 + trial\_type * ROI \tag{3.7}$$
$$+(1 + trial\_type * ROI|Subject) \tag{3.8}$$

3.3.6 Results

Our experiment measured the eye and body movements as participants performed a sorting task in a virtual environment. The participants sorted objects based on the color and/or shape where we modulated the task complexity into EASY and HARD trials. We further divided the trials into planning and execution epochs where participants planned the selection of the target objects to grasp and then executing the action of displacing it to target shelves, respectively. In this section, we report the behavioral and oculomotor differences of the subjects for the two task types (EASY, HARD), and the planning and execution epochs.

Task based Behavioral Differences

In the present study, the primary object related action was to repeatedly pickup objects and place them at a desired locations until they were sorted according to the sorting task. To account for the behavioral differences between the task complexities, we used the measure of trial duration and the number of object displacements required to completed the tasks. Figure 3.6A shows the differences in EASY and HARD trials based on the time taken to finish the sorting task. A two-sample independent t-test showed that the trial duration for the two trial types were significantly different ($t = -10.13, p < 0.001$) where EASY trials that required the objects to be sorted based on a single feature were shorter ($Mean = 54.12 seconds, SD = 13.33$) as compared to HARD trials ($Mean = 111.57 seconds, SD = 36.92$) where subjects had to sort taking into account both features (color and shape) of the objects.

Figure 3.6B shows the comparisons in the object displacements made by the subjects and the optimal number of displacements as elicited by the optimal model for both EASY and HARD trials. Subjects made lower number of object displacements in the EASY trials ($Mean = 10.2, SD = 1.99$) compared to HARD trials

($Mean = 15.52, SD = 2.59$). In the EASY trials the model required lower number of object displacements ($Mean = 9.42, SD = 1.48$) displacements, whereas, in the HARD trials, model required higher number of displacements ($Mean = 11.24, SD = 2.77$). We compared the human and model performance for the two trial types using independent t-tests. In the EASY trials, there was a significant difference between the model and human object displacements ($t = 3.64, p < 0.001$). In the HARD trials, there was also a significant difference between the model and human performance ($t = 10,61, p < 0.001$). This indicates that the participants did not plan their actions optimally and might have used other heuristics to complete the task.

o check if there were any noticeable heuristics applied by the participants, we looked into the propensity to pick-up and drop-off objects to preferred locations on the shelf. In Figure 3.6C shows that subjects preferred to pickup objects from the right-most column and bottom-most row of the shelf. Figure 3.6D shows that subjects had a propensity to drop the objects leaving out the right column and bottom row of the shelf for both EASY and HARD trials. Given the sorting tasks where subjects were presented with random initial configurations of the objects on the shelf locations, we did not expect any systematic spatial biases at play. Further, the expectation was that the subjects would move the objects randomly and not display a preference for object pickup and drop-off locations. This shows that subjects systematically, displace the objects leftward and upward employing an arbitrary heuristic to complete both task types. As the objects are instantiated on the shelf randomly, an optimal strategy would not show this behavior. We can conclude from the above that subjects offset their cognitive load of optimally completing the task by employing simple heuristics. In other words, in lieu of optimally performing the task and finishing it in a shorter time, subjects preferred to offload both cognitive effort on the environment by adopting a more sub-optimal strategy..

3.3.7 Action Locked Gaze Control

We investigated the task complexity based differences in the the average oculomotor control over the seven ROIs time locked to the grasp onset (time when the hand makes contact with the current target object). For the analysis we chose the time period from 2s before action onset to 2s after. Figure 3.7 shows the time course of proportion of fixations on the seven ROIs as described above in section 3.3.5 for the two trial types. The cluster permutation analysis of the time course over the ROIs for the EASY and HARD tasks revealed several time periods where the proportion of fixations were different.

The proportion of fixations on the previous target object differed between EASY and HARD trials for three time periods. The first time period spanned from -1.25s to -0.75s (p-value<0.001) with lower proportion of fixations in the HARD tasks indicating allocation of gaze to other ROIs task-relevant to the action sequence. The second significant time period was from -0.25s to 1s (p-value<0.001) where the proportion of fixations were higher in the HARD tasks. This time period spans the start of the grasp onset till the execution of the object displacement strongly indicating that these fixations are related to monitoring of the execution of the current action and could be classified as 'checking' fixations. The third significant time period was from 1.5s to 2s (p-value<0.001) with higher proportion of fixations in the HARD trials. These differences might be constitutive of further 'checking' fixations in the case of HARD trials while executing the object displacement.

The proportion of fixations on the previous target shelf were lower in the HARD trials compared to the EASY trials from -1.75s to -1.5s (p-value<0.001) before grasp onset. These differences suggest allocation of gaze to ROIs relevant to the task sequence happens earlier in the HARD tasks as more planning is required.

There were differences in the proportion of fixations on the current target object from 1s to 1.5s (p-value < 0.001) after grasp onset with lower proportion of fixations in the HARD tasks in this time period. These differences indicate that towards the end of action execution in HARD tasks fixations are allocated towards other task-relevant ROIs in the action sequence. Interestingly, there are no differences in the proportion of fixations on the current target object before the object displacement is initiated, suggesting that task complexity does not play a role in 'directing' fixations.

There were higher proportion of fixations on the current target shelf in the HARD trials compared to EASY trials from -1.75s to -1s (p-value < 0.001) before grasp onset. These differences imply that some proportion of fixations are utilized to plan the current task, well before action has been initiated. These fixations could be classified as 'locating' fixations or look-ahead fixations which are predominantly present due to the complexity of the task.

Similarly, there were a higher proportion of fixations on the next target object from

-1.75s to -0.25s (p-value < 0.001) and a lower proportion from 1.75 s to 2s (p-value < 0.001). The differences in the first time period suggest that these fixations serve as 'locating' fixations to execute the next object displacement in the sequence. The presence of higher proportion of this fixations in the HARD tasks compared to EASY tasks indicates a need to plan the actions more thoroughly. The differences in the latter time period show a lower proportion of fixations for the HARD tasks indicating that prior locating fixations made it unnecessary to allocate attention to that object of interest. Conversely, in the EASY trials, the lower proportion of these locating fixations indicate an ad hoc gaze allocation for action execution.

There were also a higher proportion of fixations on the next target shelf from -0.5s to 0.75s (p-value < 0.001) in the HARD tasks. These fixations are made in concert with the onset of the action execution and indicate that these fixations are a play the role of both 'checking' fixations to monitor the task progression as well as 'locating' fixations to queue the locations in the scene that are important for the next action in the sequence.

Finally, the proportion of fixations on the other objects and shelves in the scene were higher in the HARD tasks from 1s to 2s (p-value < 0.001) after grasp onset. These differences indicate search behavior towards the end of the current task execution. Given the task complexity of the HARD trials, this search behavior might function to queue in further objects or shelves of interest in the subsequent action sequence.

Spatio-temporal Gaze Control in Action Planning and Execution

The above results illustrate the average spatial and temporal aspects of attention during action planning and execution. However, the scanning behavior of subjects while they perform each action is "averaged out". In order to study the scanning behavior while subjects plan and execute an action, we computed transition matrices to capture fixations to and from each of the seven ROIs as described in Section 3.3.5. Figure 3.8A shows the exemplar transition matrix for a planning epoch in an EASY trial. With the transition matrices we wanted to capture the gaze guidance behavior of the subjects while they plan and execute the actions. The relative net-transitions within the planning and execution epochs of a trial tell us the different functions of gaze guidance behaviors exhibited of the subjects. With higher relative net transitions, we expect higher gaze guidance to the current task-relevant ROIs, i.e, subjects perform saccades only for guiding their hand or body towards the current target object and less so for searching for task-relevant objects or monitoring the task. If subjects perform a search and fixate on multiple ROIs in an epoch, we would expect lower relative net transitions indicating a pattern of fixations related multiple task relevant schemas that compete for selection.

In order to show the differences in mean gaze guidance behavior in a trial we used a linear mixed effects model (section 3.3.5) with relative net transitions as the dependent variable and trial type (EASY, HARD), epoch type (PLANNING, EXECUTION) and number of object displacements as independent variables. As the independent variables were effect coded, the regression coefficients could be directly interpreted as main effects. Figure 3.8B illustrates the effects of trial type, epoch type and number of object displacements on the relative net transitions. There was a significant main effect of factor trial type (HARD - EASY) β = -0.05 (95%CI = [-0.06, -0.03], t(77.5)=-5.72), with a p-value < 0.001 showing that HARD trials had lower relative net transitions than EASY trials. There was also a significant main effect of factor epoch type (PLANNING - EXECUTION) β = 0.03 (95%CI = [0.00, 0.05], t(46.57)=-2.16), with a p-value = 0.03 showing that PLANNING epochs had higher relative net transitions than EXECUTION epochs. There was a significant effect of number of object displacements in a trial β = 0.003 (95%CI = [0.00, 0.01], t(57.64)=2.83), with a p-value = 0.006 showing that a one unit increase in the number of object displacements in a trial led to increase in relative net transitions by a factor of 0.003. There was a significant interaction between trial type and epoch type β = -0.05 (95%CI = [-0.08, -0.01], t(53.07)=-2.5), with a p-value = 0.01 showing that PLANNING epochs in HARD trials had lower relative net transitions. There was a significant interaction between trial type and number of object displacements β = -0.009 (95%CI = [-0.01, -0.01], t(45.82)=-4.31), with a p-value < 0.001 showing that a one unit increase in the number of object displacements in HARD trials led to increase in relative net transitions by a factor of -0.009. There was no significant interaction between epoch type and number of object displacements β = 0.001 (95%CI = [0.00, -0.01], t(43.20)=0.64), with a p-value = 0.52. There was a significant interaction between trial type, epoch type and number of object displacements β = -0.01 (95%CI = [-0.02, 0.00], t(42.67)=-2.20), with a p-value = 0.03 showing that a one unit increase in the number of object displacements in HARD trials and for PLANNING epochs led to increase in relative net transitions by a factor of -0.01.

The analysis above lends further evidence that task complexity had a significant effect on the gaze guidance behavior at the level of action planning and execution. The lower relative net transitions in the HARD tasks in general are indicative of competition between action schemas either for searching task-relevant ROIs or for monitoring the task progression. The higher relative net transitions in the EASY trials suggest saccades were primarily made towards the current task-relevant objects for directing or guiding the body or hand towards the object of interest. The significant correlation of the object displacements and the relative net transitions reveal that gaze allocation predominantly occurred in a just-in-time manner supporting the sub-optimal behavior exhibited by the subjects as well.

To further disentangle the effect of task complexity on the order of gaze allocation to the task-relevant ROIs, we were interested in the latency of the first fixations to these . We used linear mixed effects regression to model the median time to first fixation on the 7 ROIs in each trial as described in section 3.3.5. We modeled the latency of the first fixations for the planning and execution epochs separately. Figure 3.8C shows the distribution of the normalized time to first fixations on the seven ROIs for the action planning and execution epochs.

In the action planning epoch, the time to first fixation on the previous target object was significantly earlier than the first fixation on the current target object β = -0.35 (95%CI = [-0.37, -0.32], t(49.67)=-29.01), with a p-value < 0.001. Similarly, the time to first fixation on the previous target shelf was significantly earlier than the first fixation on the current target object β = -0.35 (95%CI = [-0.37, -0.33], t(65.96)=-37.82), with a p-value < 0.001. The time to first fixation on other ROI was significantly earlier than the first fixation on the current target object β = -0.33 (95%CI = [-0.35, -0.31], t(72.37)=-37.03), with a p-value < 0.001. There was also a significant time difference between the first fixation on the next target object and the current target object β = -0.06 (95%CI = [-0.08, -0.05], t(47.36)=-6.96), with a p-value < 0.001. There was a significant time difference between the first fixation on the next target shelf and the current target object β = -0.05 (95%CI = [-0.07, -0.02], t(49.25)=-3.81), with a p-value < 0.001. Finally, there was also a significant time difference between the first fixation on the current target shelf and the current target object β = -0.04 (95%CI = [-0.06, -0.02], t(50.17)=-3.39), with a p-value = 0.001. When taking into account task complexity, there was a a significant interaction in the latency between previous target object and current target object β = -0.06 (95%CI = [-0.09, -0.02], t(123.78)=-3.11), with a p-value = 0.002. Similarly, there was also a significant interaction of trial type and time difference between first fixation between previous target shelf and current target object β = -0.04 (95%CI = [-0.07, -0.01], t(403.67)=-2.31), with a p-value = 0.02. There was also a significant interaction of trial type and latency between other ROI and current target object β = -0.05 (95%CI = [-0.08, -0.02], t(486.20)=-3.23), with a p-value = 0.001. There was also a significant interaction between trial type and latency between next target object and current target object β = -0.05 (95%CI = [-0.09, -0.02], t(83.82)=-3.16), with a p-value = 0.002. There was no significant interaction between trial type and the latency between next target shelf and current target object β = -0.009 (95%CI = [-0.05, 0.04], t(52.01)=-0.39), with a p-value = 0.69. There was also no significant interaction between trial type and delay between current target shelf and current target object β = -0.002 (95%CI = [-0.04, 0.04], t(55.39)=-0.10), with a p-value = 0.91.

Taken together, the irrespective of task complexity, the action planning epochs show a systematic progression of fixations from one ROI to another. This structured temporal sequence of fixations with a defined temporal window shows that the look-ahead fixations pertaining to the future action relevant ROIs are not incidental and part of the cognitive schema to accomplish the task. Moreover, given the task complexity, the temporal profiles of these look-ahead fixations can change and occur slightly earlier.

In the action execution epoch, the time to first fixation on the other ROI was significantly later than the first fixation on the current target object β = 0.17 (95%CI = [0.14, 0.19], t(50.47)=11.40), with a p-value < 0.001. Similarly, the time to first fixation on the previous target object was significantly later than the first fixation on the current target object β = 0.033 (95%CI = [0.30, 0.35], t(61.73)=24.57), with a p-value < 0.001. The time to first fixation on previous target shelf was also significantly later than the first fixation on the current target object β = 0.33 (95%CI = [0.30, 0.36], t(52.69)=23.59), with a p-value < 0.001. There was a significant time difference between the first fixation on the next target object and the current target object β = 0.36 (95%CI = [0.33, 0.39], t(58.75)=24.56), with a p-value < 0.001. There was a significant time difference between the first fixation on the next target shelf and the current target object β = 0.43 (95%CI = [0.40, 0.46], t(54.68)=28.49), with a p-value < 0.001. Finally, there was also a significant time difference between the first fixation on the current target shelf and the current target object β = 0.42 (95%CI = [0.39, 0.44], t(64.75)=32.13), with a p-value < 0.001. When taking into account task complexity, there was no significant interaction in the latency between other ROI and current target object β = 0.008 (95%CI = [-0.03, 0.05], t(136.42)=0.40), with a p-value = 0.69. There was no significant interaction between trial type and latency between previous target object and current target object β = 0.02 (95%CI = [-0.02, 0.07], t(116.24)=0.96), with a p-value = 0.34. There was no significant interaction between trial type and the latency between previous target shelf and current target object β = 0.01 (95%CI = [-0.04, 0.05], t(98.32)=0.49), with a p-value = 0.62.There was no significant interaction of trial type and time difference between first fixation between next target object and current target object β = -0.002 (95%CI = [-0.06, 0.05], t(60.67)=-0.10), with a p-value = 0.92. There was also no significant interaction of trial type and latency between next target shelf and current target object β = -0.01 (95%CI = [-0.07, 0.05], t(63.53)=-0.35), with a p-value = 0.73. There was a significant interaction between trial type and delay between current target shelf and current target object β = 0.05 (95%CI = [0.00, 0.09], t(122.10)=2.15), with a p-value = 0.03.

In sum, irrespective of the task complexity, the action execution epochs show a systematic sequence of first fixations on the seven ROIs. The fixations on the previous action related object and shelf show that a monitoring of the current action with respect to the previous action might be at play, where these fixations are made to confirm the choice of the current target shelf and might serve as look-back fixations. Similarly, fixations on the target object and shelf for the next action might serve as look-ahead fixations. However, with increasing task complexity, the temporal sequence of the first fixations remained unchanged except for the later gaze allocation to the current target shelf in the case of HARD trials. One may posit here that the added task complexity does not affect the temporal sequence of the gaze allocation during action execution.

3.3.8 Discussion and Conclusion

In the present study we investigated the spatio-temporal aspects of gaze control while action execution and action planning with varying task complexity. We report five main findings with this study. First, we found subjects used ad-hoc spatial heuristics to reduce the solution search space resulting in moderately increased number of object displacements as compared to the optimal model. Second, fixations locked to the action initiation, revealed a prevalence of look-ahead and task monitoring fixations in the HARD tasks but not in the EASY task. Third, based on the scan path transitions, we observed greater relative net transitions in the EASY trials compared to HARD trials. The lower proportion of net transitions to and from the ROIs in HARD trials is further evidence for an extensive search behavior and alternating action guiding and monitoring fixations. Fourth, the relative timing of the first fixations on the immediate task-relevant ROIs in the action planning and execution phase revealed a systematic sequence of fixations leading up to the immediate task-relevant ROI, indicating a just-in-time strategy of action supporting fixations. Finally, task complexity affected the systematic temporal sequence of fixations in the action planning phase but not in action execution phase. In sum, our findings reveal a structured effect of task complexity on the spatio-temporal features of relevant action supporting cognitive schemas.

The central aim of the our study is to generalize the cognitive mechanisms of gaze control for tasks that are not over-learned and routine. Building on the work of M. Land et al. (1999) where eye movements were studied while preparing tea, the tasks in our study are novel in the sense that they do not have an inherent action sequence attached to them. Our experimental setup provided a way to capture oculomotor behavior for tasks with varying complexity and that did not have a strict action sequence. By studying the dispersion of fixations on the previous, current, and next task relevant ROIs, we show a structured sequence of fixations that can be classified into look-ahead fixations, directing fixations, guiding fixations, checking fixations, etc. Our results generalize the occurrence of these fixations to the action sequence and the task complexity by doing away with object identity.

M. Land et al. (1999) proposed the various functions of eye movements from locating, directing, guiding, checking during tea-making. In our study, we also showed the occurrence of these fixations time-locked to the action initiation. Importantly, our study shows that task complexity affects the proportion and timing of fixations for locating/look-ahead as well as for checking the task progression. While look ahead fixations occurred predominantly before the initiation of the immediate action, the checking fixations occurred in parallel with the action execution. Interestingly, directing and guiding fixations were not affected by task complexity indicating that these fixations are central to the action repertoire.

B. Sullivan et al. (2021) recently showed the various timescales at which look-ahead fixations can occur while assembling a tent. They reported look-ahead fixations attributed to the current sub-task and within 10 seconds of motor manipulation. In our study, we found look-ahead fixations on the upcoming action related to the drop-off location of the current object displacement as well as towards the next object displacement sub-task. Our results show that these fixations are more salient in the more complex tasks. As evidenced by the relative net transitions to the task-relevant ROIs, there are more transitions in the action planning phase of the cognitively demanding tasks but they are not made at random. Further, we also show the systematic latency of these fixations to the immediate action. Hence, the look-ahead fixations described in our study are part of an elaborate cognitive schema and less likely due to incidental fixations on the task based ROIs.

M. F. Land and Furneaux (1997) have further elaborated on the schemas that direct both eye movements and the motor actions. They proposed a chaining mechanism by which one action leads to another where the eyes supply constant information to the motor system via a buffer. In our analysis, the strict temporal structure of the first fixations on the ROIs lends further evidence of such a cognitive buffer that facilitates moving from one action to another where interspersed eye movements queue-in the objects and locations necessary for current and future actions. Importantly, the final first fixation is always the object or location necessary for the immediate sub-task, indicating the universality of the 'just-in-time' nature of these action-related fixations. Taken together, our study further corroborates the cognitive schemas that sequentially support action planning and execution.

Our results show higher task solutions with increasing task complexity. To assess the sorting behavior of the participants we compared their object displacement behavior with a greedy depth-first search algorithm which optimizes for the shortest path to the solution. Studies in human performance in reasoning tasks as well as combinatorial optimization problems (MacGregor & Chu, 2011) have revealed that humans solve these tasks in a self-paced manner rather than being dependent on the size of the problem. Pizlo and Li (2005) found that subjects do not perform an implicit search of the problem state space where they plan the moves without executing them, where longer solution times would lead to shorter solution paths. Instead, they showed that humans break the problem down to component sub-tasks which gives rise to a low-complexity relationship between the problem size and time to solution. Further, Pizlo and Li (2005) show that instead of using implicit search, subjects use simpler heuristics to decide the next move. To this effect, while subjects in our study are sub-optimal compared to a depth-first search algorithm and use an arbitrary spatial heuristics during the task, humans in general are prone to use non-complex heuristics that favor limited allocation of resources in the working memory. Hence, the higher time duration and number of object displacements shown by the participants do not necessarily demonstrate a lack of planning.

From the perspective of embodied cognition (Ballard et al., 2013; Van der Stigchel, 2020; M. Wilson, 2002), humans off-load cognitive work on the environment and the environment is central to the cognitive system. The behavioral results show that subjects use spatial heuristics to complete the tasks indicating they exploit the external world to reduce the cognitive effort of selecting optimal actions. Further, Droll and Hayhoe (2007) have suggested that the just-in-time strategy to lower the cognitive cost of encoding objects in the world into the visual working memory. In the embodied cognition framework, cognition is situated and for producing actions. With our study, we show that task complexity affected various spatio-temporal features of gaze control in the action planning and execution phase. These findings are also in line with P. König et al. (2016) illustrating that eye movements reveal much about the cognitive state.

In conclusion, in the present study we investigated the oculomotor responses to novel task scenarios that varied in complexity. Our results generalize past work on action-oriented eye movements to tasks that are not routine or over-learned such as tea-making, sandwich-making, hand-washing, etc. We show that eye movements support various functions of locating object of interest, directing and guiding the body and hands, as well as monitoring the task progression. Furthermore, we show how fixations are tightly coupled with the action sequences in the task. More importantly we show the prevalence of cognitive schemas that are affected by the task complexity even when sub-optimal decisions are made by the actor.

Acknowledgement

We are grateful for the financial support by the German Federal Ministry of Education and Research for the project ErgoVR (Entwicklung eines Ergonomie-Analyse-Tools in der virtuellen Realität zur Planung von Arbeitsplätzen in der industriellen Fertigung)-16SV8052.

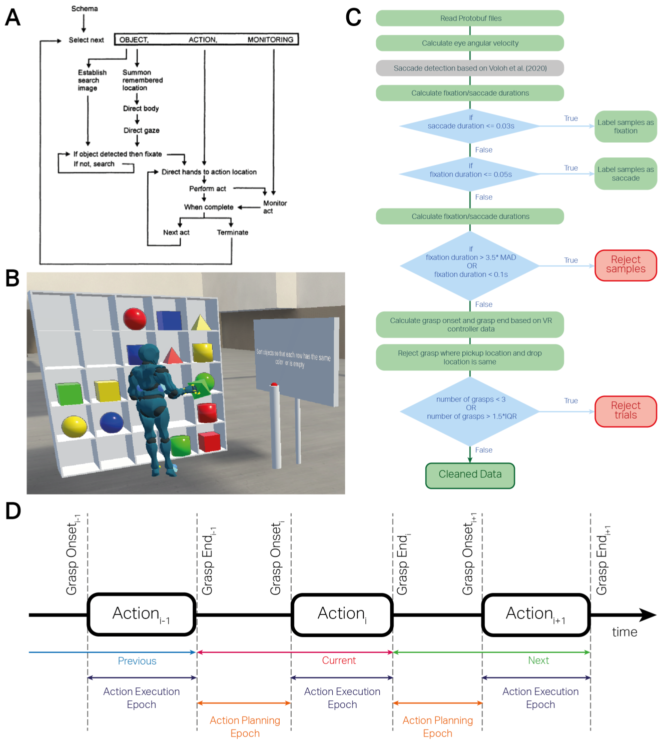

Figure 3.5: **A.** Schematic of motor and gaze control during performance of natural tasks M. F. Land and Hayhoe (2001). **B.** Experimental Task. In a virtual environment participants sorted 16 different objects based on 2 features color or shape while we measured their eye and body movements. The objects were randomly presented on a 5x5 shelf at the beginning of each trial and were cued to sort objects by shape and/or color. Trials where objects were sorted based on just one object feature (color or shape) were categorized as EASY trials. Conversely, in the trials where sorting was based on both features (color and shape) were categorized as HARD trials. All participants performed 24 trials in total (16 easy trials and 8 hard trials) with no time limit. **C.** Data preprocessing steps for fixation/saccade detection and data rejection. **D.** Action execution and planning epochs. In order to study the function of eye movements we divided each trial into action planning and action execution epochs. The action execution epochs start from grasp onset till grasp end for each object displacement, whereas the action planning epochs start from grasp end of previous object displacement and grasp onset of current object displacement.

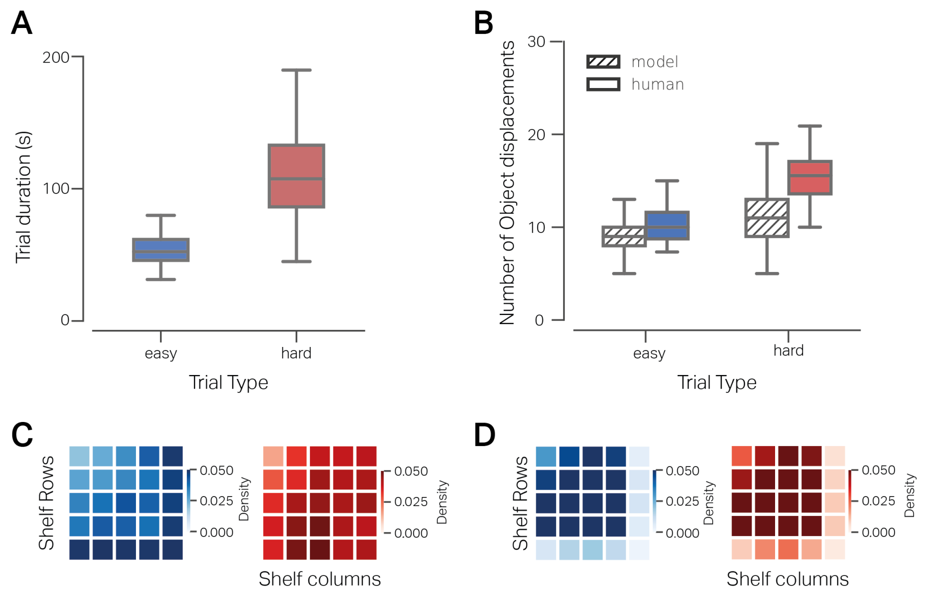

Figure 3.6: **A.** Trial duration of the EASY and HARD trials. The boxplots show the inter-quartile range (IQR) of the duration of the trials for the two different trial types for all trials and participants. The whiskers represent 1.5 times the IQR. **B.** Distribution of number of object displacements for EASY (blue) and HARD (red) trials. The colored box plots show the inter-quartile range of the number of object displacements made by subjects per trial and per participant. The whiskers represent 1.5 times the IQR. The dashed box plots show optimal number of displacements required to sort the objects for a model computed with a depth-first search algorithm for 5000 random trial initialization for each trial type. **C.** Propensity of picking up objects from a preferred location on the 5x5 shelf locations with blue heatmaps showing probability density over EASY trials and red heatmaps showing density over HARD trials. The probability density shows that subjects have a propensity to pickup objects from the rightmost column and bottom column for EASY trials (left) and conversely, in the HARD trials (right) subjects pickup objects from central locations. **D.** Propensity of dropping-off objects to a preferred location on the 5x5 shelf locations with blue the heatmap showing probability density over EASY trials and red heatmap showing density over HARD trials. The probability density shows that subjects display a systematic propensity to place the objects every where other than the bottom row or rightmost column.

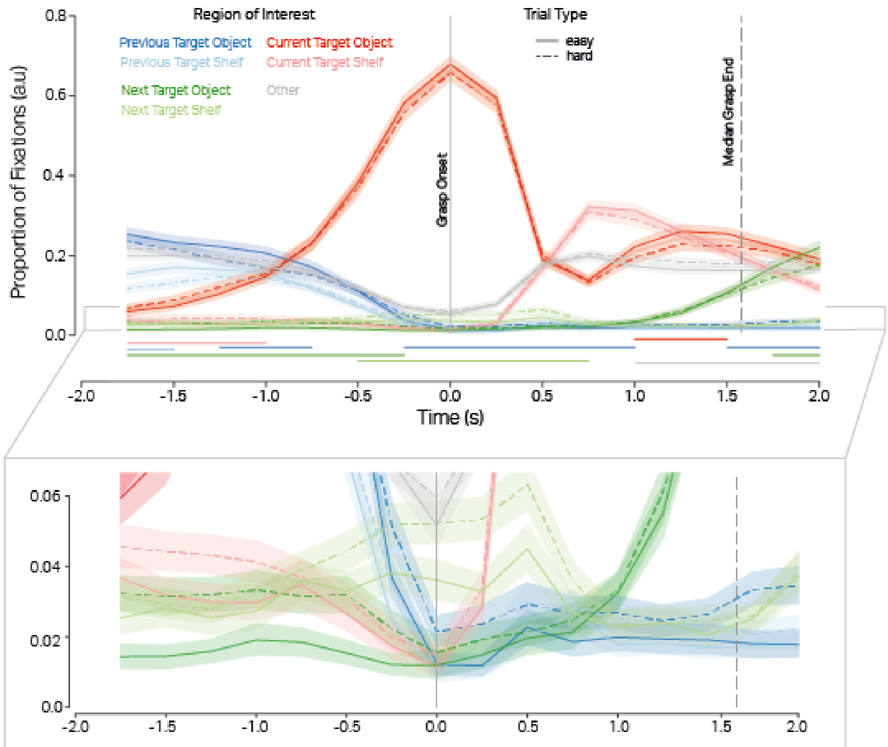

Figure 3.7: Proportion of fixations time-locked to the object displacement initiation (grasp onset) at time=0 and 2 seconds before and after on the seven regions of interest and for the EASY (solid trace) and HARD (dashed trace) tasks. In each time bin of 0.15s the proportion of fixations on all 7 ROIs for a trial type add up to 1. The dashed vertical line denotes the median end of the action execution phase The shaded regions show 95% confidence interval of the mean proportion of gaze at each time-bin across all subjects. The proportion of fixations in each The horizontal traces at the bottom correspond to the significant time periods where the proportion of fixations on an ROI were different for the EASY and HARD trials.

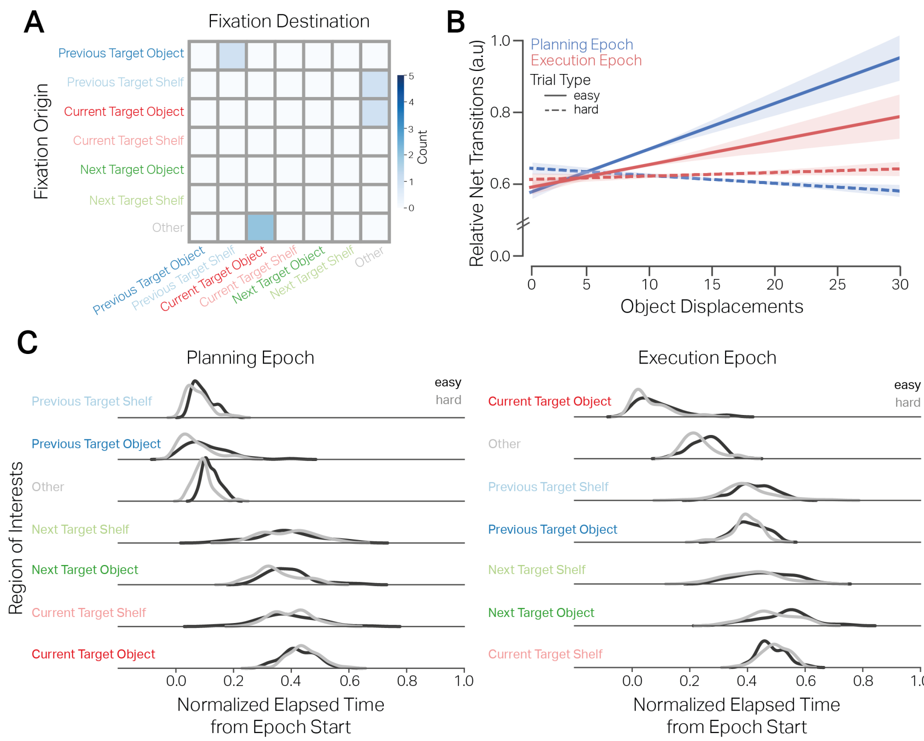

Figure 3.8: **A.** Exemplar transition matrix for gaze switching in a planning epoch. The ordinate defines the origin of the gaze and the abscissa defines the destination of the gaze. **B.** Regression fit over the fixed-effects of trial types (EASY, HARD), epoch types (PLANNING, EXECUTION) and object displacements on the relative net transitions. The traces denote the the regression fit and the shaded region denotes 95% confidence interval. **C.** Distributions of median time to first fixation on the 7 ROI for the action planning and execution epochs and trial types. The ordinate is sorted in ascending order of latency from epoch start for both planning and execution epochs.

4 Ecological Validity

4.1 Layman's summary

We have developed large-scale realistic virtual environments and investigated the validity of virtual reality experimentation compared to traditional laboratory. But what about the experiments not possible even in real life? In previous parts, we have investigated the new level of understanding that virtual reality experiments can offer. We have delved deep into the "How" of designing and conducting virtual reality experiments, as well as the debate of "what ecological validity" means in cognitive science studies and related fields. The final goal of the cognitive scientist is to study human cognitive processes in their entire complexity inside complex and dynamic environments, no matter whether ecological validity refers to the mundane reality, the closeness of the experimental cues to real life, or the environment itself. Although feasible in real life, there might also be situations in some experiments that are impossible to conduct under lab or real-life circumstances. The reason might lie in technological shortcomings to ethical issues. Nonetheless, these experiments, for instance, research on trolley dilemma, were traditionally performed by reading and imagining a scenario followed by answering questions from a questionnaire.

In a traditional pen and paper situation, almost everyone will go with the utilitarian approach: saving more lives. However, in virtual reality, manipulation of prospective shows us the difference between perception and certainty of our choice despite the fact that most participants would choose the utilitarian choice. In this experiment on trolley dilemma with manipulation of perspective, meaning changing the position of participant from being just an observer, to passenger of the car or being among the group of people that the car hits, the finding indicates the strong role of perspective in our understanding of the situation. Needless to say, these studies would be impossible without using virtual reality.

The same logic will go for just observing participants' behavior in virtually simulated realistic environments. Utilizing the environment created in the project Westdrive, we managed to drive more than 20000 participants through an arguably short but immersive drive through the virtual city. Manipulating the type of the car from a typical self-driving car to one that talks to the passenger and well-known taxi, we could observe the change in people's attitude and acceptance toward the observed vehicle. This change was visible not only through a self-reported short questionnaire but also in the participants' behavior from the observation of their head movement. Here the participants were passive observers yet were immersed in the experience. In the end, just a short ninety-second drive using an immersive virtual reality experience and just observing their behavior and evaluating their reports could already broaden our view on public acceptance toward the self-driving car.

4.2 Moral Judgements on Actions of Self-driving Cars

This section was published as a peer reviewed article in Frontiers in Frontiers in psychology together with Noa Kallioinen, Maria Pershina, Jannik Zeiser, Achim Stefan, Gordon Pipa and Peter König. See Publication List for details.

4.2.1 Abstract

Self-driving cars have the potential to greatly improve public safety. However, their introduction onto public roads must overcome both ethical and technical challenges. To further understand the ethical issues of introducing self-driving cars, we conducted two moral judgement studies investigating potential differences in the moral norms applied to human drivers and self-driving cars. In the experiments, participants made judgements on a series of dilemma situations involving human drivers or self-driving cars. We manipulated which perspective situations were presented from in order to ascertain the effect of perspective on moral judgements. Two main findings were apparent from the results of the experiments. First, human drivers and self-driving cars were largely judged similarly. However, there was a stronger tendency to prefer self-driving cars to act in ways to minimise harm, compared to human drivers. Second, there was an indication that perspective influences judgements in some situations. Specifically, when considering situations from the perspective of a pedestrian, people preferred actions that would endanger car occupants instead of themselves. However, they did not show such a self-preservation tendency when the alternative was to endanger other pedestrians to save themselves. This effect was more prevalent for judgements on human drivers than self-driving cars. Overall, the results extend and agree with previous research, again contradicting existing ethical guidelines for self-driving car decision making and highlighting the difficulties with adapting public opinion to decision making algorithms.

4.2.2 Introduction

Self-driving cars are rapidly becoming a reality. In 2016, car manufacturer Tesla announced that all of its current cars were being equipped with the hardware necessary for autonomous driving(The Tesla Team, 2016). Since then, Tesla has incrementally enabled autonomous and assisted driving features via software updates(The Tesla Team, 2019). Other manufacturers have since been following suit (Mercer and Macaulay, 2019, see) and the use of partially self-driving cars, such as these, is expected to increase within the next 20 years.

A major argument supporting the development of self-driving cars is the expected reduction in the number of traffic accidents. For example, close to 90% of the more than 300,000 traffic accidents resulting in injuries to people in Germany in 2017 were caused by driver misconduct or error, such as ignoring right of way, inappropriate following distance or speed, overtaking faults and driving under the influence of alcohol (Statistisches Bundesamt, 2018, p. 49). Similar observations have been made in both the United Kingdom and the United States (Department for Transport, 2013; National Highway Traffic Safety Administration, 2008). These errors and misconduct can potentially be mitigated by the introduction of self-driving cars, which highlights their potential to improve public safety.

However, the expected reduction of accidents will need time to be realised. Recently published statistics by the California Department of Motor Vehicles shows that self-driving car prototypes are involved in accidents at a similar rate as human drivers (Favarò et al., 2017). Other reports give somewhat more favorable numbers with a reduction of accident rates by about one third (Marshall, 2018; Thomas, 2018). The discrepancy to the optimistic forecasts cited above stems in part from an increase of, for example, unexpected breaking resulting in rear-end collisions, and the fact that even when an accident is not caused by a self-driving car, it might still be involved in it. Thus, during a multi-year introduction period, self-driving cars will be involved in a substantial number of accidents and unexpected situations. Unexpected traffic situations are often highly complex and require split-second decisions. For this reason, human drivers are not generally expected to be able to respond optimally and may be excused for making wrong decisions (Trappl, 2016). Self-driving car control systems, on the other hand, can potentially estimate the outcome of various options within milliseconds and take actions that factor in an extensive body of research, debate, and legislation (Lin, 2015). The actions taken in such situations have potentially harmful consequences for car occupants, other traffic participants, and pedestrians. Therefore, it is important to carefully consider the ethics of how self-driving cars will be designed to make decisions, an issue that is the topic of current debate (Dietrich and Weisswange, 2019; Keeling et al., 2019; Nyholm, 2018a, 2018b).

Comprehensive guidelines for ethical decision making for self-driving cars have been provided by the ethics commission of the German Federal Ministry of Transport and Digital Infrastructure, 2017. These guidelines speak out against a standardised procedure of decision making in dilemma situations (guideline 8). In cases of unavoidable accidents, “any distinction based on personal features (age, gender, physical or mental constitution) is strictly prohibited” and “[those] parties involved in the generation of mobility risks must not sacrifice non-involved parties” (guideline 9). These guidelines greatly add to the discussion and can inform the development of decision making systems. However, it is far from obvious that a practical implementation of these guidelines would garner public consensus.

As pointed out by Shariff et al., 2017, and further evident by the number of studies focusing on public opinion (Gkartzonikas and Gkritza, 2019, see) the introduction of self-driving cars requires acceptance from the public. Empirical research investigating public perception and beliefs can be useful for highlighting areas problematic for the acceptance of self-driving cars into public traffic. Such research in the area of ethical decision making for self-driving cars has primarily focused on human decision making as a basis. In a typical experiment, participants make decisions pertaining to hypothetical dilemma situations in which harm is unavoidable. Situations of this kind, known as trolley dilemmas (Jarvis Thomson, 1985), involve two groups of people, one of which must be endangered to spare the other. The utility of trolley dilemmas does not lie in their use as blueprints for crash optimizations (Holstein and Dodig-Crnkovic, 2018). Rather, they are an effective means to elucidate which ethical values are potentially conflicting in accident scenarios and to allow for the design of self-driving cars informed by human values (Gerdes et al., 2019; Keeling, 2019). As argued by J. Bonnefon et al., 2019, trolley dilemmas should not be understood primarily as simulations of real-life scenarios, but as representations of conflicts that emerge on a statistical level: The introduction of self-driving cars will likely put different people at risk compared to today. For example, would it be acceptable that due to self-driving cars, fewer people are harmed in traffic, but those who are harmed are more likely to be pedestrians than car occupants?

Moral dilemma studies can be grouped broadly into two paradigms: those that investigate moral judgements (what people claim are the right actions) and those that investigate moral actions (what people actually do in a given situation). An analysis of more than 40 million judgements on vignettes describing hypothetical dilemma situations concluded that people generally prefer self-driving cars to endanger fewer lives, endanger animals over people and endanger older people over younger people (Awad et al., 2018). Other moral judgement studies include simulation studies by H. Wilson et al., 2019 and Wintersberger et al., 2017 and vignette-based studies by Rhim et al., 2020, Smith, 2019, Meder et al., 2018, J.-F. Bonnefon et al., 2016 and J. Li et al., 2016. Importantly, J.-F. Bonnefon et al., 2016 found a discrepancy between what people deemed acceptable for self-driving cars to do in dilemma situations and their willingness to purchase cars that would act accordingly. Specifically, people considered it more morally acceptable for self-driving cars to endanger fewer lives, even at the expense of the occupants' lives, but preferred to purchase cars that would protect occupants. Martin et al., 2017 suggested that this discrepancy may be resolved if people explicitly consider the situations from both the perspectives of car occupants and pedestrians. Borenstein et al., 2019 highlighted that the perspectives of pedestrians and other non-occupants is overshadowed by the focus on car occupants in the literature, but are equally important.

Studies of moral action have used virtual reality environments to determine how human drivers would act when faced with dilemma situations. In these studies, participants were put in the perspective of drivers and controlled the steering of virtual vehicles when facing such dilemma situations. Skulmowski et al., 2014b placed participants in the role of train drivers and found participants generally preferred to save the greater number of lives. Sütfeld et al., 2017 found that the behavior of participants in the role of car drivers could be well described by a value-of-life model, such that people are valued more than animals and younger people are valued more than older. S. Li, Zhang, et al., 2019, Faulhaber et al., 2019 (further elaborated by Bergmann et al., 2018b) showed that car drivers also tend to act in ways that endanger fewer lives, even at the expense of their own. Ju et al., 2019 found that personality characteristics predict the likelihood of drivers endangering themselves. Furthermore, Luzuriaga et al., 2019 directly compared actions chosen by participants tasked with programming a self-driving car with actions made by participants in a driving simulator. They found that participants programming a self-driving car more readily endangered car occupants to save pedestrians, than participants driving in a simulator. Thus, our knowledge of how humans act in critical situations in virtual reality is increasing.

While the results of these moral judgement and moral action studies have been generally consistent, there are important distinctions between the approaches needing consideration before making strong conclusions. First, there is growing evidence of discrepancies between what people consider to be the right action in moral dilemmas and what they would actually do (e.g. FeldmanHall et al., 2012; Francis et al., 2016; N. Gold et al., 2015; Patil et al., 2014; Tassy et al., 2013). Additionally, what is generally considered ethical for human drivers may not be the same for self-driving cars. Furthermore, the perspective from which the situations are presented may affect how they are evaluated.

To address aforementioned issues, we conducted two studies in the moral judgements paradigm which allowed us to investigate moral beliefs about self-driving cars and human drivers in dilemmas situations from different perspectives. In both studies, we recorded judgements pertaining to virtual dilemma situations involving either self-driving cars or human drivers. We included the perspectives of car occupants, uninvolved observers and pedestrians, which to our knowledge, no previous studies have done. Study 1 employed virtual reality to investigate judgements in specific dilemma situations, while Study 2 used simplified animations and varied aspects of the situations in a more fine-grained manner.

4.3 Study 1: Moral Judgements in Virtual Reality

In this study, we addressed the effects of perspective (passenger, pedestrian or observer) and type of motorist (human driver or self-driving car) on moral judgements in immersive virtual environments. We investigated three different scenarios, all involving the choice between endangering one of two groups of virtual avatars. The scenarios were designed to be morally ambiguous to avoid ceiling or floor effects. We hypothesised a self-preservation effect, such that, independent of the type of motorist, participants would be less likely to judge actions that endangered their own virtual avatars as more acceptable.

4.3.1 Materials and method

Participants

184 people (96 male, 88 female) voluntarily participated in the virtual reality experiment. Participants were recruited through social media, university mailing lists, word of mouth or were directly approached. Participants could earn experiment participation credits required for some university programs, but no monetary incentive was provided. Participants were required to be at least 18 years old with

native-level German and gave written informed consent after being briefed on the content of the experiment. Exclusion criteria included having experienced previous car-related trauma, being prone to motion sickness and having a history of epileptic seizures. The study was approved by the ethics review board at Osnabrück University, Germany. Descriptive statistics of the participants are shown in Table S1.

Materials

The stimuli consisted of six pairs of virtual reality animations, each approximately 30 seconds in duration, created with Unity (Unity Technologies, 2018). Each scenario involved a car with two occupants: driver and passenger (human driver condition) or two passengers (self-driving car condition). The car drove in the middle of a road and encountered a dilemma situation in which it could veer either to the left or the right, endangering one of two groups of avatars. Animations depicting both possible actions were shown in sequence.

To prevent unnecessary distress, the animations and sound effects in the virtual environment ceased immediately before the car would be involved in a collision. A braking sound effect was played in the moments before the animations ended to demonstrate that the car attempted, but was unable, to stop before impact. Participants had no control over the car or avatars, but could freely observe the virtual environment. If the motorist was a self-driving car, the steering wheel of the car was absent and a label was shown at the front of the car indicating that it was self-driving in order to remind participants during the course of the experiment. Three different scenarios were investigated: child pedestrians versus adult pedestrians; pedestrians on the road versus pedestrians on the sidewalk; and car occupants versus pedestrians. Each scenario included two different trials.

In the child pedestrians versus adult pedestrians scenario the car either veered towards a group of pedestrians including children or a group of only adult pedestrians. The two trials differed by group size, but the ratio was static. In the smaller groups trial, there was one child (and an adult viewpoint avatar) in one group and two adults (and an adult viewpoint avatar) in the other group; in the larger groups trial, there were two children (and an adult viewpoint avatar) in one group and 4 adults (and an adult viewpoint avatar) in the other group.

In the pedestrians on the road versus pedestrians on the sidewalk scenario, the car veered towards either adult pedestrians standing on the sidewalk or adult pedestrians standing on the road. The two trials differed by group size, but the ratio was static. In the smaller groups trial, there was one pedestrian on the sidewalk and two pedestrians on the road; in the larger groups trial, there were two pedestrians on the sidewalk and four on the road.

In the car occupants versus pedestrians scenario, the car veered towards either the pedestrians on the road or an obstacle that would endanger the lives of the car occupants. Instead of varying by the size of the groups, the two trials differed by the type of obstacle. In the parked van trial, the car would veer towards a large van parked on the side of the road, whereas in the cliff trial, the car would veer towards a cliff edge. Both variations of these scenarios are equivalent in the implied outcome: either car occupants or pedestrians will be harmed. While Faulhaber et al., 2019 only investigated endangering car occupants in the context of a cliff setting, we wanted to contrast this scenario with a less extreme setting. By having the car veer towards a parked van, harm towards car occupants is still implied, but the scenario is overall more integrated into a typical traffic setting.

We chose these specific types of scenarios as they allow us to contribute to related findings and discussions in recent literature. The influence of potential victims' ages has been investigated by Awad et al., 2018, Faulhaber et al., 2019 (further elaborated by Bergmann et al., 2018b), and Sütfeld et al., 2017. The potential protection afforded to pedestrians on a sidewalk has been studied in Faulhaber et al., 2019 (further elaborated by Bergmann et al., 2018b). The issue of prioritising car occupants or pedestrians has been theoretically discussed by Gogoll and Müller, 2016 and Lin, 2015, and implemented in a multitude of experiments including Ju et al., 2019, Awad et al., 2018, Faulhaber et al., 2019 (further elaborated by Bergmann et al., 2018b), Wintersberger et al., 2017 and J.-F. Bonnefon et al., 2016. The three scenarios are conceptually depicted in Figure 4.2 and details of the trials for each scenario are shown in Table 4.31.

As described, the numbers of lives at risk were unequal in the first two scenarios. There were twice as many pedestrians on the road compared to the sidewalk, and twice as many adults as children. These particular ratios were chosen based on the results from the study reported by Faulhaber et al., 2019, which were further elaborated by Bergmann et al., 2018b. The number of car occupants and pedestrians at risk were equal in the car occupants versus pedestrians scenario.

This ratio was anticipated to best elicit differences between the car occupant and pedestrian perspectives, as, barring any intrinsic bias towards pedestrians or car occupants, both should be equally valued.

Design

We employed a 4 (perspective) × 2 (motorist-type) between-participants factorial design. The two levels of motorist-type were self-driving car and human driver. The four levels of perspective were passenger, observer, pedestrian in the smaller group and pedestrian in the larger group. We used a between-participant design to prevent experimental confounds such as recognition of the trials and attempts to be self-consistent. As decisions made during previous trials could be easily recalled, we considered that a within-participant design would not have allowed us to distinguish whether participants were influenced more by the experimental manipulations or by their previous responses. Thus, variables were manipulated in such a way that each participant saw all trials from the same perspective and involving the same motorist-type. To control for gender effects such as those described by Skulmowski et al., 2014b, the genders of all human avatars in the virtual environment were matched to each participant.

Procedure

Participants were assigned via permuted block randomisation to one of the eight conditions corresponding to the combinations of perspective and motorist-type (e.g. observer & human driver; car occupant & self-driving car). Participants of the smaller and larger pedestrian groups shared the same car occupants versus pedestrians trials as there was only one pedestrian group involved in those scenarios. Participants completed a practice trial and a control trial before the experimental trials. The six experimental trials as well as animations within each trial were shown in random order; trials were separated by distraction tasks. After viewing a pair of animations, participants could replay the pair as many times as they wanted. Participants were then asked to choose which of the two actions of the motorist they considered to be more acceptable by selecting the corresponding outcome image. In accordance with Mandel and Vartanian, 2007, after making each judgement, participants indicated how confident they were in it on a scale from 0 (not confident at all) to 100 (very confident). Decision confidence in moral dilemmas has also been previously investigated by M. Lee et al., 2018, Parkinson et al., 2011 and Royzman et al., 2014, as it gives further information than merely the binary choice. Specifically, the confidence ratings provide information on how conflicted participants were about the corresponding judgements. High scores on confidence indicate more robust judgements than lower scores. Thus, the proportions of judgements and the corresponding confidence levels should be considered in parallel.

After the experiment ended, participants completed a short questionnaire on demographics, driving experience, prior knowledge of self-driving cars and their attitudes towards them. Furthermore, as a manipulation check, participants reported which party in the situation they identified most with while watching the animations by responding to the question “while watching the animations, which party did you identify most strongly with?”. The options were the pedestrians, the car occupants or the observer. Finally, they were asked whether the motorist was a human driver or a self-driving car. Those participants who failed the control task or were not able to recollect the correct motorist-type in the self-driving car condition were excluded.

Statistical Analysis

Statistical analyses were conducted in R (R Core Team, 2018) using *lme4*(Bates et al., 2015) for model fitting. Significance testing was performed using parametric bootstrapping with *afex* (Singmann et al., 2018) and *emmeans* (Lenth, 2018) was used for follow-up multiple comparisons on the estimated marginal means (*EMM*s).

Two models were computed for each of the three scenarios: one for the prediction of judgements (which of the two actions was considered more acceptable); the other for participants’ self-reported confidence in their own judgements. Judgements, based on perspective and motorist-type, were modelled by logit mixed models. As there were two trials per participant for each dilemma, random by-participant intercepts were included in all models. This corresponds to the maximal random effects structure as described by Barr et al., 2013 and Barr, 2013. Significance testing using Type-III sums of squares was performed by parametric bootstrapping with 1000 simulations. Confidence, based on judgement, perspective, motorist and trial was modelled by linear mixed models. Significance testing using Type-III sums of squares was performed using Kenward-Roger test. Along with trial (smaller groups/larger groups in the first two scenarios, parked van/cliff in the third scenario), the following covariates were included: gender, age, positive opinion of self-driving cars, visual acuity, education level and driving experience. Models without covariates are reported in the supplementary material, but did not result in different conclusions. Results for the three scenarios are reported separately.

4.3.2 Study 1 Results

Manipulation check

To determine whether varying the visual perspective affected which party participants self-identified with, we performed a chi-squared test of independence, comparing participants' self-identification with the perspective from which they experienced the situations (Table S2). The majority of participants identified most strongly with the perspective from which they experienced the scenarios $\chi^2(24, N = 184) = 114.11$, $p < .0001$. Follow up Bonferroni-adjusted comparisons showed all three perspective groups had significantly different patterns of responses from each other (all $p < .0001$) (Table S3). Thus, the manipulation check indicates that in most cases participants identified with the intended perspective.

Children versus adults

Next, we investigated the influence of perspective and motorist-type on judgements on the children versus adults scenario. According to model predictions, endangering the larger group, which consisted of only adult pedestrians, was considered more acceptable than endangering the smaller group, which consisted of adults and children (probability = 0.71). Figure 4.3A depicts the predicted probability of judgements and levels of confidence separated by perspective and motorist-type based on the statistical model. There were no significant effects of perspective or motorist-type on judgements (Table 4.32). The predicted mean self-confidence in judgements (on a 0 to 100 scale) was 49.92, however it varied considerably between conditions. There was a significant main effect of perspective ($p = .0017$) moderated by judgement ($p = .0222$) on self-reported confidence in judgements (Table 4.33). Within those who chose endangering the larger group (of only adults) as more acceptable, participants in the observer perspective had significantly lower confidence in their choices (*EMM* = 35.86) than either the pedestrian with children (*EMM* = 58.57) or the pedestrian with adults (*EMM* = 55.62) perspectives, p = .0178, p = .0358, respectively. Within those who chose endangering children as more acceptable, participants in the pedestrian with children perspective had significantly greater confidence (*EMM* = 71.87) than either the observer (*EMM* = 36.13), the passenger (*EMM* = 41.92) or the pedestrian with adults (*EMM* = 42.34), p = .0003, p = .0161, p = .0045, respectively (Tables S4 and S5). Thus observers had among the lowest confidence regardless of judgement.

Sidewalk versus road

In the second scenario, we tested small groups of pedestrians on the sidewalk against larger groups of pedestrians on the road. Overall, endangering the smaller group was considered more acceptable than endangering the larger group (probability = 0.84). Thus, participants overwhelmingly considered that endangering fewer pedestrians was more acceptable, despite those pedestrians being situated on a sidewalk. Mean confidence (on a 0 to 100 scale) was 62.44 and, thus, considerably greater than in the children versus adults scenario. Figure 4.3B depicts the predicted probability of judgements and levels of confidence separated by perspective and motorist-type based on the model. There were no significant effects of perspective or motorist type on judgements (Table 4.32). However, there was a significant effect of gender, such that females (probability = 0.004) were less likely to consider endangering the larger group of pedestrians (on the road) as more acceptable than males (probability = 0.034). Self-reported confidence depended on judgement (Table 4.33), such that choosing endangering pedestrians on the sidewalk as more acceptable was associated with greater confidence (*EMM* = 68.88) than choosing endangering pedestrians on the road (*EMM* = 60.93), p = .0332 (Tables S6 and S7). Thus, the observed differences in confidence matches the bias in judgement in the sidewalk versus road scenario.

Car occupants versus pedestrians

Finally, we investigated a scenario in which endangering car occupants was contrasted with endangering pedestrians. As the two trial types for this scenario were conceptually different, an interaction with trial type was included in the model.

For the parked van trial, the vast majority preferred to endanger the car occupants (probability = 0.99). In the cliff trial however, this was much less likely (probability = 0.53). Mean confidence was also different: 67.08 for the parked van trial and 43.62 for the cliff trial. Figure 4.3C depicts the predicted probability of judgements and levels of confidence separated by perspective and motorist-type for the parked van trial and Figure 4.3D depicts the same for the cliff trial.

There was a significant main effect of trial-type. Participants were more likely to consider endangering the car occupants as more acceptable in the van trial than the cliff trial, $p = .0010$. As falling off a cliff is more likely to result in injury or death than colliding with a parked van, the judgements by participants appear to take into account the degree of potential harm.

Furthermore, there was a significant trial-type × perspective interaction. In the cliff trial, passengers were significantly less likely than either observers (odds-ratio = 5.303, $p = .0047$) or pedestrians (odds-ratio = 3.584, $p = .0118$) to consider endangering the car occupants (including themselves) as more acceptable. This indicates a self-preservation effect.

Statistical analysis of self-reported confidence was performed only for pedestrians and car occupant perspectives as there were no responses preferring to endanger pedestrians in the observer perspective. There were main effects of trial ($p = .0052$) and judgement ($p = .0002$), moderated by a trial × judgement interaction ($p = .0011$), on self-reported confidence. Confidence when preferring to endanger car occupants was lower in the cliff trial (*EMM* = 47.8) than the parked van trial (*EMM* = 75.2), $p < .0001$. This was not the case for preferring to endanger pedestrians (*EMMs* = 50.4 and 55.2, respectively, $p = .7582$) (Table S11). Note that there were no observers who preferred endangering pedestrians in the parked van trial, so the confidence could not be estimated and the follow up comparisons for endangering pedestrians only considered the responses of the other perspectives.

4.3.3 Study 1 Discussion

For the three scenarios, patterns of judgements aligned with actions taken in similar dilemma studies reported by Faulhaber et al., 2019 (further elaborated by Bergmann et al., 2018b) and Sütfeld et al., 2017: participants generally preferred motorists to risk the lives of adult pedestrians rather than child pedestrians, despite endangering more lives by doing so; it was highly acceptable for a motorist to swerve onto a sidewalk in order to endanger fewer pedestrians; and there was a tendency to protect pedestrians over car occupants. However, it seems that the perceived danger to the car occupants plays a role; participants were less likely to accept a car veering towards a cliff edge, than a car veering towards a parked van.

Only in the cliff trial of the car occupants versus pedestrians scenario did we observe a main effect of perspective on judgements There was disagreement between the car occupant and pedestrian perspectives. Car occupants preferred the car to remain on course and endanger the pedestrians, rather than veering towards a cliff edge, while pedestrians preferred the opposite. Interestingly, observers appear to agree with the pedestrians in this case. This corresponds to a self-preservation effect for both car occupants and pedestrians. However, it is important to notice that this effect only arose when the situation clearly pitted the lives of car occupants against the lives of pedestrians. It was not prevalent between pedestrians, nor in the parked van trial (which may have been considered as less dangerous for the car occupants).

The collection of self-reported confidence allowed for a more fine-grained analysis by enabling effects that were not prevalent in the primary forced-choice response data to be investigated. Specifically, there was an effect of perspective in the children versus adults scenario: observers were among the lowest in confidence, regardless of judgement, despite there being no significant difference in judgements themselves. This is noteworthy as the uninvolved observer is often considered as an "objective" viewpoint (Coeckelbergh, 2016). One might then expect the observer perspective to be associated with high confidence, but this is not apparent here.

4.4 Study 2: Moral Judgements on Simplified Animations

Our second study builds on the first investigating the influence of perspective and motorist with the addition of investigating the influence of the number of lives at risk and the presence of a sidewalk. We used an online deployment platform and presented the scenarios in the form of simplified animations. Rather than offering an immersive experience, the goal of using simplified animations was to illustrate the scenarios while prompting participants to evaluate them from a particular perspective. We consider the use of animations to be a natural extension of the combination of simplified images and textual vignettes, as used in previous studies (Awad et al., 2018; J.-F. Bonnefon et al., 2016; J. Li et al., 2016). As such a combination has been shown by Sachdeva et al., 2015 to sufficiently manipulate perspective in moral dilemmas, simplified animations should similarly prompt participants to consider situations from the presented perspective. Nevertheless, a manipulation check was included in the analysis to confirm that such an effect occurred.

We tested whether increasing the number of lives at risk by staying on course increases the acceptability of swerving to endanger a single life. Further, we tested whether swerving onto a sidewalk would be less acceptable than swerving onto another road. We hypothesised that perspective would influence judgements, such that participants would be less likely to consider endangering their own avatars as the more acceptable action.

4.4.1 Materials and Method

Participants

368 people (176 male, 191 female, 1 other) voluntarily participated in this online animation-based experiment. Participants indicated their age groups, the median of which was 18–29 years old. Participants were recruited through social media, university mailing lists and word of mouth. 24 different countries were represented, with major participation from Germany, Armenia, Australia and Russia. The study was approved by the ethics review board at Osnabrück University, Germany. Descriptive statistics of the participants are given in Table S12.

Materials

The stimuli consisted of animations of five seconds in length made with Blender (Blender Online Community, 2018). Each animation depicted a car travelling over a hill. Immediately after the hill, the car encountered a dilemma situation. It could either stay on course and risk the lives of pedestrians on the road or swerve to the side. Depending on the scenario, swerving would direct the car either into a single pedestrian (on a road or a sidewalk) or the side of a passing freight train. The animations ended shortly before impact to avoid unnecessary distress for participants (Figure 4.4). To manipulate the perspective, each animation depicted a scenario from either a bird's-eye view; a first-person perspective of a pedestrian; or a first-person perspective of the car occupant (Figure 4.4).

Design

Two scenarios were investigated in this study (pedestrian versus pedestrian; car occupants versus pedestrians). While the two associated designs differed in important ways, the general framework was the same. Four different lives-at-risk situations were investigated; swerving always endangered a single life, but staying on course endangered from 1 to 4 lives, depending on the trial.

For the pedestrians versus single pedestrian scenario we employed a 2 (motorist-type) × 4 (perspective) × 2 (road-type) × 4 (lives-at-risk) mixed factorial design. There were two levels of motorist-type (self-driving car, human driver), and four of perspective (car occupant, pedestrian-straight-ahead, pedestrian-on-the-side, observer). All participants saw the two levels of road-type (split-road, road-with-sidewalk) and lives-at-risk (1 versus 1, 2 versus 1, 3 versus 1, and 4 versus 1). Motorist-type and perspective were manipulated between-participants, while road-type and lives-at-risk were manipulated within-participants. Thus, each participant witnessed all pedestrians versus single pedestrian scenario from a single perspective involving a single motorist-type.

For the pedestrians versus car occupant scenario we employed a 2 (motorist-type) × 3 (perspective) × 4 (lives-at-risk) mixed factorial design. Motorist-type had two levels (self-driving, human-driven) and perspective had three levels (car occupant, pedestrian straight ahead, observer). All participants saw all four levels of lives-at-risk (1v1, 2v1, 3v1 and 4v1). Motorist-type and perspective were manipulated between-participants, while lives-at-risk was manipulated within-participants. Thus, each participant witnessed all occupant versus pedestrian dilemmas from a single perspective involving a single motorist-type.

Procedure

Participants were given a link to an animation-based online survey, created and hosted on LabVanced, an online platform for social science experiments (Finger et al., 2017). Upon starting the study, participants were randomly allocated into one of the eight conditions described above, corresponding to the combinations of motorist-type and perspective in the larger design. Participants in observer and car occupant perspectives were presented both scenarios, as described above. However, the participants allocated to the pedestrian on-the-side perspective did not view the pedestrians versus car occupant scenario, as there was no corresponding viewpoint in these animations. A single trial consisted of a pair of animations depicting the same situation. One animation showed the car staying on course, the other showed it swerving to the side. The order of the two animations was counterbalanced across trials. After viewing the pair of animations, images of the final frames of each animation were presented side-by-side. Participants were asked to choose which of the two actions was more acceptable by clicking on the corresponding image. Throughout the trials, a textual notice reminded participants about both the perspective from which they are viewing the scenarios and the type of motorist depicted.

All experimental trials were completed in random order. The experiment always began with a control trial; participants who failed it were excluded. After the experimental block, participants completed a short questionnaire on demographics, driving experience, prior knowledge of self-driving cars and opinion toward them. Furthermore, participants were asked whether they identified more with the pedestrians or the car occupant while watching the animations with the question: “while watching the animations, which party did you most strongly identify with?” The options were: the car, the pedestrians.

Statistical analysis

As with the first study, statistical analyses were conducted in R (R Core Team, 2018) using *lme4* (Bates et al., 2015) for model fitting. Significance testing was performed using likelihood ratio tests with *afex* (Singmann et al., 2018) and *emmeans* (Lenth, 2018) was used for follow-up multiple comparisons on the estimated marginal means (*EMMs*).

Following the study design, the two scenarios were analysed individually. For both, we modelled the likelihood of choosing swerving to the side as more acceptable than staying on course based on lives-at-risk, road-type, perspective and motorist-type, using generalised linear mixed models with logit link functions. To control for individual differences, we implemented maximal random-effects structures as suggested by Barr et al., 2013 and Barr, 2013. In the pedestrian versus pedestrian dilemmas, due to convergence issues, the maximal random effects structure was replaced with a sub-maximal structure, without the random slope for lives-at-risk. The following covariates were included in all models: gender, age, knowledge of self-driving cars, and opinion of self-driving cars.

4.4.2 Study 2: Results

Similar to Study 1, we first performed a manipulation check to determine if the perspective from which participants viewed the scenarios affected with which party they identified most strongly. The omnibus goodness-of-fit test was significant, $\chi^2(24, N = 350) = 60.66$, $p < .0001$. The majority of participants in the pedestrian or car occupant perspectives identified most strongly with the corresponding perspective. Approximately equal numbers of participants in the observer perspective identified with car occupants and pedestrians (Tables S13 and S14). Thus, the manipulation check indicates that in most cases participants identify with the allocated perspective and the observer perspective was not biased.

Next, we investigated the effects of the perspective, motorist-type, road-type and lives-at-risk on judgements on the pedestrian versus pedestrian dilemma (Table 4.34). There was a significant main effect of lives-at-risk ($p < .0001$). With increasing imbalance of the number of pedestrians endangered, the probability of swerving changed steeply from close to 0.0 to nearly 1.0. Further, we observed a significant main effect of road-type ($p = .0002$). Participants tended to perceive swerving as more acceptable when swerving onto another road (probability = 0.88) than onto a sidewalk (probability = 0.76), odds-ratio = 2.50 (Table S16).

Generally, increases in lives-at-risk were positively associated with the probability of preferring to swerve (the more lives at risk by staying, the higher the probability of preferring to swerve). However, the nuances of this relationship depended on perspective and motorist-type and their interaction (Table S17). Lives-at-risk interacted with perspective ($p < .0001$) and we observed a three-way interaction of lives-at-risk × perspective × motorist-type (p = .0152) (Figure 4.5). Specifically, comparing the case of 2v1 lives-at-risk, the probability of swerving was higher for car occupant and observer perspectives than for pedestrian perspectives. Furthermore, there was a difference in the case of 2v1 lives-at-risk from the pedestrian-straight-ahead perspective between human driver and self-driving car. Follow up comparisons of the lives-at-risk × perspective × motorist-type interaction indicated that in all except one condition, acceptability of swerving was significantly higher at 2v1 compared to 1v1 lives-at-risk, all $p < .0001$ (Table S18). The exception to this was for participants who judged human drivers from the perspective of pedestrians-straight-ahead. In their case, this increase only occurred at 3v1 lives-at-risk (odds-ratio = 31.67, $p < .0001$). This indicates that perspective may affect how human drivers' actions are perceived, and at which point it is considered appropriate for them to intervene.

In the next scenario, car occupants were weighed against pedestrians. There was a significant main effect of lives-at-risk ($p < .0001$) and a significant lives-at-risk × perspective × motorist-type interaction (p = .0288) (Table 4.34). Preferring to swerve was generally positively associated with lives-at-risk. In all conditions, swerving was significantly more acceptable at 4v1 lives-at-risk compared to 1v1 lives-at-risk (all $p < .05$). For judgements on self-driving cars this increase occurred between 1v1 and 2v1 lives-at-risk, while for judgements on human drivers, this point depended on perspective. For those in the car occupant perspective, there was no significant difference between 1v1 and 2v1 lives-at-risk (p = .0604), but there was a significant difference between 1v1 and 3v1 conditions (odds-ratio = 68.02, p = .0001). For both observers and pedestrians, this occurred only after 4v1 lives-at-risk, odds-ratios = 20.42 (p = .0011) and 11.97 (p = .0136), respectively. However, in the latter case, this was due to the already high acceptability of swerving at 1v1 lives-at-risk (probability = 0.68). These results are depicted in Figure 4.6. Thus, moral judgements were rather similar in the case of self-driving cars, and were dependent on perspective only in the case of human drivers.

4.4.3 Study 2: Discussion

In this study we observed that increasing the number of people in the direct path of a car led to higher acceptability of swerving to endanger a single life. Generally, when two or more pedestrians were in danger, the probability of preferring to swerve was substantially higher than when there was only a single pedestrian in danger. This is in line with previous studies, reporting a high sensitivity of participants to the number of lives at risk. Further, we observe that swerving onto a sidewalk was less acceptable than swerving onto a connecting road. However, this effect was overshadowed by the preference to minimise the number of lives endangered. Additionally, we observed other differences between judgements on human drivers and self-driving cars. When swerving would endanger a pedestrian, there was general agreement between perspectives for self-driving cars to minimise the number of lives endangered. However, for human drivers, this was not the case. Those in the perspective of pedestrians in the direct path of a car only accepted a human driver swerving when three or more pedestrians would be otherwise endangered. All other perspectives considered it more acceptable when there were two pedestrians in the direct path of a car (Figure 4.5). When swerving would endanger car occupants, there was general agreement between perspectives on what self-driving cars should do. It was more acceptable for self-driving cars to minimise harm while protecting their occupants when all else was equal. However, there was disagreement between perspectives about which action was more acceptable for human drivers to take. Those in the observer perspective only considered it more acceptable for drivers to endanger themselves when faced with four pedestrians on the road. Conversely, those in the pedestrian perspective already considered it more acceptable for drivers to swerve when there was a single pedestrian at risk (Figure 4.6A). Similar to Study 1, this indicates a self-preservation effect for pedestrians, however only for judgements on human drivers.

4.4.4 General Discussion

In both studies, we found that judgements on self-driving cars do not seem to differ substantially from those on human drivers. In cases where there is a discrepancy, it seems to be due to a stronger preference for self-driving cars to minimise harm. Based on this result, it seems that people generally expect self-driving cars to follow the same traffic regulations as human drivers. The experiments revealed that differences between perspectives occur in situations where lives of car occupants are weighed against those of pedestrians. Results from Study 1 show that perspective seems to affect the acceptability of a car driving off a cliff: passengers are less likely to prefer swerving off a cliff than observers or pedestrians. Study 2 indicates disagreement between perspectives when considering at which point human drivers should intervene and endanger their own lives for the greater good. Additionally, perspective seems to affect confidence: people who observe a collision from a detached point of view seem to be less confident in their judgements. Although there are many commonalities in the judgements from different perspectives, the identified discrepancies should be taken into consideration in further research.

Results from our studies on moral judgement generally align with those from previous studies of moral action, in which participants were in the roles of drivers in similar dilemma scenarios (Faulhaber et al., 2019; Sütfeld et al., 2017). This indicates that the discrepancy between moral action and moral judgement shown by e.g. Francis et al., 2016 may not be extremely pronounced in driving-related dilemmas presented in virtual environments. Thus, previous studies on the topic should be considered equally relevant irrespective of whether they focused on moral judgement or action.

One of the more controversial aspects of introducing self-driving cars may concern the endangering of pedestrians on sidewalks. According to our results, pedestrians on a sidewalk seem to be offered more protection than pedestrians on the road when the numbers of lives at risk are equal (Figure 4.5). However, this protection is overshadowed by the preference to endanger fewer lives. This opposes prominent ethical guidelines such as those issued by the ethics commission of the German Federal Ministry of Transport and Digital Infrastructure, 2017, which states that non-involved parties (e.g. pedestrians on a sidewalk) should not be endangered. Similar divergence occurs when a dilemma involves clearly risking the lives of car occupants or children, as there is no general agreement between people's judgements on what is considered more acceptable. However, the guidelines state that personal features, such as age, should not be taken into consideration in unavoidable accident situations. While ethical guidelines are important to consider, another aspect to consider is legality. In research by S. Li, Zhang, et al., 2019 and Awad et al., 2018, the legal liability of different parties involved in a situation (for example whether pedestrians were crossing legally or not) was shown to affect judgements. However these studies did not consider the interplay between the type of motorist, perspective and legality, something that future research should aim to elucidate.

Our studies aimed to expand understanding of moral psychology in the context of artificial intelligence. This research assists in determining criteria that self-driving car decision making needs to meet in order to be commonly accepted. However, we want to stress that responses to simplified dilemma situations should not be the basis for legal or ethical regulations. Furthermore, in agreement with Nyholm, 2018a, and Keeling, 2017, we believe empirical research alone cannot answer the ethical question of how self-driving cars should be programmed to behave. Nevertheless, we believe the results provide insights into the public's preferences regarding the decision making of self-driving cars and potential conflicts that may arise. The results from our studies point to specific questions warranting further investigation and attention in the debate surrounding the introduction of self-driving cars. In particular, these relate to the lack of agreement regarding specific dilemmas, apparent discrepancies between public opinion and ethical guidelines, the effects of perspective, the identified self-preservation effect and the albeit slight differences between judgements on self-driving cars and human drivers. These findings all highlight issues with creating decision making algorithms that attempt to simultaneously consider intuitions, ethical guidelines and legal regulations.

4.4.5 Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

4.4.6 Author Contributions

PK, GP and AS conceived of the initial research idea. NK, FNN, MP and JZ participated in planning the research and designing the experiments. PK, GP and AS gave feedback to the experimental designs. NK, FNN and MP participated in the creation of the materials. NK, FNN, MP and JZ participated in data collection. NK and MP analysed the collected data. NK, FNN, MP and JZ interpreted the results and wrote the manuscript. PK, GP and AS provided feedback and edited the manuscript.

4.4.7 Funding

We gratefully acknowledge financial support by the European Commission (H2020 FETPROACT-2014, SEP-21014273, socSMCs, ID: 641321, PK), by Deutsche

Forschungsgemeinschaft (DFG) and the Open Access Publishing Fund of Osnabrück University.

4.4.8 Acknowledgments

This paper is based on the work done in a student-run research project. The authors gratefully thank Jean-Philipp Almstedt, Linus Edelkott, David Finger, Kimberly Gerbaulet, Gayane Ghazaryan, Anastasia Mukhina, Iryna Ruda and Robert Sartorius for their valuable contributions to the project.

4.4.9 Supplemental Data: Extended procedure descriptions

Study 1

Participants were randomly assigned to one of the eight conditions corresponding to the combinations of perspective and motorist-type (e.g. observer & human driver; car occupant & self-driving car). Participants of the smaller and larger pedestrian groups shared the same pedestrians versus car occupants trials as there was only one pedestrian group involved in those scenarios.

Participants were asked to observe the environment carefully before pressing a button to begin each animation. There was a pause after the animation and sound stopped and participants could freely inspect the scene again before pressing a button to continue. The order of the animations within a trial was randomised for each participant. After viewing a pair of animations, participants could replay the pair in the same order as originally shown or continue to the response screen. There was no limit on the number of times the animations could be replayed before responding.

The response screen consisted of side-by-side images of the final frames of each animation. Participants were asked to choose which of the two actions of the motorist they considered to be more acceptable by selecting the corresponding outcome image. Depending on the motorist-type, the question either mentioned that the actions were taken by a person driving the car or a self-driving car. After making the judgement, participants indicated how confident they were with their choice on a scale from 0 (not confident at all) to 100 (very confident).

Participants completed a practice trial and a control trial before the experimental trials. The practice trial involved a scenario with a single pedestrian on one side of the road and a group of five pedestrians on the other side. This task allowed the participants to become accustomed to the immersion of VR and the controls.

The control trial consisted of a single pedestrian on one side and empty road on the other. The intention of the control trial was to check whether participants considered swerving to empty road more acceptable than endangering a single person. Participants were excluded if they considered swerving to the pedestrian as more acceptable than swerving to the empty side of the road, as this behaviour indicated a tendency to risk a life for no purpose or a misunderstanding of the task.

To separate each trial, distraction tasks were presented between them. These consisted of simple mathematical equations which participants had to indicate whether they were correct or incorrect. The distraction trial lasted 20 seconds regardless of the number of responses in that time. There was no visible countdown in order to avoid stress. The distraction trial ended once the participant responded to the final question displayed within the 20 seconds. No data from the distraction task was recorded or analysed.

After completing the practice and control trials, participants completed the six experimental trials in random order, separated by distraction tasks.

After the experiment ended, participants completed a short questionnaire on demographics, driving experience, prior knowledge of self-driving cars and their attitudes towards them. Furthermore, as a manipulation check, participants reported which party in the situation they identified most with while watching the animations: the pedestrians, the car occupants or the observer. Finally, they were asked whether the motorist was a human driver or a self-driving car. Participants in the self-driving car condition that could not recollect their motorist-type were excluded, since this indicated they did not fully understand the instructions. We chose to not equally exclude the participants of the human driver condition as they were not briefed about different motorist-type conditions and only presented the term 'self-driving car' in the post-study questionnaire. On completion of the questionnaire, participants were debriefed. The entire procedure took approximately 25–30 minutes.

Figure 4.1: The screenshot depicts the decision screen presented to participants in the virtual reality study after have seen the animations. Two images of each choice were shown to show the scene from different angles. The scenario depicted here is a practice trial involving the choice between endangering a group of pedestrians or a single pedestrian. **(Left)** Self-driving car (with a relaxing second pedestrian) is veering towards one pedestrian; **(B)** Self-driving car is veering towards a group of pedestrians. Instructions translated from German: "Which option of the autonomous vehicle is more acceptable? (Left/Right) – To alternate between options, move the analog stick 'left' or 'right' respectively. Subsequently, continue with 'A'."

Study 2

Participants were given a link to an animation-based online survey, created and hosted on LabVanced, an online platform for social science experiments. Upon starting the study, participants were randomly allocated into one of the eight conditions mentioned before, corresponding to the combinations of motorist-type and perspective in the larger design. Participants were presented scenario situations of both types, as described above. However, the participants allocated to the pedestrian on the side perspective did not view occupant-pedestrian scenarios, as there was no corresponding viewpoint in these animations.

A single trial consisted of a pair of animations depicting the same moral scenario situation. One animation showed the car staying on course, the other showed the car swerving to the side. The order of the two videos was counterbalanced across trials. After viewing the pair of animations, images of the final frames of each animation were presented side-by-side. Participants were asked to choose which of the two actions was more acceptable by clicking on the corresponding image. Depending on the motorist-type, the question either stated the actions were made by a person driving the car or a self-driving car. Below the question was a reminder of which perspective the animations were viewed from. Participants could click a button to replay the pair of animations in the same order as originally shown. There was no limit to the number of times the animations could be replayed before making a choice nor the amount of time participant needed to answer the up coming question.

The first trial was always a control task. The presented scenario involved a single pedestrian in the path of the car and a clear road to the side. Participants were excluded if they considered “stay” as more acceptable than “swerve”, as this indicated either a tendency to risk a life for no purpose or a misunderstanding of the task.

All participants completed eight trials of pedestrian vs. pedestrian scenarios, corresponding to the combinations of the levels of the road-type and lives-at-risk. An additional four trials were occupant vs. pedestrian scenarios, corresponding to the four different levels of lives-at-risk. However, participants assigned to the pedestrian on the side perspective only completed the trials relating to the pedestrian vs. pedestrian scenario. This was due to no corresponding viewpoint in the occupant vs. pedestrian scenario. Thus, participants completed a total of either eight or twelve trials. All experimental trials were completed in random order.

After the experimental block, participants completed a short questionnaire on demographics, driving experience, prior knowledge of self-driving cars and attitudes toward them. Furthermore, participants were asked whether they identified more with the pedestrians or the car occupant while watching the animations. The entire study procedure took approximately five minutes.

4.4.10 Study 1 Supplementary Tables

Table 4.1: Descriptive statistics of the participants in Study 1.

motorist: human driver				
	observer	passenger	pedestrian (smaller group)	pedestrian (larger group)
gender (male:female)	12:11	12:11	12:11	12:11
age (M, SD)	23.57 (4.41)	21.87 (3.31)	23.35 (6.12)	23.04 (2.99)
driving experience				
no experience	6	3	3	4
1–5 years	6	15	16	9
5–10 years	10	4	2	9
≥10 years	1	1	2	1
education				
no higher education achieved	18	21	20	13
undergraduate education	3	2	3	10
post-graduate education	2	0	0	0
opinion of self-driving cars (M, SD)	3.52 (1.04)	3.48 (1.24)	3.48 (1.31)	3.48 (1.27)
visual acuity				
normal vision	17	13	15	9
corrected vision	5	6	6	12
uncorrected vision	1	4	2	2

motorist: self-driving car				
	observer	passenger	pedestrian (smaller group)	pedestrian (larger group)
gender (male:female)	12:11	12:11	12:11	12:11
age (M, SD)	23.30 (3.72)	21.70 (2.36)	22.43 (2.73)	24.26 (6.08)
driving experience				
no experience	4	4	2	3
1–5 years	9	14	13	11
5–10 years	8	5	7	6
≥10 years	2	0	1	3
education				
no higher education achieved	14	19	21	15
undergraduate education	6	4	2	6
post-graduate education	3	0	0	2
opinion of self-driving cars (M, SD)	3.30 (1.18)	3.43 (1.27)	3.52 (1.20)	3.3 (1.02)
visual acuity				
normal vision	13	13	15	10
corrected vision	9	7	7	10
uncorrected vision	1	3	1	3

Table 4.2: Contingency table for manipulation check (Study 1).

	identification		
perspective	observer	passenger	pedestrian
observer	31	10	5
passenger	8	35	3
pedestrian	33	3	56

Table 4.3: Follow up comparisons for manipulation check (Study 1).

	comparison	p	
1	observer vs. passenger	$< .01$	**
2	observer vs. pedestrian	$< .01$	**
3	passenger vs. pedestrian	$< .01$	**

Note: * $p < .05$, ** $p < .01$, *** $p < .001$.

4.5 Study 2 supplementary tables

contrast	odds ratio	SE	df	z lives-at-risk	p	
perspective = car occupant						
motorist = self-driving car						
1v1 / 2v1	0.0066	0.0043	Inf	−7.715	<.0001	***
1v1 / 3v1	0.0018	0.0013	Inf	−8.377	<.0001	***
1v1 / 4v1	0.0026	0.0019	Inf	−8.257	<.0001	***
2v1 / 3v1	0.2661	0.1588	Inf	−2.218	.1182	
2v1 / 4v1	0.3893	0.2216	Inf	−1.657	.3466	
3v1 / 4v1	1.4633	0.9111	Inf	0.611	.9285	
perspective = observer						
motorist = self-driving car						
1v1 / 2v1	0.0097	0.0057	Inf	−7.869	<.0001	***
1v1 / 3v1	0.0046	0.0030	Inf	−8.315	<.0001	***
1v1 / 4v1	0.0017	0.0013	Inf	−8.574	<.0001	***
2v1 / 3v1	0.4685	0.2592	Inf	−1.370	.5180	
2v1 / 4v1	0.1747	0.1108	Inf	−2.752	.0302	*

contrast	odds ratio	SE	df	z lives-at-risk	p	
3v1 / 4v1	0.3730	0.2421	Inf	−1.519	.4257	
perspective = pedestrian forward						
motorist = self-driving car						
1v1 / 2v1	0.0320	0.0188	Inf	−5.869	<.0001	***
1v1 / 3v1	0.0027	0.0020	Inf	−7.884	<.0001	***
1v1 / 4v1	0.0033	0.0024	Inf	−7.814	<.0001	***
2v1 / 3v1	0.0851	0.0518	Inf	−4.050	.0003	***
2v1 / 4v1	0.1040	0.0615	Inf	−3.830	.0007	***
3v1 / 4v1	1.2224	0.7812	Inf	0.314	.9893	
perspective = pedestrian side						
motorist = self-driving car						
1v1 / 2v1	0.0395	0.0228	Inf	−5.594	<.0001	***
1v1 / 3v1	0.0344	0.0202	Inf	−5.744	<.0001	***
1v1 / 4v1	0.0073	0.0051	Inf	−7.040	<.0001	***
2v1 / 3v1	0.8705	0.4623	Inf	−0.261	.9938	
2v1 / 4v1	0.1847	0.1140	Inf	−2.737	.0315	*
3v1 / 4v1	0.2122	0.1313	Inf	−2.506	.0590	
perspective = car occupant						
motorist = human						
1v1 / 2v1	0.0024	0.0016	Inf	−9.147	<.0001	***
1v1 / 3v1	0.0006	0.0005	Inf	−9.708	<.0001	***
1v1 / 4v1	0.0008	0.0006	Inf	−9.667	<.0001	***
2v1 / 3v1	0.2689	0.1532	Inf	−2.306	.0966	
2v1 / 4v1	0.3171	0.1762	Inf	−2.067	.1641	
3v1 / 4v1	1.1793	0.7115	Inf	0.273	.9929	
perspective = observer						
motorist = human						
1v1 / 2v1	0.0061	0.0040	Inf	−7.821	<.0001	***
1v1 / 3v1	0.0032	0.0022	Inf	−8.189	<.0001	***
1v1 / 4v1	0.0038	0.0026	Inf	−8.096	<.0001	***
2v1 / 3v1	0.5151	0.2996	Inf	−1.141	.6643	
2v1 / 4v1	0.6159	0.3528	Inf	−0.846	.8324	
3v1 / 4v1	1.1958	0.7140	Inf	0.299	.9907	
perspective = pedestrian forward						

contrast	odds ratio	SE	df	z lives-at-risk	p	
motorist = human						
1v1 / 2v1	0.4276	0.1922	Inf	−1.890	.2324	
1v1 / 3v1	0.0316	0.0174	Inf	−6.277	<.0001	***
1v1 / 4v1	0.0119	0.0075	Inf	−7.036	<.0001	***
2v1 / 3v1	0.0738	0.0386	Inf	−4.981	<.0001	***
2v1 / 4v1	0.0279	0.0168	Inf	−5.943	<.0001	***
3v1 / 4v1	0.3775	0.2199	Inf	−1.672	.3383	
perspective = pedestrian side						
motorist = human						
1v1 / 2v1	0.0642	0.0370	Inf	−4.768	<.0001	***
1v1 / 3v1	0.0390	0.0237	Inf	−5.339	<.0001	***
1v1 / 4v1	0.0330	0.0205	Inf	−5.503	<.0001	***
2v1 / 3v1	0.6081	0.3496	Inf	−0.865	.8229	
2v1 / 4v1	0.5148	0.3003	Inf	−1.138	.6657	
3v1 / 4v1	0.8465	0.5028	Inf	−0.281	.9923	

Note: * $p < .05$, ** $p < .01$, *** $p < .001$.

lives-at-risk	P(“swerve”)	SE	df	lower asymptotic 95% CI	upper asymptotic 95% CI
perspective = car					
motorist = self-driving					
1v1	0.1275	0.0686	Inf	0.0418	0.3284
2v1	0.9569	0.0286	Inf	0.8506	0.9886
3v1	0.9881	0.0092	Inf	0.9472	0.9974
4v1	0.9827	0.0127	Inf	0.9288	0.9960
perspective = observer					
motorist = self-driving					
1v1	0.1089	0.0591	Inf	0.0357	0.2873
2v1	0.9262	0.0454	Inf	0.7736	0.9787
3v1	0.9640	0.0246	Inf	0.8695	0.9908
4v1	0.9863	0.0107	Inf	0.9387	0.9970
perspective = pedestrian (forward)					
motorist = self-driving					
1v1	0.1153	0.0637	Inf	0.0369	0.3072
2v1	0.8030	0.0981	Inf	0.5473	0.9322
3v1	0.9796	0.0151	Inf	0.9165	0.9952

lives-at-risk	P("swerve")	SE	df	lower asymptotic 95% CI	upper asymptotic 95% CI
4v1	0.9751	0.0178	Inf	0.9028	0.9940
perspective = pedestrian (side)					
motorist = self-driving					
1v1	0.1737	0.0922	Inf	0.0563	0.4255
2v1	0.8417	0.0912	Inf	0.5816	0.9532
3v1	0.8593	0.0838	Inf	0.6108	0.9597
4v1	0.9664	0.0252	Inf	0.8624	0.9925
perspective = car					
motorist = human					
1v1	0.0339	0.0207	Inf	0.0101	0.1080
2v1	0.9361	0.0384	Inf	0.8063	0.9810
3v1	0.9820	0.0128	Inf	0.9298	0.9956
4v1	0.9788	0.0147	Inf	0.9200	0.9946
perspective = observer					
motorist = human					
1v1	0.0540	0.0320	Inf	0.0165	0.1630
2v1	0.9028	0.0587	Inf	0.7145	0.9718
3v1	0.9475	0.0351	Inf	0.8188	0.9863
4v1	0.9378	0.0406	Inf	0.7940	0.9833
perspective = pedestrian (forward)					
motorist = human					
1v1	0.2078	0.1006	Inf	0.0734	0.4649
2v1	0.3802	0.1426	Inf	0.1579	0.6676
3v1	0.8926	0.0635	Inf	0.6941	0.9682
4v1	0.9565	0.0301	Inf	0.8419	0.9891
perspective = pedestrian (side)					
motorist = human					
1v1	0.2694	0.1261	Inf	0.0951	0.5641
2v1	0.8518	0.0868	Inf	0.5988	0.9568
3v1	0.9043	0.0612	Inf	0.7027	0.9742
4v1	0.9178	0.0541	Inf	0.7324	0.9785

Note: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4.4: Estimated marginal means for self-reported confidence in judgements — children versus adults scenario (Study 1).

perspective	EMM	SE	df	lower 95% CI	upper 95% CI
judgement = endanger larger group (adults)					
observer	37.4348	6.9340	186.98	23.7558	51.1138
passenger	50.6119	7.7075	181.57	35.4042	65.8196
pedestrian (smaller group)	55.8615	8.0671	181.26	39.9441	71.7790
pedestrian (larger group)	53.7088	7.2938	181.68	39.3174	68.1002
judgement = endanger smaller group (children)					
observer	37.3025	7.3825	218.66	22.7525	51.8525
passenger	46.8662	8.2803	202.90	30.5397	63.1928
pedestrian (smaller group)	69.7271	8.5895	206.26	52.7926	86.6615
pedestrian (larger group)	39.8304	8.0855	226.03	23.8978	55.7630

Table 4.5: Follow up comparisons for self-reported confidence in children versus adults scenario (Study 1).

contrast	estimate	SE	df	t	p	
judgement = endanger larger group (adults)						
observer – passenger	−13.1771	7.2467	211.35	−1.818	.2676	
observer – pedestrian (smaller group)	−18.4267	7.0260	205.78	−2.623	.0459	*
observer – pedestrian (larger group)	−16.2740	6.8486	207.04	−2.376	.0849	
passenger – pedestrian (smaller group)	−5.2496	7.1644	214.91	−0.733	.8838	
passenger – pedestrian (larger group)	−3.0969	7.1190	207.64	−0.435	.9724	
pedestrian (smaller group) – pedestrian (larger group)	2.1528	6.9085	200.55	0.312	.9895	
judgement = endanger smaller group (children)						
observer – passenger	−9.5637	8.1973	260.82	−1.167	.6484	
observer – pedestrian (smaller group)	−32.4246	8.3792	276.43	−3.870	.0008	***
observer – pedestrian (larger group)	−2.5279	8.3234	286.57	−0.304	.9903	
passenger – pedestrian (smaller group)	−22.8608	8.2415	258.10	−2.774	.0301	*
passenger – pedestrian (larger group)	7.0358	8.5637	273.21	0.822	.8442	
pedestrian (smaller group) – pedestrian (larger group)	29.8966	8.7240	285.26	3.427	.0039	**

Note: * $p < .05$, ** $p < .01$, *** $p < .001$.

4.5.1 Data Availability Statement

The datasets for the studies are available as supplementary materials.

Table 4.6: Estimated marginal means for self-reported confidence in sidewalk versus road scenario (Study 1).

judgement	EMM	SE	df	lower 95% CI	upper 95% CI
endanger smaller group (sidewalk)	68.8838	5.8744	162.94	57.2840	80.4837
endanger larger group (road)	60.9278	6.6278	203.52	47.8599	73.9957

Table 4.7: Follow up comparisons for self-reported confidence in sidewalk versus road scenario (Study 1).

contrast	estimate	SE	df	t	p	
endanger smaller (sidewalk) – endanger larger (road)	7.9561	3.7212	337.76	2.138	.0332	*

Note: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4.8: Estimated probability of preferring to endanger car occupants in car occupants versus pedestrians scenario (Study 1).

trial	P(“endanger car occupants”)	SE	df	lower 95% CI	upper 95% CI
parked van	0.9145	0.0403	Inf	0.7956	0.9671
cliff edge	0.5531	0.1078	Inf	0.3448	0.7442

contrast	odds ratio	SE	df	z	p	
parked van / cliff edge	8.6435	2.9567	Inf	6.305	<.0001	***

Table 4.9: Follow up comparisons for estimated probability of preferring to endanger car occupants in car occupants versus pedestrians scenario (Study 1).

Note: * $p < .05$, ** $p < .01$, *** $p < .001$.

trial	EMM	SE	df	lower 95% CI	upper 95% CI
parked van	64.5887	5.5074	205.55	53.7304	75.4470
cliff edge	48.4315	5.2839	188.95	38.0085	58.8546

Table 4.10: Estimated marginal means for self-reported confidence in car occupants versus pedestrians scenario (Study 1).

trial	EMM	SE	df	lower 95% CI	upper 95% CI
judgement = endanger pedestrians					
parked van	55.2472	15.9957	330.19	23.7809	86.7134
cliff	50.3588	7.2760	259.22	36.0312	64.6864
judgement = endanger car occupants					
parked van	75.2332	5.2659	192.35	64.8470	85.6195
cliff	47.8217	5.8862	237.91	36.2260	59.4175

Table 4.11: Estimated Marginal Means for self-reported confidence separated by trial and judgement in car occupants versus pedestrians scenario (Study 1).

Note that for the parked van trial, there were no observers preferring to endanger pedestrians, so the EMMs for that trial only consider the other perspectives.

Table 4.12: Descriptive statistics of the groups in car occupants versus pedestrians scenario (Study 2).

age group	gender	knowledge of self-driving cars	driving experience	country
18–29: 142	female:151	no: 23	0: 52	Germany: 55
30–39: 43	male: 130	yes:259	5: 54	Armenia: 30
40–49: 18	NA's: 1		6–10: 35	Australia: 25
50–59: 18			10+: 95	Russia: 10
60–69: 24			NA's: 37	(Other): 19
70–79: 3				NA's: 142
80–89: 1				

Table 4.13: Contingency table for manipulation check (Study 2).

	identification	
	car occupant	pedestrian
perspective		
car occupant	73	28
observer	53	42
pedestrian (in car's path)	21	65
pedestrian (to side of car's path)	23	63

Table 4.14: Follow up comparisons for manipulation check (Study 2).

comparison	p	
car vs. observer	.02	
car vs. pedestrian (in car's path)	<.01	**
car vs. pedestrian (to side of car's path)	<.01	**
observer vs. pedestrian (in car's path)	<.01	
observer vs. pedestrian (to side of car's path)	<.01	**
pedestrian (in car's path) vs. pedestrian (to side of car's path)	.86	

Note: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4.15: Follow up comparisons for sidewalk versus road scenarios (Study 2).

contrast	estimate	SE	df	z	p	
road − sidewalk	0.8677	0.2576	Inf	3.369	.0008	***

Note: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4.16: Estimated marginal means for acceptability of swerving in sidewalk versus road scenarios (Study 2).

scenario	EMM	SE	df	lower asymptotic 95% CI	upper asymptotic 95% CI
road	2.0196	0.3237	Inf	1.3852	2.6540
sidewalk	1.1519	0.3579	Inf	0.4504	1.8534

Table 4.17: Follow-up comparisons of lives-at-risk in pedestrians versus pedestrians scenario (Study 2).

contrast	odds ratio	SE	df	z lives-at-risk	p	
1v1 / 2v1	0.0206	0.0054	Inf	−14.706	<.0001	***
1v1 / 3v1	0.0061	0.0019	Inf	−16.347	<.0001	***
1v1 / 4v1	0.0043	0.0014	Inf	−16.739	<.0001	***
2v1 / 3v1	0.2950	0.0606	Inf	−5.940	<.0001	***
2v1 / 4v1	0.2092	0.0452	Inf	−7.241	<.0001	***
3v1 / 4v1	0.7089	0.1549	Inf	−1.574	.3933	

Note: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4.18: Follow-up comparisons of lives-at-risk in pedestrians versus pedestrians scenario.

contrast	P(*swerve*)	SE	df	z lives-at-risk	p	
1v1 / 2v1	0.0206	0.0054	Inf	−14.706	<.0001	***
1v1 / 3v1	0.0061	0.0019	Inf	−16.347	<.0001	***
1v1 / 4v1	0.0043	0.0014	Inf	−16.739	<.0001	***
2v1 / 3v1	0.2950	0.0606	Inf	−5.940	<.0001	***
2v1 / 4v1	0.2092	0.0452	Inf	−7.241	<.0001	***
3v1 / 4v1	0.7089	0.1549	Inf	−1.574	.3933	

Note: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4.19: Estimated marginal means for Lives-At-Risk factor in pedestrians versus pedestrians scenario (Study 2).

lives-at-risk	P(swerve)	SE	df	lower asymptotic 95% CI	upper asymptotic 95% CI
1v1	.1170	0.0346	Inf	0.0642	0.2036
2v1	.8653	0.0397	Inf	0.7672	0.9260
3v1	.9561	0.0153	Inf	0.9140	0.9780
4v1	.9685	0.0114	Inf	0.9365	0.9846

Table 4.20: Preference of swerving for lives-at-risk × perspective interaction in pedestrians versus pedestrians scenario

lives-at-risk	P (“swerve”)	SE	df	lower asymptotic 95% CI	upper asymptotic 95% CI
perspective = car					
1v1	0.0668	0.0304	Inf	0.0268	0.1571
2v1	0.9474	0.0260	Inf	0.8664	0.9804
3v1	0.9854	0.0084	Inf	0.9554	0.9953
4v1	0.9809	0.0106	Inf	0.9441	0.9936
perspective = observer					
1v1	0.0771	0.0339	Inf	0.0317	0.1753
2v1	0.9152	0.0394	Inf	0.7997	0.9669
3v1	0.9565	0.0224	Inf	0.8846	0.9844
4v1	0.9705	0.0161	Inf	0.9161	0.9900
perspective = pedestrian fwd					
1v1	0.1561	0.0616	Inf	0.0688	0.3163
2v1	0.6126	0.1088	Inf	0.3917	0.7952
3v1	0.9523	0.0240	Inf	0.8761	0.9825
4v1	0.9671	0.0174	Inf	0.9098	0.9884
perspective = pedestrian side					
1v1	0.2178	0.0837	Inf	0.0961	0.4216
2v1	0.8468	0.0680	Inf	0.6643	0.9392
3v1	0.8837	0.0549	Inf	0.7272	0.9559
4v1	0.9472	0.0285	Inf	0.8543	0.9821

Table 4.23: Follow-up comparisons for lives-at-risk × perspective interaction for pedestrians versus pedestrians scenario.

contrast	odds ratio	SE	df	z lives-at-risk	p	
lives-at-risk= 1v1						
car occupant / observer	0.8579	0.4594	Inf	−0.286	.9918	
car occupant / pedestrian (in car's path)	0.3873	0.2149	Inf	−1.710	.3186	
car occupant / pedestrian (to side of car's path)	0.2573	0.1505	Inf	−2.320	.0935	
observer / pedestrian (in car's path)	0.4514	0.2491	Inf	−1.441	.4735	
observer / pedestrian (to side of car's path)	0.2999	0.1748	Inf	−2.066	.1644	
pedestrian (in car's path) / pedestrian (to side of car's path)	0.6643	0.3761	Inf	−0.722	.8882	
lives-at-risk = 2v1						
car occupant / observer	1.6701	0.9872	Inf	0.868	.8216	
car occupant / pedestrian (in car's path)	11.3976	6.7101	Inf	4.133	.0002	***
car occupant / pedestrian (to side of car's path)	3.2605	2.0813	Inf	1.852	.2493	
observer / pedestrian (in car's path)	6.8247	3.9824	Inf	3.291	.0055	**
observer / pedestrian (to side of car's path)	1.9523	1.2394	Inf	1.054	.7176	
pedestrian (in car's path) / pedestrian (to side of car's path)	0.2861	0.1710	Inf	−2.094	.1550	
lives-at-risk = 3v1						
car / observer	3.0675	2.0122	Inf	1.709	.3190	
car / pedestrian (in car's path)	3.3778	2.2723	Inf	1.809	.2688	
car / pedestrian (to side of car's path)	8.8704	6.1554	Inf	3.145	.0090	**
observer / pedestrian (in car's path)	1.1011	0.7140	Inf	0.149	.9988	
observer / pedestrian (to side of car's path)	2.8917	1.9156	Inf	1.603	.3769	
pedestrian (in car's path) / pedestrian (to side of car's path)	2.6261	1.7060	Inf	1.486	.4458	
lives-at-risk = 4v1						
car / observer	1.5594	1.0308	Inf	0.672	.9077	
car / pedestrian (in car's path)	1.7467	1.1731	Inf	0.830	.8400	
car / pedestrian (to side of car's path)	2.8616	2.0073	Inf	1.499	.4381	
observer / pedestrian (in car's path)	1.1201	0.7601	Inf	0.167	.9983	
observer / pedestrian (to side of car's path)	1.8351	1.2945	Inf	0.861	.8251	
pedestrian (in car's path) / pedestrian (to side of car's path)	1.6383	1.1199	Inf	0.722	.8883	

Note: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4.24: Follow-up comparisons for lives-at-risk in car occupants versus pedestrians scenario (Study 2).

contrast	odds ratio	SE	df	z lives-at-risk	p	
1v1 / 2v1	0.2037	0.0673	Inf	−4.813	< .0001	***
1v1 / 3v1	0.0540	0.0220	Inf	−7.164	< .0001	***
1v1 / 4v1	0.0299	0.0133	Inf	−7.871	< .0001	***
2v1 / 3v1	0.2650	0.0924	Inf	−3.807	.0008	***
2v1 / 4v1	0.1468	0.0559	Inf	−5.038	<.0001	***
3v1 / 4v1	0.5541	0.2037	Inf	−1.606	.3751	

Note: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4.25: Preference of swerving for different lives-at-risk in car occupants versus pedestrians scenario (Study 2).

lives-at-risk	P(swerve)	SE	df	lower asymptotic 95% CI	upper asymptotic 95% CI
1v1	0.3442	0.1330	Inf	0.1419	0.6249
2v1	0.7204	0.1215	Inf	0.4413	0.8937
3v1	0.9067	0.0541	Inf	0.7349	0.9715
4v1	0.9461	0.0337	Inf	0.8276	0.9847

Table 4.26: Follow up comparison of motorist-type × Lives-At-Risk factor interaction (Study 2).

contrast	odds ratio	SE	df	z	p	
motorist = self-driving						
2v1 / 1v1	9.5327	4.5049	Inf	4.771	<.0001	***
3v1 / 1v1	35.3768	19.8466	Inf	6.357	<.0001	***
4v1 / 1v1	50.2008	30.0540	Inf	6.541	<.0001	***
motorist = human						
2v1 / 1v1	2.1436	0.8766	Inf	1.865	0.1578	
3v1 / 1v1	6.6640	3.0346	Inf	4.165	0.0001	***
4v1 / 1v1	16.1888	8.2770	Inf	5.446	<.0001	***

Note: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4.27: Preference for swerving based on motorist-type (Study 2).

lives-at-risk	P(“swerve”)	SE	df	lower asymptotic 95% CI	upper asymptotic 95% CI
motorist = self-driving					
1v1	0.2913	0.1331	Inf	0.1041	0.5924
2v1	0.7967	0.1085	Inf	0.5131	0.9358
3v1	0.9357	0.0435	Inf	0.7794	0.9836
4v1	0.9538	0.0329	Inf	0.8267	0.9889
motorist = human					
1v1	0.4308	0.1621	Inf	0.1717	0.7344
2v1	0.6187	0.1570	Inf	0.3058	0.8567
3v1	0.8346	0.0953	Inf	0.5659	0.9513
4v1	0.9246	0.0506	Inf	0.7475	0.9807

Table 4.28: Follow up comparisons for lives-at-risk × perspective× motorist-type in pedestrian versus car occupant scenario (Study 2).

contrast	odds ratio	SE	df	z	p	
perspective = car						
motorist = self-driving						
2v1 / 1v1	8.2638	6.6516	Inf	2.624	.0244	*
3v1 / 1v1	15.8926	13.7236	Inf	3.203	.0040	**
4v1 / 1v1	33.9867	32.2364	Inf	3.717	.0006	***
perspective = observer						
motorist = self-driving						
2v1 / 1v1	9.1710	7.1300	Inf	2.850	.0124	
3v1 / 1v1	35.0391	31.8736	Inf	3.910	.0003	***
4v1 / 1v1	12.3135	9.8459	Inf	3.140	.0049	**
perspective = pedestrian						
motorist = self-driving						
2v1 / 1v1	14.8127	12.6942	Inf	3.145	.0048	**
3v1 / 1v1	101.2472	104.1660	Inf	4.488	<.0001	***
4v1 / 1v1	462.8984	576.8877	Inf	4.925	<.0001	***
perspective = car						
motorist = human						
2v1 / 1v1	5.7257	4.3716	Inf	2.285	.0604	
3v1 / 1v1	68.0244	69.9305	Inf	4.105	.0001	***
4v1 / 1v1	29.5189	27.3060	Inf	3.659	.0007	***
perspective = observer						
motorist = human						
2v1 / 1v1	2.7827	2.0350	Inf	1.399	.3644	
3v1 / 1v1	3.6255	2.6863	Inf	1.738	.2029	
4v1 / 1v1	20.4203	17.2913	Inf	3.562	.0011	**
perspective = pedestrian						
motorist = human						
2v1 / 1v1	0.7835	0.5459	Inf	−0.350	.9496	
3v1 / 1v1	2.9083	2.1650	Inf	1.434	.3453	
4v1 / 1v1	11.9686	10.5255	Inf	2.823	.0136	*

Note: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4.29: Estimated marginal means for lives-at-risk × perspective × motorist-type interaction in car occupants versus pedestrians scenario

contrast	odds ratio	SE	df	z lives-at-risk	p	
perspective = car occupant						
motorist = self-driving car						
1v1 / 2v1	0.1210	0.0974	Inf	−2.624	.0432	*
1v1 / 3v1	0.0629	0.0543	Inf	−3.203	.0074	**
1v1 / 4v1	0.0294	0.0279	Inf	−3.717	.0012	**
2v1 / 3v1	0.5200	0.4258	Inf	−0.799	.8551	
2v1 / 4v1	0.2431	0.2131	Inf	−1.613	.3710	
3v1 / 4v1	0.4676	0.4111	Inf	−0.865	.8231	
perspective = observer						
motorist = self-driving car						
1v1 / 2v1	0.1090	0.0848	Inf	−2.850	.0227	*
1v1 / 3v1	0.0285	0.0260	Inf	−3.910	.0005	***
1v1 / 4v1	0.0812	0.0649	Inf	−3.140	.0092	**
2v1 / 3v1	0.2617	0.2225	Inf	−1.577	.3918	
2v1 / 4v1	0.7448	0.5733	Inf	−0.383	.9809	
3v1 / 4v1	2.8456	2.4293	Inf	1.225	.6109	
perspective = pedestrian						
motorist = self-driving car						
1v1 / 2v1	0.0675	0.0579	Inf	−3.145	.0090	**
1v1 / 3v1	0.0099	0.0102	Inf	−4.488	<.0001	***
1v1 / 4v1	0.0022	0.0027	Inf	−4.925	<.0001	***
2v1 / 3v1	0.1463	0.1281	Inf	−2.195	.1245	
2v1 / 4v1	0.0320	0.0350	Inf	−3.149	.0089	**
3v1 / 4v1	0.2187	0.2310	Inf	−1.439	.4750	
perspective = car occupant						
motorist = human						
1v1 / 2v1	0.1747	0.1333	Inf	−2.285	.1014	
1v1 / 3v1	0.0147	0.0151	Inf	−4.105	.0002	***
1v1 / 4v1	0.0339	0.0313	Inf	−3.659	.0014	
2v1 / 3v1	0.0842	0.0801	Inf	−2.602	.0458	
2v1 / 4v1	0.1940	0.1673	Inf	−1.902	.2272	
3v1 / 4v1	2.3044	2.1254	Inf	0.905	.8021	
perspective = observer						
motorist = human						
1v1 / 2v1	0.3594	0.2628	Inf	−1.399	.4997	
1v1 / 3v1	0.2758	0.2044	Inf	−1.738	.3038	
1v1 / 4v1	0.0490	0.0415	Inf	−3.562	.0021	**
2v1 / 3v1	0.7675	0.5581	Inf	−0.364	.9835	
2v1 / 4v1	0.1363	0.1096	Inf	−2.479	.0632	
3v1 / 4v1	0.1775	0.1413	Inf	−2.172	.1309	
perspective = pedestrian						
motorist = human						
1v1 / 2v1	1.2763	0.8892	Inf	0.350	.9853	
1v1 / 3v1	0.3438	0.2560	Inf	−1.434	.4780	
1v1 / 4v1	0.0836	0.0735	Inf	−2.823	.0246	*
2v1 / 3v1	0.2694	0.2009	Inf	−1.758	.2936	
2v1 / 4v1	0.0655	0.0580	Inf	−3.078	.0112	*
3v1 / 4v1	0.2430	0.2127	Inf	−1.616	.3694	

Note: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4.30: Preference for swerving for lives-at-risk × perspective × motorist-type interaction in car occupants versus pedestrians scenario.

lives-at-risk	P(swerve)	SE	df	lower asymptotic 95% CI	upper asymptotic 95% CI
perspective = car occupant					
motorist = self-driving car					
1v1	0.3681	0.2336	Inf	0.0752	0.8065
2v1	0.8280	0.1496	Inf	0.3804	0.9742
3v1	0.9025	0.0954	Inf	0.5249	0.9873
4v1	0.9519	0.0521	Inf	0.6798	0.9946
perspective = observer					
motorist = self-driving car					
1v1	0.4146	0.2438	Inf	0.0900	0.8353
2v1	0.8666	0.1220	Inf	0.4509	0.9809
3v1	0.9613	0.0424	Inf	0.7271	0.9957
4v1	0.8971	0.0986	Inf	0.5179	0.9861
perspective = pedestrian					
motorist = self-driving car					
1v1	0.1614	0.1338	Inf	0.0270	0.5720
2v1	0.7403	0.1900	Inf	0.2913	0.9518
3v1	0.9512	0.0504	Inf	0.6987	0.9939
4v1	0.9889	0.0137	Inf	0.8856	0.9990
perspective = car occupant					
motorist = human					
1v1	0.3694	0.2323	Inf	0.0766	0.8053
2v1	0.7704	0.1824	Inf	0.3077	0.9620
3v1	0.9755	0.0285	Inf	0.7939	0.9976
4v1	0.9453	0.0580	Inf	0.6569	0.9936
perspective = observer					
motorist = human					
1v1	0.1715	0.1428	Inf	0.0281	0.5973
2v1	0.3656	0.2305	Inf	0.0759	0.8016
3v1	0.4288	0.2436	Inf	0.0965	0.8406
4v1	0.8087	0.1601	Inf	0.3574	0.9698
perspective = pedestrian					
motorist = human					
1v1	0.6844	0.2191	Inf	0.2289	0.9406
2v1	0.6295	0.2357	Inf	0.1899	0.9249
3v1	0.8632	0.1238	Inf	0.4470	0.9801
4v1	0.9629	0.0408	Inf	0.7347	0.9959

4.5.2 Tables

Table 4.31: Outline of trials for Study 1.

Scenario	Trial	Groups at risk
children vs. adults	smaller groups larger groups	1 child (+ viewpoint avatar†) vs. 2 adults (+ viewpoint avatar†) 2 children (+ viewpoint avatar†) vs. 4 adults (+ viewpoint avatar†)
sidewalk vs. road	smaller groups larger groups	1 adult on sidewalk vs. 2 adults on road 2 adults on sidewalk vs. 4 adults on road
car occupants vs. pedestrians	parked van cliff	2 adult car occupants vs. 2 adults on road 2 adult car occupants vs. 2 adults on road

Note: † To avoid the artificiality of presenting the scenarios from the perspective of a child, additional adult avatars were added to both groups in the children vs. adults scenario, from which the pedestrian perspectives were presented.

Table 4.32: Predictors of judgements based on separate logit mixed models for each scenario (Study 1). p-values are calculated by parametric bootstrapping with 1000 samples.

	χ^2	df	p	
children versus adults scenario				
perspective	2.92	3	.5205	
motorist-type	3.57	1	.0991	
trial	1.22	1	.2475	
perspective × motorist-type	1.60	3	.7293	
gender	0.58	1	.4635	
age	0.38	1	.5972	
positive opinion of self-driving cars	11.33	4	.0639	
education	4.47	2	.1968	
driving experience	5.60	3	.2070	
visual acuity	6.05	2	.0859	
sidewalk versus road scenario				
perspective	6.94	3	.0986	
motorist-type	3.70	1	.0744	
trial	5.11	1	.0543	
perspective × motorist-type	5.50	3	.1698	
gender	5.15	1	.0253	*
age	0.65	1	.4200	
positive opinion of self-driving cars	7.51	4	.0866	
education	4.37	2	.1512	
driving experience	6.06	3	.1040	
visual acuity	3.81	2	.1170	
car occupants versus pedestrians scenario				
perspective	5.12	2	.1399	
motorist-type	3.45	1	.0909	
trial	68.89	1	.0010	**
perspective × motorist-type	3.43	2	.2452	
perspective × trial	8.58	2	.0170	*
motorist-type × trial	2.64	1	.1515	
perspective × motorist-type × trial	6.48	2	.0630	
gender	0.05	1	.8417	
age	0.62	1	.4754	
positive opinion of self-driving cars	5.40	4	.3083	
education	1.98	2	.4230	
driving experience	3.28	3	.4210	
visual acuity	5.68	2	.0960	

Note: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4.33: Predictors of self-reported confidence based on separate linear mixed models for each scenario (Study 1). p-values are calculated by Kenward-Roger test.

	num df	den df	F	p	
children versus adults scenario					
perspective	3	169	5.27	.0017	**
motorist-type	1	169	1.50	.2230	
decision	1	325	0.09	.7600	
trial	1	180	0.24	.6275	
perspective × motorist-type	3	170	0.55	.6509	
perspective × judgement	3	322	3.25	.0222	*
motorist-type × decision	1	329	1.55	.2139	
perspective × motorist-type × judgement	3	320	2.25	.0823	
gender	1	164	0.04	.8500	
age	1	159	1.68	.1970	
positive opinion of self-driving cars	4	161	0.52	.7180	
education	2	164	0.13	.8825	
driving experience	3	161	0.28	.8373	
visual acuity	2	163	0.63	.5337	
sidewalk versus road scenario					
perspective	3	191	2.30	.0791	
motorist-type	1	191	0.03	.8542	
judgement	1	338	4.57	.0332	*
trial	1	180	1.73	.1900	
perspective × motorist-type	3	190	1.92	.1279	
perspective × judgement	3	332	0.78	.5044	
motorist-type × judgement	1	338	2.47	.1170	
perspective × motorist-type × judgement	3	332	2.12	.0979	
gender	1	164	2.95	.0875	
age	1	160	0.02	.8910	
positive opinion of self-driving cars	4	161	1.10	.3607	
education	2	161	0.23	.7982	
driving experience	3	161	0.50	.6810	
visual acuity	2	160	2.86	.0603	
car occupants versus pedestrians scenario					
perspective	2	250	1.07	.3457	
motorist-type	1	284	0.20	.6534	
judgement	1	326	13.77	.0002	***
trial	1	248	7.93	.0052	**
perspective × motorist-type	2	232	0.19	.8263	
perspective × judgement	2	327	1.69	.1866	
motorist-type × judgement	1	322	0.68	.4118	
perspective × trial	2	242	2.49	.0852	
motorist-type × trial	1	258	0.00	.9652	
judgement × trial	1	298	10.81	.0011	**
perspective × motorist-type × judgement	2	321	0.16	.8508	
perspective × motorist-type × trial	2	236	0.18	.8339	
perspective × judgement × trial	2	287	0.49	.6112	
motorist-type × judgement × trial	1	301	0.07	.7974	
perspective × motorist-type × judgement × trial	1	303	0.17	.6827	
gender	1	164	0.54	.4627	
age	1	164	0.51	.4752	
positive opinion of self-driving cars	4	164	1.21	.3074	
education	2	161	4.06	.0191	*
driving experience	3	165	0.53	.6639	
visual acuity	2	166	0.17	.8457	

Note: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4.34: Predictors of judgements based on separate logit mixed models for each scenario (Study 2). p-values are calculated via likelihood ratio tests.

	df	χ^2	χ^2 df	p	
pedestrians versus single pedestrian scenario					
lives-at-risk	46	899.92	3	< .0001	***
perspective	46	2.99	3	.3928	
motorist-type	48	2.19	1	.1389	
road-type	48	9.87	1	.0017	**
lives-at-risk × perspective	40	70.19	9	< .0001	***
lives-at-risk × motorist-type	46	1.72	3	.6316	
perspective × motorist-type	46	0.96	3	.8108	
lives-at-risk × road-type	46	2.97	3	.3956	
motorist-type × road-type	48	0.98	1	.3214	
lives-at-risk × perspective × motorist-type	40	20.47	9	.0152	*
lives-at-risk × motorist-type × road-type	46	0.84	3	.8409	
first animation	48	0.01	1	.9305	
positive opinion of self-driving cars	45	12.92	4	.0117	*
knowledge of self-driving cars	48	1.29	1	.2566	
pedestrians versus car occupant scenario					
lives-at-risk	28	123.35	3	< .0001	***
perspective	29	1.95	2	.3767	
motorist-type	30	0.94	1	.3319	
lives-at-risk × perspective	25	7.13	6	.3086	
lives-at-risk × motorist-type	28	6.93	3	.0742	
perspective × motorist-type	29	2.36	2	.3079	
lives-at-risk × perspective × motorist-type	25	14.07	6	.0288	*
first animation	30	0.01	1	.9190	
positive opinion of self-driving cars	27	10.20	4	.0371	*
knowledge of self-driving cars	30	5.71	1	.0168	*

Note: * $p < .05$, ** $p < .01$, *** $p < .001$.

4.5.3 Figures

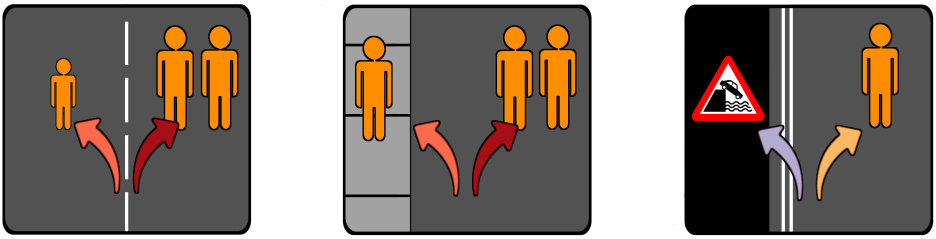

Figure 4.2: Pictorial representations of the three scenarios in Study 1. The relative numbers of orange figures in each scenario represent the ratios between the two groups at risk (assuming a single car occupant). The arrows indicate possible car actions and are colored corresponding to the graphs in Figure 4.3. **(A)** Children versus adults scenario; **(B)** Sidewalk versus road scenario; **(C)** Car occupants versus pedestrians scenario.

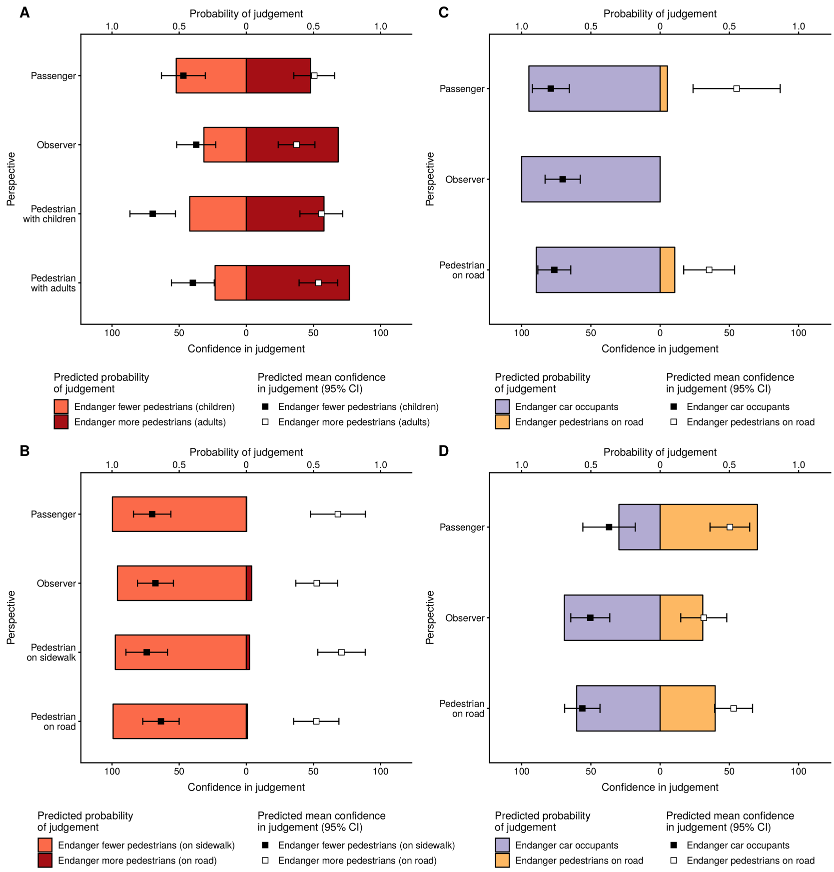

Figure 4.3: Model predictions for judgements and confidence (Study 1). Colored bars indicate the predicted probability of making particular judgements (indicated on the top x-axis) and are colored corresponding to the actions shown in Figure 1. Black and white squares with error bars indicate predicted mean self-reported confidence (95% CI) in the judgements made on a 0–100 scale (indicated on the bottom x-axis). As there were no significant effects of motorist-type, predictions are only separated by perspective. **(A)** Children versus adults scenario; **(B)** Sidewalk versus road scenario; **(C)** Car occupants versus pedestrians scenario (parked van trial) — note there were no observers who preferred endangering pedestrians, so the confidence in that case could not be estimated; **(D)** Car occupants versus pedestrians scenario (cliff trial).

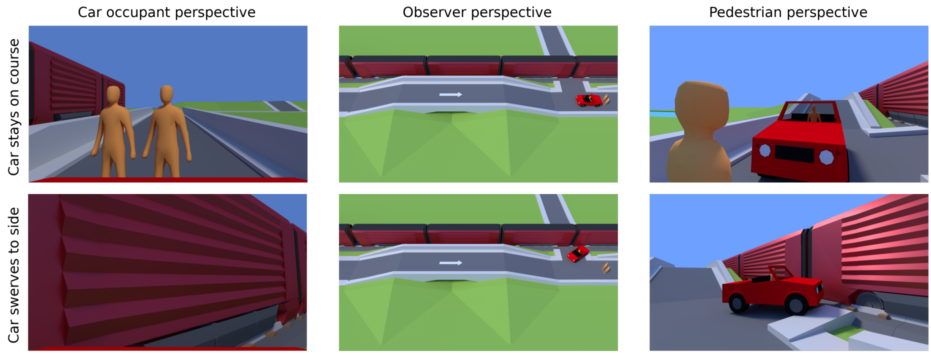

Figure 4.4: Final frames from animations for the pedestrians versus car occupant scenario (Study 2). The car either stays on course, endangering two pedestrians (top row), or swerves into a freight train, endangering the car occupant (bottom row). Different perspectives are shown: car occupant perspective (left column), observer perspective (middle column), pedestrian perspective (right column). Images depict 2v1 lives-at-risk (2 pedestrians vs. 1 car occupant). The animations used graphical models based on those by Jim van Hazendonk (https://racoon.media/) and Clint Bellanger (http://clintbellanger.net/).

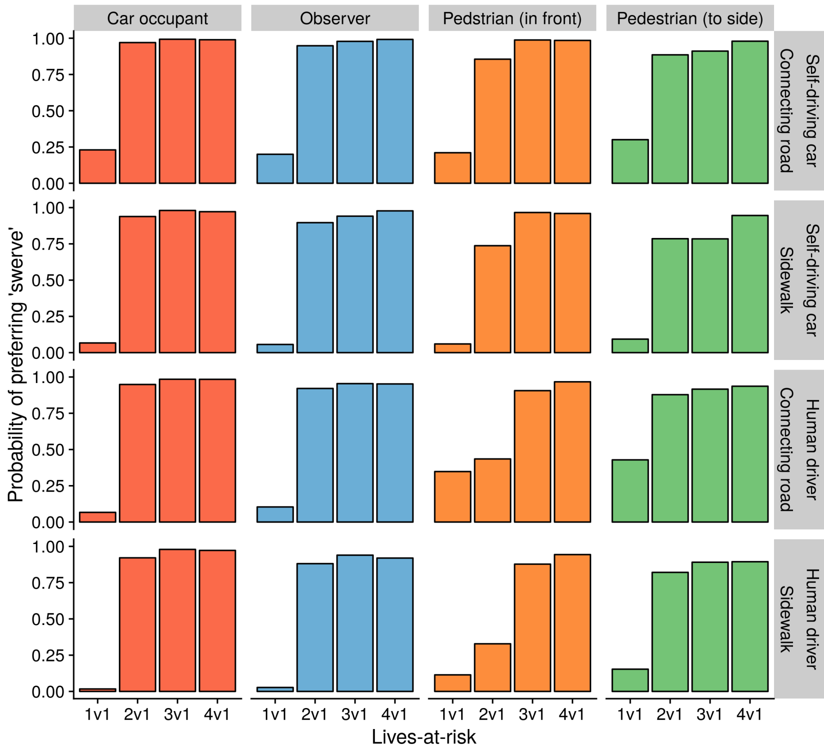

Figure 4.5: Model predictions for judgements on the pedestrians versus single pedestrian scenario (Study 2). Height of bars indicate the probability of choosing 'swerve' (endanger a single pedestrian to the side) as more acceptable. Different perspectives are separated in columns, combinations of motorist-type and road-type are separated in rows.

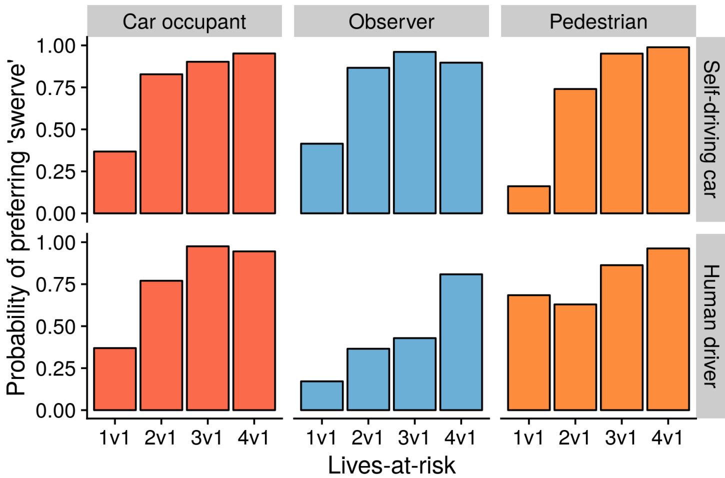

Figure 4.6: Model predictions for judgements on the pedestrians versus car occupant scenario (Study 2). Height of bars indicate the probability of choosing 'swerve' (endanger the car occupant) as more acceptable. Different perspectives are separated in columns, different motorist-types are separated in rows.

4.6 Talking cars, doubtful users - a population study in virtual reality

This section was submitted as a peer reviewed article in IEEE Transactions on Human-Machine Systems together with Shadi Derakhshan Maximilian Alexander Wächter, Ashima Keshava, Artur Czeszumski, Hristofor Lukanov, Marc Vidal de Palol, Gordon Pipa and Peter König. See Publication List for details.

4.6.1 abstract

ADVs are a significant development in our society, and their acceptance will largely depend on trust. This study investigates strategies to increase trust and acceptance by making the cars' decisions transparent. We created a virtual reality experiment with a self-explaining autonomous car, providing participants with verbal cues about crucial traffic decisions. First, we investigated attitudes toward self-driving cars in 7850 participants by a simplified version of the Technology Acceptance Model questionnaire. Results revealed that female participants show less acceptance than male participants, and there is a general decrease in acceptance with increasing age. A self-explaining car impacts trust and perceived usefulness positively. Surprisingly, it negatively influences the intention to use and perceived ease of use. This implies that trust is dissociated from the other items of the questionnaire. Secondly, we analyzed behavioral data of 26750 participants to investigate the effect of self-explaining systems on head movements during the virtual reality drive. We observed significant differences in head movements during the critical events and the baseline periods of the drive between the three driving conditions. Further, we demonstrated positive correlations between head movement parameters and the TAM scores, where trust showed lowest correlation. This is further evidence for the dissociation of trust from the other TAM factors. These results demonstrate the benefits of combining subjective data obtained by questionnaires with objective behavioral data. Overall, the outcome indicated a partial dissociation of self-stated trust from the intention to use and objective behavioral data.

4.6.2 Introduction

Autonomous driving vehicles (ADVs) are the primary goal of most car manufacturers Hars, 2016. The development seems to be cumulative since more and more functionalities are automated in new cars Dajsuren and van den Brand, 2019. One primary reason why ADVs are of value is the possibility of eliminating human driving error, which is responsible for 93% of road accidents Allahyari et al., 2008; T. Johnson, 2013. Further, ADVs are safer as they are faster and more precise in the dynamic driving task as well as in object and event detection Carranza-García et al., 2021; Papadoulis et al., 2019; SAE Internation, 2014; Schoettle, 2017. As technical developments in the field are fast and continuously improving, there is little doubt that self-driving cars will have a significant impacts on our society Chehri and Mouftah, 2019. These can ranges from drastically decreasing greenhouse gas emissions to reducing the number of traffic-related injuries. Which consequently might lead to possible reshaping the infrastructure of our current cities Benleulmi and Blecker, 2017; Chehri and Mouftah, 2019; Othman, 2021; Ryan, 2020. Thus, introducing ADVs into our daily lives appears as a highly desirable goal.

Trust and acceptance of potential customers define the extent to which ADVs are used for individual transportation Howard and Dai, 2014b; Krueger et al., 2016. According to current research, there is a limited willingness among potential customers to use ADVs C. Lee et al., 2019; C. Lee et al., 2017; M. et al., 2015; Ryan, 2020; Ward et al., 2017. Various surveys have shown that most potential buyers are unwilling to use an ADV at all or to use it to its full extent Othman, 2021; Rezaei and Caulfield, 2020b. Primary reasons for the lack of trust and acceptance are the fear of system malfunctions and the hesitation of giving complete control to the car C. Lee et al., 2019; Othman, 2021; Szikora and Madarász, 2017. The further reluctance of potential customers may stem from low technology self-efficacy Czaja et al., 2001, meaning that people do not feel confident enough to operate an ADV proficiently M. König and Neumayr, 2017. Since trust and acceptance are shaped by knowledge and experience, the cause of this reluctance might be rooted in the lack of transparency in and knowledge of the ADVs' decision-making. It might not be clear on what basis an ADV is deciding on. Not knowing what the artificial agent perceives or reasons, directly influences the concerns of safety Forster et al., 2017; Koo et al., 2016. Therefore it is crucial to find measures that are able to increase the trust and acceptance of ADVs.

According to the Technology Acceptance Model (TAM), perceived usefulness and ease of use are cognitive responses to new technology and predict the intention of using it Davis and Venkatesh, 1996. Consequently, a low intention to use makes the future application of ADVs questionable Bergmann et al., 2018a; Howard and Dai, 2014b; Rezaei and Caulfield, 2020a. Belanche and colleagues developed a research model Belanche et al., 2012 expanding the TAM by adding trust as a component. They found a causal relationship between trust and all three elements of the original TAM Belanche et al., 2012. Therefore, trust can be seen as a critical factor for the acceptance of a new technology C. Lee and Coughlin, 2015; Lüders and Brandtzæg, 2017; Wintersberger and Riener, 2016. Earlier studies investigated trust modulating factors, such as human-machine communication style, feedback, and anthropomorphic features in automation Hoff and Bashir, 2015; Seppelt and Lee, 2019; Wintersberger et al., 2019; Wintersberger et al., 2020. Hoff and Bashir Hoff and Bashir, 2015 suggested that trust in an ADV is an accumulation of personal tendencies, environment, and user's perception of the autonomous system. Lee and See J. D. Lee and See, 2004 argue that the perceived homogeneity of communication style, feedback, and anthropomorphic features shape trust levels. The shared statement in all these findings is that the user should perceive the system as reliable and trustworthy. Moreover, previous research showed that excess information about the driving state of a car is perceived as distracting or unpleasant. Howard and Dai, 2014b; Koo et al., 2015; Krueger et al., 2016; Othman, 2021; Rezaei and Caulfield, 2020b; Ryan, 2020. The desired amount of information by the ADV may be the key to understanding trust and, consequently, acceptance of self-driving cars Du et al., 2019.

Research on trust as a high-level cognitive phenomenon relies heavily on self-reported data. A review of Raats and colleagues of 258 experiments on trust in ADVs revealed that 84% used questionnaires as the assessment method. Only 4.7% of the studies used observations as a data-gathering tool Raats et al., 2020. An objective form of data is needed since the participant's self-assessment is often biased by self-perception or socially desired behavior Choi and Pak, 2005. For this, we propose head movement data, since humans represent their cognitive states implicitly based on body language, facial expression, gaze direction, and movement of the head Newen et al., 2018; Zhao et al., 2013. Even though previous research has established the link between gaze shift and cognition Yarbus, 1967, several studies showed that head rotation corresponds to the visual gaze Fang et al., 2015; Yarbus, 1967 and both coordinate cognitive processes M. F. Land, 2004; Proudlock et al., 2003. This coordination exists in orientation, which means that the head and eyes move in the same direction. Thereupon, the head orientation provides information about the center of attention Fang et al., 2015. Behavioral data such as head movements are unconscious, fine-grained, and provide continuous information that can be used to access cognitive processes like trust Grafsgaard et al., 2012; Lu and Sarter, 2019.

To gain insight into modulating factors of acceptance towards ADVs and their representation in users' head movements, we used a previously developed large-scale virtual reality (VR) experiment called WestdriveNezami et al., 2020. We expected to find significant differences in the participants' attitudes and significant differences between different age groups and genders. Additionally, we predicted differences in head movements between different conditions. We expected to find significant variance in head movement patterns and head angular velocity as an effect of transparent communication of an ADV. We assumed that the self-reported acceptance in conjunction with head movements enables more objective insights into modulating factor of acceptance.

4.6.3 Materials and Methods

We gathered data from visitors in the German Ministry of Education over six months, and in a traveling exhibition (MS-Wissenschaft) over the course of a full summer. Participants experienced a 90-second drive in a virtual environment called Westdrive, covering 2,5 km^2 with more than 100 cars and 150 pedestrians. Participants experienced a single trial in one of three driving conditions. The first condition was a fully autonomous car with an anthropomorphic voice assistant system (AVAS) giving information about critical traffic events and the corresponding decisions of the car. The second condition was an ADV with a radio broadcast playing through the whole trial. In the third condition, a female Taxi-Driver drove the participant through the city. Here, the TaxiDriver responded verbally to the surrounding traffic. We gathered objective and subjective data in the form of head orientation and head angular velocity, as well as by an adaptation of the TAM questionnaire Davis and Venkatesh, 1996.

During each trial, participants were confronted with three critical traffic events without the possibility to intervene (Figure 4.7). The duration of the events was the time between entering and exiting the event objects to the participants' view. In the first event, a jogger crossed the road directly in front of the car. In the second event, a high-speed car took the right of way at an intersection and the third event included a slow pedestrian crossing the street. The onset and end of the events were the same for all the participants in all conditions. At none of the events, the participants' cars hit any event objects. In the AVAS condition, the ADV gave short information about the critical event situation. This happened at the spawning of the critical traffic objects to warn at the earliest possible point. The design of these events was based on previous research that showed feedback should include the reason why an ADV decides in a specific way Koo et al., 2015. Also, additional information should be provided while interacting with vulnerable road users when their intentions are not clear but can influence the car's behavior Wintersberger et al., 2020. These events were implemented to test the participants reactions as a passive passenger in critical situations. They were designed to test whether the in-vehicle communication can alter behavioral reactions and acceptance.

Before starting the trial and data recording, participants were asked to adjust the HMD by themselves. This adjustment phase was not limited in time and was not counted as part of the experiment. Afterwards, participants were informed about the study procedure and the goal of this study. Because of possible symptoms of cybersickness, the participants were informed in the introduction that they could remove the HMD at any given time. If so, this data has been excluded from the analysis.

The simplified questionnaire consists of three questions from the original TAM in perceived usefulness, ease of use, the intention of use, and one additional question on perceived trust. It also included questions on age, gender, aviophobia, driving experience, amount of gaming hours per week and the number of exposures to virtual reality before the experiment. The questionnaire has been answered in the Likert scale, with numbers from 0 (strongly disagree / dislike) to 100 (strongly agree / like) indicated by thumb icons of like and dislike.

The used experimental setup consists of two HTC Vive pro HMDs and lighthouses version 1 for tracking head position and rotation while seated in the car. The VR computers were equipped with Nvidia Geforce RTX 2080Ti GPUs, 16Gb of RAM, and Intel Xeon W-2133 CPU @ 3.60Ghz core, resulting in an average frame rate of 25,2 fps. Additionally, the setup used two raspberry pies and touch monitors for web-based questionnaires. For analyses, Python 3.6, pandas 0.24.2, NumPy 1.16.4, Scipy 1.7.2, statsmodels 0.10.0, as well as SPSS 29 were used. All plots were created using Matplotlib 3.1.0 combined with seaborn 0.9.0. Data-driven prepossessing on questionnaire data was performed with the OPTBIN algorithm

Figure 4.7: Three scripted critical events occurred during the ride from top to bottom: Pedestrians running on the street from left to right, fast cars cutting in the self-driving car path, and pedestrians walking in the middle of the road.

Knuth, 2013 using histogram-based age binning.

Analysis of the data

Head movement data were obtained from 26750 participants and the questionnaire was answered by a fraction of them. Elimination of incomplete answers resulted in 7850 data sets.

First, we focused on the analysis of the questionnaire data. Of the complete data set, 4464 participants identified as male, 3386 as female. By using optimal binning Knuth, 2013, we divided participants into five age groups. The cleaned data set consisted of 2812, 1513, 1883, 582, and 86 in the age groups <20 years, 21-40 years, 41-60 years, 61y-80 years, and above 80 years, respectively. In the AVAS, TaxiDriver and RadioTalk condition we recorded 2691, 2636, and 2509 data sets, respectively. The large number of participants in each bin allowed the use of regression-like inferential tests (i. e. MANOVA) due to their robustness against non-normalities in large data-sets Pek et al., 2018.

To investigate the effect of gender, age, and driving condition on the four aspects of the questionnaire, a one-way multivariate analysis of variance (MANOVA) has been performed. MANOVA tests the optimal linear combination of dependent variables to find significant effects. We performed a one-way MANOVA for the four TAM aspects modeled with respect to gender, age, and driving condition. Pillai's Trace test statistic uses estimated F-Values to test significance, which is robust against non-normalities. Therefore, Pillai's Trace adds an extra layer of protection against false positives Finch and French, 2013 and is a good choice for interpreting the results. To understand how different categories within each factor, e.g., male or female in gender, affect the four TAM aspects, we calculated a separate one-way analysis of the variance (ANOVA). Following, we calculated the different effect sizes (Cohen's D and Hedge's G) for each of the factors using estimated means and standard deviations reported for the category within that factor. Although both of these effect sizes are based on Cohen's suggestions, Hedge's G considers the sample sizes of the compared groups. Therefore, both effect sizes have been used to interpret the results. Further, each participant's four TAM aspects were combined into one single value. Together with the MANOVA, we were able to make statements about how gender, age, and the condition affect the questionnaire scores.

However, ANOVA can only be calculated on a single independent variable. The best way to combine the four TAM aspects into one value is by multiplying each aspect's score for a given participant by a corresponding weight and adding them all together to get a single value. This acceptance score was calculated by performing a linear discriminant function for each factor that will yielded in a different raw coefficient for each TAM aspect concerning the given factor. The linear discriminant analysis (LDA) intends to find a linear combination of features that characterizes or separates two or more classes. It expresses the dependent variable as a linear combination of the independent variables that maximizes the group differences within the dependent variable McLachlan, 1992. The raw discriminant function coefficients can be used as weights to calculate the four TAM aspects into one independent number, which we can call acceptance score.

Next, we turn to the analysis of the objective behavior. A head-mounted HMD measured the orientation and position of the participant's head in the virtual environment. We determine the head orientation in a reference frame fixed to the car. Since most interesting visual detail was placed near the ground level and all the dynamic objects of the virtual city moved along the horizontal axis, we focused on the orientation along the horizontal plane. Further, we compared the head angular velocity, meaning the change in head orientation degree over time. To examine the differences between conditions, we used one-way ANOVA followed by the Tukey Honest Significant Difference (HSD) post-hoc test. The Tukey HSD compares pairs of means to detect which of the group means are different from the others (Meandiff). With this test, we could define the separate condition that causes differences in orientation and angular velocity in specific point of time Abdi and Williams, 2010. Additionally, we calculated the Pearson Correlation between the head angular velocity and the TAM scores for each questionnaire item to check for consistencies in both subjective and objective measures.

4.6.4 Results

Questionnaire Results

The questionnaire data of the simplified TAM from 7850 participants showed a positive correlation of $r>0.4$ between the questionnaire items. Therefore, these items have to be analyzed together as multivariate dependent variables. To check validity of the assumptions, a Levene's test was performed. If the test was significant we would assume a violation of variance homogeneity in the groups. Levenes's test resulted in F-values of 1.369 for perceived Usefulness ($p = 0.089$), 2.333 for Ease of use ($p < 0.001$), 1.459 for Intention of use ($p = 0.053$) and 1.443 for Trust ($p = 0.058$). Considering the large sample size, known to reduce p-values in Levene's test, a further check of the covariance matrices for the dependent variables of the TAM concerning the main factors of gender, age group, and condition has been done. We found homogeneity of covariances, as assessed by Box's test ($p > .001$). Together, Levene's test and the covariance matrices provide essential evidence for the validity of the assumptions of MANOVA. Out of the four different null hypothesis tests of the multivariate analysis, Pillai's Trace was chosen due to its known robustness toward non-normalities in the data Ateş et al., 2019. Therefore, the multivariate analysis of variance is the prime analysis method Warne, 2014; Warne et al., 2012.

To gain deeper insights into how gender, age, and condition affect the TAM factors, a linear discriminant analysis (LDA) was used to extract each independent variable's weighted influence. Linear discriminant analysis tries to find a set of coefficients that will maximize the separability within the given independent variable. These coefficients were used to interpret the influence of each independent variable on each of the modulator factors of the TAM.

The effect of gender

First analysis checked for differences between male and female participants regarding the acceptance scores. In order to find out the influence of gender on acceptance, we performed an MANOVA with a follow-up LDA for gender. The Pillai's Trace resulted in 0.00293 ($F(4,7835) = 4.761$, $p < 0.001$) showing that there is a significant effect of gender on overall acceptance. The follow-up LDA showed that females have a lower score based on the observed discriminant coefficients. The resulting coefficients were -0.33 for the intention of use, -0.06 for perceived usefulness, -0.60 for perceived ease of use, and -0.18 for trust (Figure 4.8 a), all with a medium effect size (Cohen's $D = 0.45$). Additionally, the LDA showed that perceived usefulness and trust were less affected by gender than the intention of use and the perceived ease of use (Figure 4.8 a). These findings indicate that females and males have an almost equivalent attitude towards the perceived usefulness but differ in the perception of ease of use and, consequently, the intention of using self-driving cars. Thus, we interpreted that females anticipate difficulties in handling and therefore score lower in the intention to use.

The effect of age group

As a next step, we investigated what influence age had on the answers in the questionnaire. Similar to the effects of gender, the result of the MANOVA was paired with a follow-up LDA for the age group to find the influence of the TAM items. The resulting Pillai's Trace of 0.04561 ($F(16,313) = 24.107$, $p < 0.001$) indicated a significant effect of age group on the overall acceptance. LDA resulted in discriminant coefficients of -0.27 for the intention of use, -0.19 for perceived usefulness, -0.46 for perceived ease of use, and -0.30 for trust. These results showed that age has a negative effect on all TAM items. The age group under 20 years showed high scores in all questions, with medium effect sizes compared to the age group between 20-40 (Cohen's D = 0.50) and 61-80 (Cohen's D = 0.43) (for the full list of the effect sizes, see Appendix I Table 4.35). Similar to the analysis of gender, perceived ease of use was affected by age the most (Figure 4.8 b). Together with the decreased intention of use, it can be stated that older adults anticipate hardships in using this technology, therefore showing lower scores in the intention to use. However, the overall acceptance scores increased again beyond 80 years of age, especially in the perceived usefulness (Figure 4.8 b). This is also reflected in smaller effect sizes between the age group below 20 and above 80 years (Cohen's D = 0.18) (Table 4.35). Concluding, data showed the highest acceptance in the age group below 20 years, with a general decrease of acceptance until 80 years.

The effect of condition

A central hypothesis of the study was, that compared to a traditional taxi the acceptance level of ADVs is reduced, but can be partially recovered by making the decisions of the ADV transparent. The result of MANOVA showed that the condition had a significant effect with Pillai's Trace of 0.00259 ($F(8,15672) = 2.541$, $p = 0.009$). The LDA for condition resulted in coefficients of -1.12 for the intention of use, 0.99 for perceived usefulness, -0.33 perceived ease of use, and 0.44 for trust, with overall small effect sizes in all comparisons (Cohen's D = <0.11) (Appendix I Table 4.36). While the effect on trust and ease of use is negligible between conditions, differences could be found in intention to use and perceived usefulness. The AVAS condition had a slightly higher median score in the perceived usefulness of 71 compared to 69 in the TaxiDriver (Figure 4.8 c). Here, the AVAS condition resulted in a lower median score of 65 than the TaxiDriver with a median of 68 (Figure 4.8 a). We concluded that there were no adverse effects of the condition on the ease of use like in gender and age. Still, there was a negative effect on the intention to use such technology independently of gender and age. These results already accommodated that additional factors besides age and gender negatively influenced the intention of use. This observed effect could also be due to technology aversion, which had already been mentioned in the effects of age and gender. It can be summarized that there was a small positive effect of in-car communication methods on accepting ADVs regarding the ease of use and a small negative effect regarding the intention of use.

The Interaction effect of gender and age group

While investigating the effects of gender, age and condition, it became clear that these factors separately did not explain all variance observed in the data. There was a significant interaction effect of gender and age group with Pillai's Trace of 0.00498 ($F(16,31352) = 2.441$, $p = 0.001$). According to the follow-up LDA, there was a negative effect for the intention of use and perceived ease of use (both -0.73) and a positive effect on the perceived usefulness (0.22) and trust (0.55) in the questionnaire items. Here, the effect sizes were most notably between the age groups 21-60 years compared to under 20 years and above 60 for each gender (Appendix I Table 4.37). These results support findings of the previous analyses on gender and age. In addition, it could be shown that the interaction of gender and age had a significant influence on the acceptance of ADVs. In addition, the largest effect sizes ($0.5 <$ Cohens' $D <= 0.9$) resulted by comparing female participants in the age between 21-60 years against the male participants in the same range. Participants below 20 years had the highest TAM scores, and females between the ages of 21-60 years showed the lowest TAM scores. Although there is was a decrease in all TAM factors in both genders for ages between 21-60 years old compared to the below 20 years, female participants showed stronger decreases in TAM scores (Figure 4.9). This accounts especially for the intention of use and perceived ease of use. Once again, as age increases for people between ages 21-80 years, we can also observe TAM scores. In conclusion, gender and age group interaction significantly affect all TAM factors, specifically negative influences on the intention of use and perceived ease of use for ADVs, but a positive effect on perceived usefulness and trust. This means that although the ADV was seen as

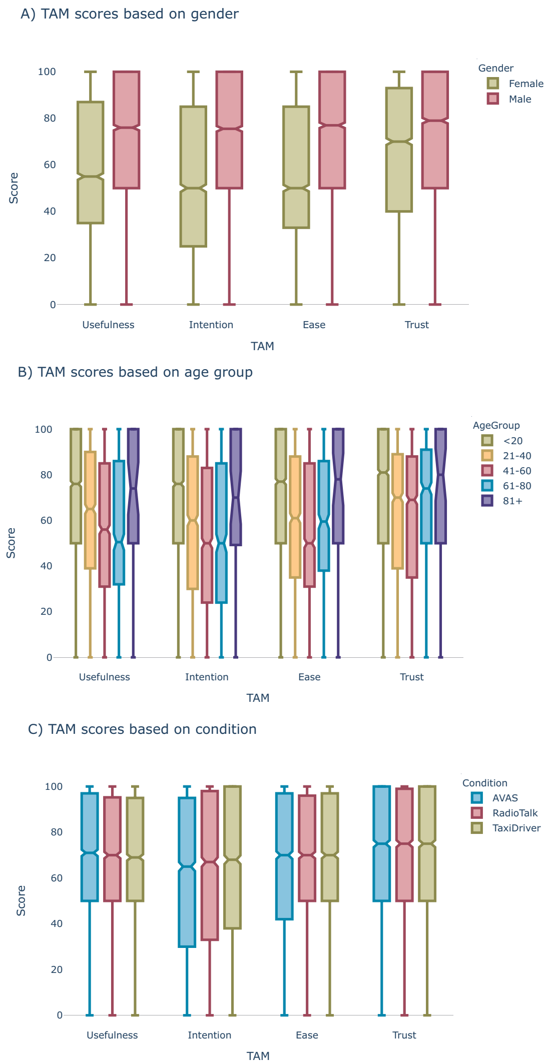

Figure 4.8: The descriptive categorical plot of the mean questionnaire answers for each of the main factors a) gender, b) age group, and c) condition.

useful and trustworthy, there were still other hidden factors that decreased the ease of use and the intention to use it. Consequently, the demographic factors of age, gender and the interaction of these two, have much more impact on the items of the TAM questionnaire. The positive effects of a self-explanatory ADV were not sufficient to compensate for the negative influence of demographics on ease of use and, accordingly, intention to use.

4.6.5 Behavioral Results

Identification of critical events

As a first step of the head movement analyses, we investigated whether participants' behavior differs during the critical events from the baseline parts of the drive. The initial analysis was performed regardless of the driving condition. We considered the mean orientation and variance of head orientation over all participants as the relevant dependent variables. Collapsing the data over conditions, we tested whether the mean of head orientation in each frame was significantly different from the distribution resulting from a permutation over time (permutation test). Head orientation differed significantly from baseline at the end on the first and second and at the very end of the third event. (Figure 4.10). Further, we observed differences early in the trial, when participants where intensively looking around inside the car. Additionally, three other significant intervals were observed. During these times pedestrians were visible on the sidewalk in crowded places of the city. We assume that this is related to a need of information to assess the situation. A final period of deviant head orientation is observed at the very end of the drive, when participants prepared to exit the car in a crowded area. By applying this method we are confident that an additional measurement of head movements is a valid approach to enhance subjective data. Overall, compared to the baseline head orientation, the three critical events showed significant differences in participant behavior regardless of the effect of conditions. These differences were not limited to the critical event intervals but identified in additional areas of the trial.

The effect of condition

As a next step, we consider how much of the observed variance was related to the effect of condition. We investigated whether participants' behavior objectively

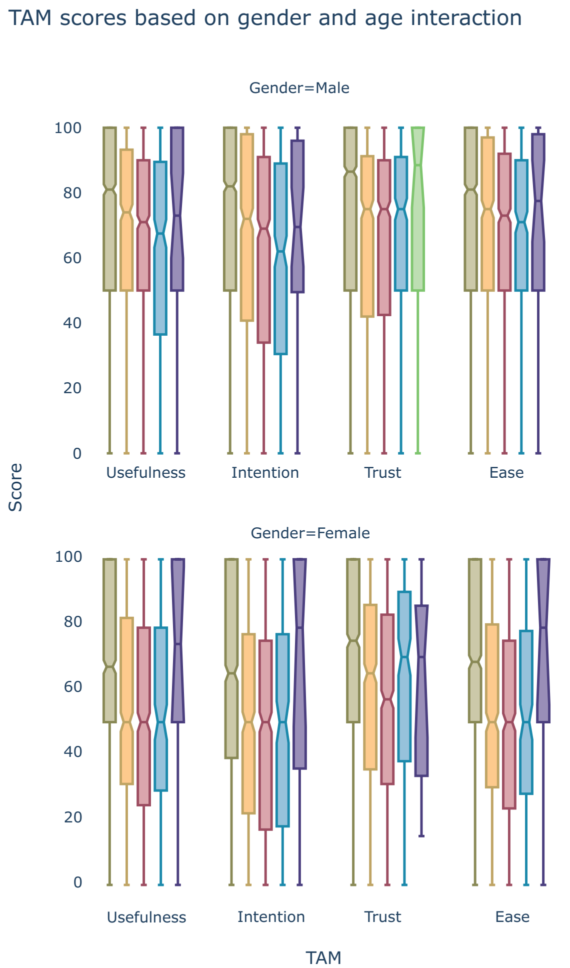

Figure 4.9: Mean of answers for questions regarding the usefulness, intention, trust, and ease for age group and gender combined

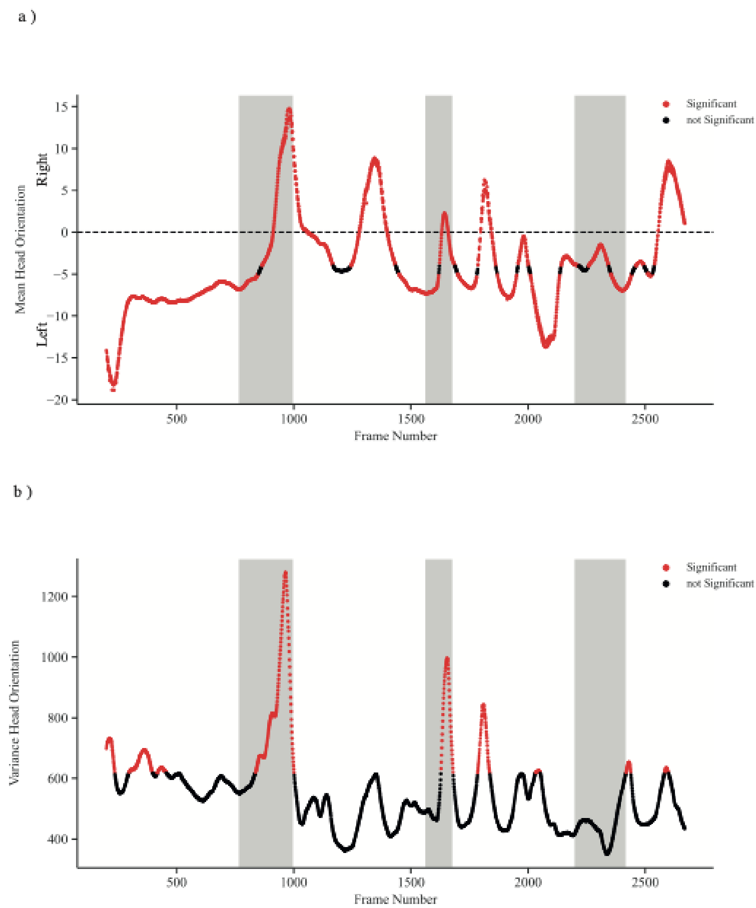

Figure 4.10: Time intervals of significant differences in head orientation after the permutation test (n=1000). Shaded areas represent the critical traffic event intervals. a) Mean of head orientation over all subjects. Each point indicates an average of the head orientation across the participant within each frame. b)Variance of head orientation over all subject. The red areas indicate the intervals where there was a difference in head orientation between all participants. Please note that the data is collapsed over condition.

differs between conditions. Differences in head orientation were seen as indicators of participants' reaction to the environment in different conditions. By visualizing the head movement data, we observed differences in the mean head orientation, over large parts of the drive and during the critical traffic events. (Figure 4.11). To see, whether these differences were significant, we calculated a one-way ANOVA based on head orientation for each frame as the dependent variable and applied a post-hoc comparison of Tukey HSD in the significant Intervals. The result of the ANOVA showed significant differences in head orientation between the three conditions during most of the drive time ($F>10$, $p < 0.05$) (Figure 4.12). Specifically, the TaxiDriver condition was significantly different from the two others over a large part of the drive. The post-hoc comparison revealed larger mean difference (Meandiff) for TaxiDriver compared to AVAS (i.e. Meandiff = 2.07, $p = 0.001$ for frame = 1300) and RadioTalk (i.e. Meandiff = 1.77, $p = 0.001$ for frame = 1300) condition. This result was mostly constant during the experimental trial, including the three critical events. At the start of the first and the second event, no significant differences between the RadioTalk and AVAS condition were found. In the third event we found differences between all three conditions. Here, participants elicited the smallest degrees of the head orientation in the AVAS condition, and the highest degree in the TaxiDriver condition. We mainly observed a higher mean head orientation in the TaxiDriver condition in the beginning of the critical traffic events. This means, the distribution of the head orientations in angular space were wider in the TaxiDriver condition compared to the two autonomous conditions in most of the trial times (Figure 4.11).

Head angular velocity

To gain deeper insights into the participants' head movement behavior, we calculated the magnitude of change in head orientation throughout time as the angular velocity. We quantify the absolute value of the angular velocity in frame n for each critical traffic event separately ($\omega_n = |\theta_n - \theta_{n-1}|/\Delta t$) where θ_n is the head orientation in frame n and $\Delta t = 0.04s$ based on the experiment's overall average frame rate. The analysis of the angular velocity showed that in the AVAS condition, participants rotated their heads significantly faster only in the first critical traffic event ($F(2,24447) = 71.35$, $p < 0.01$). In the second critical traffic event, no significant differences between conditions were found ($F(2,24447) = 2.8$, $p= 0.06$). In the third critical traffic event, the angular velocity in AVAS was significantly lower than the two other conditions ($F(2,24447) = 29.06$, $p < 0.01$) (Figure 4.13).

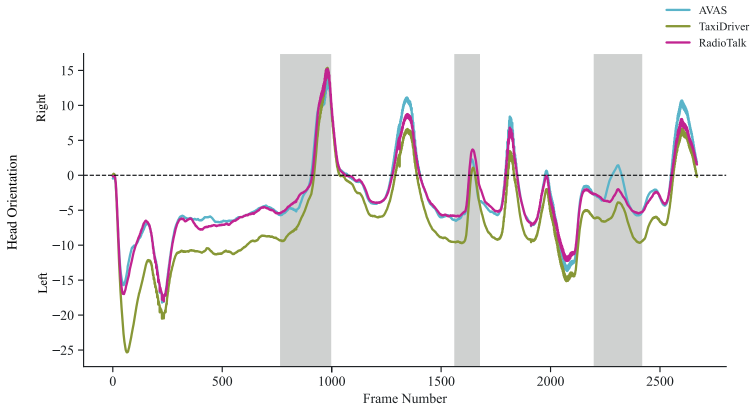

Figure 4.11: Mean head orientation in each frame divided in three conditions. The positive and negative values of the mean orientation relate respectively to the right and left directions. Shaded areas represent the critical traffic event intervals.

Overall, the data revealed that the angular velocity of head movements decreased during the experimental trial in all three conditions. However, the AVAS condition reduced the head's angular velocity to a larger degree than the other autonomous condition. With the analysis of the angular velocity, we were able to show that participants' behavior changed as an effect of self-explaining ADV over time.

4.6.6 Questionnaire comparison

The head angular velocity was an illustration of the participant's head movements behavior during the trial. Calculating the relationship between the angular velocity and the TAM items allowed us to determine if the participants self-assessed acceptance had been expressed in their previous behavior during the experimental trial. We used Pearsons' correlation for the participant's absolute head angular velocity over the entire trial and participants' respective TAM item scores. The analysis showed positive correlation between the head angular velocity and all TAM item scores in all three conditions (Figure 4.14). Comparing the TAM items, there was a lower correlation between the angular velocity and trust compared to its correlation with other TAM items. Along with the previous finding in the analysis of the questionnaire, the mismatch between trust and the other questionnaire items has been demonstrated in the correlation of the items and the angular velocity.

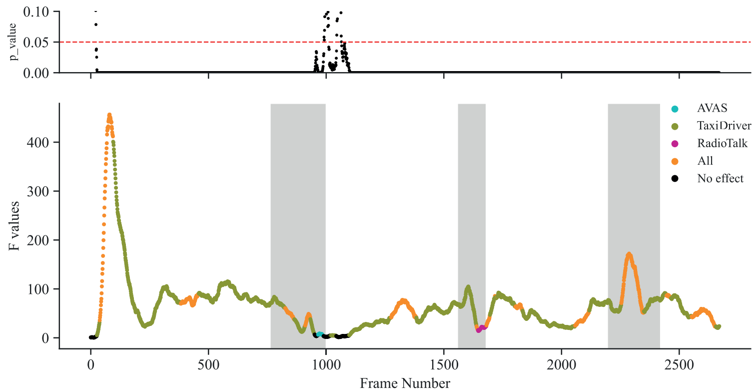

Figure 4.12: Time intervals of significantly different behavior between the three conditions. The graph depicts the P and F values of one-way ANOVA overall the experimental trial. Each dot shows the original F value of each frame. The red dash indicates the significance threshold (p < 0.05). Shaded areas represent the critical traffic event intervals. The result of Tukey's post hoc comparison is represented by different colors. Each color shows the significant variable mean(s) in cross-check.

The dissociation between trust and the other questionnaire items lead us to the assumption that trust is not an ideal item in self assessments via a questionnaire. This claim is supported by the mismatch in the self-assessment, as well object behavioral data. Therefore, we argue that the objective behavioral data was able to reflect the findings in the TAM questionnaire.

4.6.7 Discussion

The present study revealed that self-reported acceptance in conjunction with objective observation enhances the understanding of modulating acceptance factors. According to the results, subjective data from a post-experimental questionnaire and objective data from head movements during the experimental trial were largely congruent. Outcomes of investigating gender, age, and condition effect on the overall acceptance showed a lower acceptance of female participants toward ADV than males. However, this effect is even more pronounced in the intention to use ADVs. The results also indicated that people below 20 years of age have the highest acceptance toward ADV, gradually decreasing with age while increasing again above 80 years. Regarding the effect of a self-explaining ADV, we found a

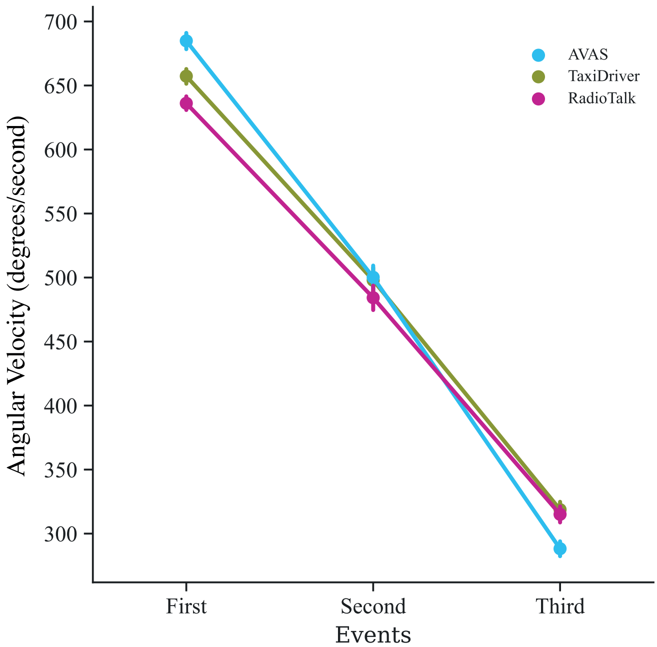

Figure 4.13: Head angular velocity for the critical event intervals. The y axis refers the rotation change divided by the number of event frames.

small positive effect in the ease of use and a small negative effect regarding the intention of use. However, age, gender, and the interaction of these two have a substantially higher impact on the questionnaire scores. Therefore, the positive effects of a self-explanatory ADV are not sufficient to compensate for the negative influence of demographics on ease of use and the intention to use. We could show that participants' head orientation differed between the conditions by analyzing the head movement data. Especially in the TaxiDriver condition, we did see significant differences over the whole drive with further differentiation between conditions during the critical events. Further, we observed a decrease in the head's angular velocity over time for all conditions. This effect was most substantial in the AVAS condition. Finally, correlating the magnitude of the participant's head angular velocity with the TAM scores showed a significant relationship between acceptance as a combination of the TAM factors, which was weaker for the factor trust.

Previous studies mainly depended on answers gathered from the potential users Howard and Dai, 2014b; Raats et al., 2020; Wintersberger et al., 2020. However, behavioral data is not as susceptible as questionnaire answers and can be used to validate possible self-assessments Davis and Venkatesh, 1996. The presented study could show a dissociation of the self-assessed trust and the other TAM items, especially with the intention to use. This observation contrasts previous research such as the work of Belanche Belanche et al., 2012. In fact in the present study it is shown that self-assessments are heavily modulated by the demographic factors such as age and gender, as well as the interaction of these two factors. Behavioral data confirmed the dissociation of trust and intention, by showing a connection between head movements and scores in intention, ease of use and perceived usefulness. Therefore we are arguing that including behavioral data is a valid approach to better understand underlying factors of acceptance and correcting potentially faulty subjective data. This is due to the fact that head movements can be considered as part of nonverbal communication in humans Mehrabian, 2017. It contains information of the participant's emotions and intentions Gunes and Pantic, 2010. For instance, head angular velocity and acceleration were higher during negative affects Hammal et al., 2015. Combining sources of subjective and objective data, make it is possible to validate the questionnaire data. In conclusion, it could be stated that the behavioral data is an important resource that can be used to validate investigations of the technology acceptance model and its underlying factors.

Due to the nature of the experiment within a public exhibition and a significant number of visitors, the technology acceptance questionnaire used in this study was a simplified version. Thus it might not grasp the entire aspect and spectrum of factors that modulate acceptance, such as technology self-efficacy, which might play an essential role in perceived ease of use. Furthermore, the questionnaire was also translated into German, and we could not validate the questionnaire before using it in the experiment. Although part of the variance in the data might be due to the translation, such effect is thought to be minuscule and negligible since our main findings align with that of the previous works Chen and Chan, 2011; Koo et al., 2016; Othman, 2021; Venkatesh, 2000. Due to the simplified nature of the study, we can not directly address and analyze the underlying information processing that influences attitude. Still, we are confident to make informed statements due to the large effects in a vast data set. Additionally, there is a possibility that cybersickness influenced the TAM scores and head movement data. Nevertheless, we tried to control as much as possible for motion or cybersickness in this trial.

The first step was to include a bigger static frame for the participants as the car interior, reducing the probability of cybersickness during the trial. Also, we only used a low-speed environment, without sharp turns to reduce cybersickness as much as possible van Emmerik et al., 2011. Further, We acknowledge that a more precise measurement instrument such as eye trackers would have enhanced the analysis and the findings. However, once more, the nature of the experiment and the absence of experimenters on-site (not counting the numerous visits for maintenance) rendered the use of such methods impossible. Another criticism could be that the experimental time was limited to 90 seconds, and each participant observed only one of the experimental conditions. However, this experience can already investigate participants' acceptance toward various in-car communications in ADVs. Additionally, the vast amount of data gathered by the experiment allowed for entirely data-driven analyses both for questionnaire and behavioral data. Therefore, the results of this study are valuable for understanding the population's acceptance of ADVs and the importance of objective measurements.

Despite these limitations, we are confident to show an effect of a self-explaining ADV based on subjective and objective data. As mentioned earlier, previous research explained trust as a combination of the communication style, feedback, and the anthropomorphic features of the ADV Belanche et al., 2012; Koo et al., 2015. In contrast, Hoff and Bashir state that trust is largely shaped by the personal traits of the usersHoff and Bashir, 2015. This is supported by newer findings in real driving scenarios, where personality traits were identified as relevant factors of trust Stephan, 2019, and were only out weighted by the actual driving performance. These factors are summarized under "dispositional trust," which consists of age and gender, and personality traits. In line with previous research our study could show that the demographic factors have a higher impact on acceptance compared to the ADVs' features.

Nevertheless, our findings are not generalizable over all demographic groups, since their communication needs are different: While we see a positive influence of the talking car in one group, the second group may view the in-vehicle information as excess and distracting. Following, the user-specific communication could increase trust in doubtful users - making them more confident to properly operate such a system since it might be able to increase system knowledge. However, there is a need for a further investigations using more extensive questionnaires to examine further modulators of acceptance, specifically trust in combination with more objective measurement instruments such as eye tracking. In the end, we argue that user specific in-vehicle communication can be useful to create guidelines for the further development of a safer and inclusive future of mobility.

4.6.8 SI: Calculated Effect Sizes for each significant factor

Table 4.35: Effect Sizes between different age groups on Intention to use, perceived usefulness, perceived ease of use and trust. Numbers in the table present the Cohen's D and in the case of difference Hedges G

age group	below 20	20 - 40	40 - 60	60 - 80	above 80
below 20	0	0.40	0.50	0.43	0.17 / 0.2
20 - 40	0.40	0	0.10	0.01	0.18 / 0.20
40 - 60	0.50	0.10	0	0.08	0.28
60 - 80	0.43	0.01	0.08	0	0.20 / 0.22
above 80	0.17 / 0.2	0.18 / 0.20	0.28	0.20 / 0.22	0

Table 4.36: Effect Sizes between different condition on Intention to use, perceived usefulness, perceived ease of use and trust. Numbers in the table present the Cohen's D and in the case of difference Hedges G

condition	AVAS	RadioTalk	TaxiDriver
AVAS	0	0.05	0.11
RadioTalk	0.99	0	0.06
TaxiDriver	0.11	0.06	0

4.6.9 Acknowledgment

The authors would like to thank all involved members of the WestDrive team that helped building software and hardware as well as acquiring data. Namely, we would like to thank: Johannes Max Pingel and Sumin Kim for helping to organize and commenting the experimental code, again Johannes Max Pingel for maintaining hardware during the duration of the exhibition. Deniz Gün and Thomas Klein for implementing the supplementary material into the exhibitions, as well as Philipp Spaniol for designing and building the exhibitions and lastly Marc Vidal de Palol and Hristofor Lukanov for implementing the Questionnaire, supplementary material,

Table 4.37: Effect Sizes between different combination of gender and age group on Intention to use, perceived usefulness, perceived ease of use and trust. Numbers in the table present the Cohen's D and in the case of difference Hedges G

gender = male					
Gender/Age Group	below 20	20 - 40	40 - 60	60 - 80	80+
Male / below 20	0	0.30	0.36	0.39	0.28 - 0.32
Male / 20 - 40	0.30	0	0.06	0.07	0
Male / 40 - 60	0.36	0.06	0	0.01	0.05
Male / 60 - 80	0.39	0.07	0.01	0	0.06
Male / 80+	0.28 - 0.32	0	0.05	0.06	0
Female / below 20	0.31	0.01	0.05	0.06	0.04
Female / 20 - 40	0.73	0.42	0.36	0.35	0.39 - 0.41
Female / 40 - 60	0.90	0.58	0.51	0.51	0.53 - 0.56
Female / 60 - 80	0.78 - 0.80	0.46	0.40	0.40	0.42 - 0.44
Female / 80+	0.29 - 0-33	0	0.03	0.05	0.01
gender = female					
Gender/Age Group	below 20	20 - 40	40 - 60	60 - 80	80+
Male / below 20	0.31	0.73	0.9	0.78 - 0.80	0.29 - 0-33
Male / 20 - 40	0.01	0.42	0.58	0.46	0
Male / 40 - 60	0.05	0.36	0.51	0.40	0.03
Male / 60 - 80	0.06	0.35	0.51	0.40	0.05
Male / 80+	0.04	0.39 - 0.41	0.53 - 0.56	0.42 - 0.44	0.01
Female / below 20	0	0.42	0.57	0.46	0.01
Female / 20 - 40	0.42	0	0.15	0.03	0.37 - 0.40
Female / 40 - 60	0.57	0.15	0	0.11	0.51 - 0.55
Female / 60 - 80	0.46	0.03	0.11	0	0.41 - 0.44
Female / 80+	0.01	0.37 - 0.40	0.51 - 0.55	0.41 - 0.44	0

and server structure. Without the support of all these wonderful members of the institute, this project would not have been realized.

4.6.10 Authors Contribution

FNN and MAW developing and designing and conducting the experiment as well as analysis of the questionnaire data. SD analysed the headtracking data and contributed to the writing process. SD, FNN and MAW share first authorship. AK and AC assisted in the analysis of the data as well as reviewing the manuscript, HL and MV has developed the questionnaire used in the experiment. PK and GP supervised the project and reviewed the manuscript and share senior authorship.

4.6.11 Conflict of Interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

4.6.12 Funding

This work is funded by the University of Osnabrück in cooperation with the graduate college “Vertrauen und Akzeptanz in erweiterten und virtuellen Arbeitswelten” (FN), the German Federal Ministry of Education and Research for the project ErgoVR (Entwicklung eines Ergonomie-Analyse-Tools in der virtuellen Realität zur Planung von Arbeitsplätzen in der industriellen Fertigung)-16SV8052(AK), as well as from the GMH foundation (MW).

4.6.13 Datasets

Code of the entire experiment conducted in this article is available in project-westdrive Gitlab repository under creative common license. furthermore all analysis scripts and output resuts, as well as raw data and demo video are available at OSF under creative common license.

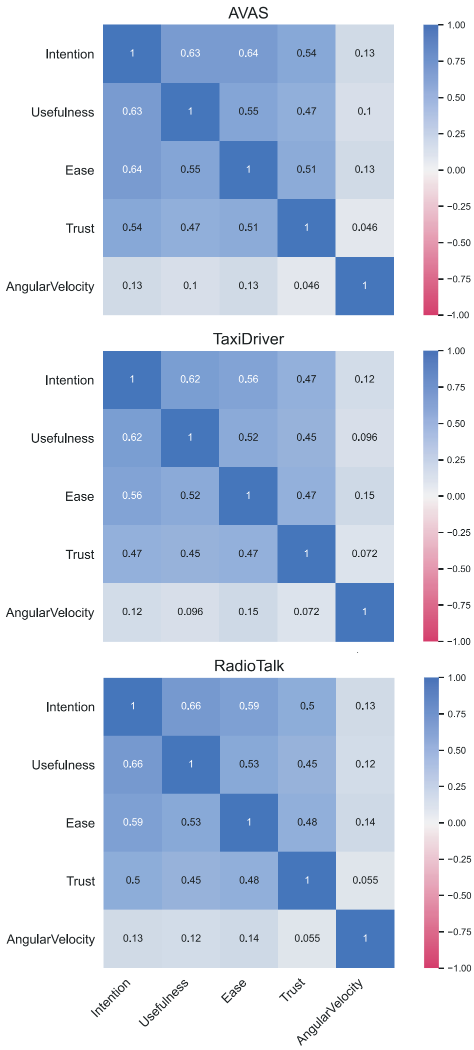

Figure 4.14: Summary of the Pearson correlation between the head angular velocity and the TAM questionnaire items. The correlation p value for Intention, usefulness, ease of use and trust are as follows for each condition: a) AVAS (Intention: $p < 0.001$, Usefulness: $p < 0.0010$, Ease of use: $p < 0.001$, Trust: $p < 0.01$) b) RadioTalk (Intention: $p < 0.001$, Usefulness: $p < 0.001$, Ease of use: $p < 0.001$, Trust: $p < 0.001$) c) TaxiDriver (Intention: $p < 0.001$, Usefulness: $p < 0.001$, Ease of use: $p < 0.001$, Trust: $p < 0.001$)

5 General Discussion

Toward real-life experimentation, this thesis tries to argue for the use of virtual reality to improve the validity of the findings of our experiments. Performing simulations, especially in driving or self-driving vehicle research, is not a new idea in science. The idea of computer simulation itself dates back to the era of World War II and the mathematicians, Jon Von Neumann and Stanislaw Ulam (University of Houston, 2000). One can argue that any experimentation is a simplified simulation of a real-world phenomenon. Some experiments can arguably already be conducted in a real-life scenario. However, that is not the case for all fields of study. For self-driving cars, simulators are often bulky, expensive, and nonetheless not that close to reality. Other studies, such as spatial navigation, can be done in real-life environments.Yet despite this possibility , in real-life experimentation, researchers will eventually need to deal with various sources of unaccounted noise. We live in a boisterous environment, and as such, there will be undoubtedly some of this noise present in the data gathered and can affect the researcher's conclusion if not careful enough. An example of such noises is the electromagnetic field produced by the alternating current, namely AC, electricity passing through the power lines, affecting the quality of EEG data. Although some of these noises are systematic and well studied, we still cannot account for all possible noise sources on the data and remove them to clean up the data. Here virtual reality offers a solution. It offers a cost-effective, realistic simulation and experimental environment that can be conducted in areas with controlled or at least known possible noises. These reasons were the primary motivation behind the development of both projects, Westdrive, and Westdrive LoopAR. This environment was designed with the open-source and open science philosophy in mind and tuned to be suitable for research on self-driving vehicles and other fields of research such as social interaction or spatial navigation. While project Westdrive showed us the limitations of a realistic simulation, such as computational limitations or the cost of a physically-based simulation project, LoopAR tried to overcome those simulations. Utilizing dynamic loading and partial in time data saving to avoid any data loss, optimizing the physics and path creation, and integrating real-world terrain data from OpenStreetMap framework, LoopAR offered a performant yet huge (25 square kilometer) environment. These projects help us learn the best practices in creating virtual worlds and designing virtual reality experiments. They also were created in order to be utilized by other researchers with few modifications to conduct studies in face perception, spatial navigation, and even more complex driving scenarios.

Investigating the benefits of experimentation using immersive virtual reality environments, one should not neglect possible challenges that arise simply from what is possible with today's hardware. Although today's virtual reality hardware is not like Sword of Damocles (Virtual Reality Society, 2017) anymore, that is so heavy that it needs to be suspended from the ceiling of the lab, a compelling computational power that enables running acceptably realistic experiences is still needed.. Furthermore, there are movements in creating standards on the software side, namely OpenVR, so that the simulation, or the environment itself, is not dependent on one specific brand of virtual reality device. There is still a significant variation of devices that differ in many fundamental aspects from each other. These differences can include but are not limited to the type of spatial tracking, the field of view, display resolution, and technology to the type of sensors used in the device itself. Here we have investigated one such aspect by comparing a new eye integrated eye tracker from Varjo VR and Finnish tech startup with the integrated eye tracker from Tobii inside the Vive pro virtual reality glasses from HTC, a well-established name among the virtual reality hardware producers. Although our tests, which were an extension of the test battery from Ehinger et al., 2019, indicate that both eye trackers are precise enough for conducting scientific experiments, there is still a significant difference in the quality of the data. Not only is one hardware inherently better than the other, but each hardware performed better in specific measurements such as blink detection, where Tobii outperforms Varjo with a large margin. Therefore when one decides on conducting their experiments using virtual reality, there should be caution and careful attention when choosing the hardware that the researcher will, conduct their experiment with in the end.

When designing and developing virtual reality-based experiments, the choice of appropriate hardware is only one part of the consideration to ensure a valid experiment. Currently, most such experiments are created using development environments and game engines, to be more precise, such as Unity3D and Unreal engine. The reasoning behind the choices is evident as both of these tools are free and can be used without any further cost if there is no financial gain planned for the project. They support almost all possible platforms as the final target of the project and the most commonly used virtual reality hardware that can be bought today. Nevertheless, one should remember that these tools were designed as game engines, and therefore they are inherently not optimized or made for scientific experimentation. They are made with maximizing the visual fidelity of the rendered scene in mind rather than precise timing or persistence between runs. Wiesing et al., 2020 showed in his work the hardware delay between the intended stimuli onset and when the stimuli are presented to the participant on the virtual reality glasses. Although these slight delays do not affect how the users interact with games, they are crucial in studies with a high temporal resolution, such as studies that use neuro-imaging techniques such as EEG. Although Wiesing et al., 2020 offers a simple solution for one of these systematic delays for Unreal engine that relates to the vertical sync of the virtual reality glasses displays, the same solution might not work for all sources of delay or even be generalized to different game engines. The different game engine uses different strategies while computing the programmed code and use different pathways for physics, graphics, and audio processing which means that the experiment might need to deal with various sources of delays, in the end, it depends on the complexity of cues. One such limitation investigated in this dissertation is the lack of proper networking solutions that prioritize simplicity, speed, and accuracy, which can be used to perform join the action, social interaction, or hyper scanning studies. Here we were able to offer a lightweight solution for such networking multi-participant needs. However, this solution will also inherit the issues mentioned earlier associated with the game engine the solution was developed for, here Unity3D, with higher complexity. Despite the great potential and what virtual reality experiments can offer regarding close to real-world experimentation plus improving the ecological validity of experiments, in the sense of mundane reality, this dissertation calls for the development of a scientific engine that functions as a foundation for the development of scientific experiments.

Although, We do hope to observe human cognition unfold in its full complexity in a dynamic, real-life environment as described by Parada, 2018, discoveries in cognitive science and similar field are made in small incremental steps. There is still a lot that we do not know yet about human cognitive processes. Even in the reductionist simple lab environment, there have been many new and valid discoveries during past decades. However, 4E cognition view as described in depth in the 4E cognition book by Newen et al., 2018 tries to compel and convince us that our cognition is not simply the process of a single isolated brain. It tries to argue that cognition emerges through our interactions and enactment with our environment and can be extended to the mundane objects in our reality.In simpler words not only every day objects such as a cellphone, a simple pen, and paper, or a printed map can affect our cognition but essentially become part of our cognitive processes. This premise is the foundation for the arguments of this dissertation, observing behavior and measuring under natural interaction.

This premise is what we have tried to investigate. When asked to lift or use a tool, observing participants' gaze behavior yields the same results as the study, which it tries to replicate. However, the original tool interaction study by M. F. Land and Hayhoe, 2001 only used viewing tools on the 2d monitor and contained no interaction with the presented use. Participants had to imagine using or lifting the presented tool. In contrast, in the study performed in virtual reality, they could lift or mimic using the tools with the help of standard controllers, which come with a full HTC Vive pro eye virtual reality headset. It is indeed a piece of good news that we could successfully replicate the results of an otherwise previously Lab-based experiment. However, this can lead to a valid concern. If we can replicate the exact findings of a traditional Lab-based experiment, why should any researcher invest the time and effort to design a virtual reality experiment if there is no new information to be learned. Although a valid question; however, as mentioned before based of 4E cognition view, our cognitive processes emerge in "natural" interaction with the environment.In other words using virtual reality might result in finding that could not be observed in the traditional lab based experiments.Here in the virtual reality based study with tool interaction, when participants use their hands, in contrast to controllers, to interact with the tools, the previously insignificant orientation of the tools presented turned into a significant effect (Keshava et al., 2021. This study also indicates that a simple act of actual motor and bodily adjustments necessary to grasp an object with different orientations can affect our eye movement and planning. This finding indicates that realism lies not only in cue or environment but rather in natural interaction with the environment. It is simply different if one should use their hands to interact with objects which require higher levels of cognitive planning than just click a mouse button or a key on the keyboard as it often is with traditional Lab experiments. Virtual reality experiments not only comply well with 4E cognition movement but also Keshava et al., 2021 showed that this realism and natural interaction indeed matter in understanding our cognitive processes.

Knowing the importance of natural interaction, we have further investigated planning as a cognitive process under a simulated sorting task closely resembling a factory. In this study resemblance and realism does not lie within simulating a factory. Instead, the realism here refers to the full-body movement involved to grasp and move virtual objects. In this planning experiment since the shelf was of actual size, people needed to grasp some objects from areas considered harder to reach than others. Interestingly, the study's primary goal was to investigate the ergonomics of the work environment and underlying mechanisms of human task planning. However, we could observe unaccounted artifacts originating from the full-body interaction. We can observe that participants are often unwilling to move the objects for far distances but instead move them only to the neighboring empty place on the shelf. The two studies combined, therefore, show the importance of the ecological validity lies within the natural interaction with the environment that will activate the entire underlying cognitive processes.

After learning the challenges and the importance of embodiment and interaction with the environment, we can finally observe behavior unfolds. There are already many works, especially by Mel Slater (Slater, 2009; Slater et al., 1994; Slater and Wilbur, 1997) on the role and depth of presence, body perception, and modification of the self-body perception in virtual reality. There are particular modifications and manipulation that are either impossible or extremely hard to perform in either reality or the lab environment. The case of the trolley dilemma, specifically with regards to self-driving vehicles, is one such example. However, with a careful presentation of environment and experimental conditions, setting participants in different perspectives of such a dilemma situation revealed a better, or relatively more complete understanding of people's moral judgments toward the correct decision in such situations. The immersion of virtual reality experiences can engage participants emotionally and directly expose them to the situation. In such cases, the behavior that emerged from the environment is hoped to approximate the natural behavior. Here, one can observe a more fine grade opinion on the right choice in a trolley dilemma scenario by asking how much participants agree with their own choice, notably when participants are put in the perspective of the group being hit by the car. Therefore it is apparent that immersion and sense of presence in the environment does affect our perception and consequently can affect our responses to the observed experimental cues or conditions.

Under the right experimental design, even being a passive viewer in an immersive experience can elicit desired behavior under experimentation. Being immersed in a virtual city, filled with dynamic actors such as other cars and pedestrians and reacting to the self-driving car's actions and explanations enriches our current understanding of the general public acceptance of self-driving cars. In the talking cars experiments, participants only observed one experimental condition in a short ninety-second drive. However, they were placed inside a real car cut in half before putting on the virtual reality glasses. Thanks to the power of simulation, although realistic and immersive, there was complete control over what every participant observed. Since the goal of this study was to compare acceptance to a well-known condition such as a ride in a taxi, the change in acceptance, notably trust and intention to use after such a short experience, was a significant discovery. However, here, we can observe the change through a simplified self-report questionnaire and the lens of behavioral data such as head movements. Altogether, these experiments testify the opportunity immersive virtual reality experiences can offer. In the end, such experiences under the presentation of valid cues and immersion can elicit real-life and natural behavior, which can help us study cognition as it is in mundane real-life situations.

5.1 A middle-ground for real-life experimentation

This dissertation suggests a solid case for experimentation following the banner of 4E cognition and closer to real-life observations. Immersiorama may be a made-up name, yet believes that immersing individuals in realistic virtual environments will consequently lead to the observation of natural behavior as a contrast to the artificial environment of our traditional laboratories. There is much that is yet not understood about human cognitive processes and cognition. However, we also cannot learn the in-depth truth about cognition if experimentation in the field always attempts to dissect the complex interconnected cognitive processes in isolated, austere environments. In addition, this dissertation's goal is not to undermine the validity and importance of low-level lab-based experimentation. Until the whole scientific community concurs to move entirely to real-life experimentation, it can offer a middle ground to perform lab-based close to reality experimentation that is more in line with the nature of our cognition.

5.2 Concluding remarks

The 4E cognition already emphasizes the environment and interactions (Newen et al., 2018). Whether one believes in the philosophy of the movement or not, there is an increasing body of evidence on how social interaction and engaging in a task with others can affect our cognitive processes down to the neuronal level (Czeszumski et al., 2019; Czeszumski et al., 2020). One should not view virtual reality just as another experimentation method or tool, but virtual reality can be view as the new lab itself. Virtual reality combined with other modern technologies such as the modern telecommunication networks and internet as well as modern computation hardware and measurement sensors offers a unique opportunity for the researcher. In this new lab, one can study human cognition with relative ease, not isolated, deprived of social interaction with just a screen and keyboard. It offers the opportunity to study cognition in its full complexity under any imaginable environment, even what is yet impossible in our reality. Indeed, there is a need for better hardware and advancement in our methods and tools. However, there is also a need for advancement to our view on studying cognition, be it cognitive science or related fields. As such, until new ideas replace that of today, virtual reality experimentation, MoBI movement, and real-world experimentation might help us gain a deeper and better understanding of our cognition.

Disclaimer

The ethics committees of the University of Osnabrück have approved all experiments reported in this dissertation per the Declaration of Helsinki. Furthermore, All measurements within the faculty in 2020 and 2021 were carried out with a hygiene concept approved by the university. I hereby confirm that I wrote this thesis independently and that I have not used resources other than those indicated. I guarantee that I significantly contributed to all materials used in this thesis. Furthermore, this thesis was neither published in Germany nor abroad, except the mentioned parts, and has not been used to fulfill any other examination requirements.

Bibliography

Abdi, H., & Williams, L. J. (2010). Tukey's honestly significant difference (hsd) test. *Encyclopedia of research design*, *3*(1), 1–5.

Abe, G., Itoh, M., & Yamamura, T. (2011). Effective and acceptable forward collision warning systems based on relationships between car-following behaviour and reaction to deceleration of lead vehicle. In P. C. Cacciabue, M. Hjälmdahl, A. Luedtke, & C. Riccioli (Eds.), *Human modelling in assisted transportation* (pp. 155–164). Springer Milan. https://doi.org/10.1007/978-88-470-1821-1_16

Ahmad, S. (2020, February 14). *HTC Vive: Vor- und Nachteile des VR-Headsets World of VR* [World of VR].

Allahyari, T., Saraji, G. N., Adi, J., Hosseini, M., Iravani, M., Younesian, M., & Kass, S. J. (2008). Cognitive failures, driving errors and driving accidents. *International journal of occupational safety and ergonomics*, *14*(2), 149–158.

Anthes, C., Garcia-Hernandez, R. J., Wiedemann, M., & Kranzlmuller, D. (2016). State of the art of virtual reality technology. *2016 IEEE Aerospace Conference*, 1–19. https://doi.org/10.1109/aero.2016.7500674

Arbib, M. A., Bonaiuto, J. B., Jacobs, S., & Frey, S. H. (2009). Tool use and the distalization of the end-effector. *Psychological research*, *73*(4), 441–462. https://doi.org/10.1007/s00426-009-0242-2

Ateş, C., Kaymaz, Ö., Kale, H. E., & Tekindal, M. A. (2019). Comparison of test statistics of nonnormal and unbalanced samples for multivariate analysis of variance in terms of Type-I error rates. *Comput. Math. Methods Med.*, *2019*, 2173638.

Audi. (2017). *Audi technology portal - audi q7 traffic jam assist* [Audi technology portal].

Awad, E., D'Souza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., & Rahwan, I. (2018). The Moral Machine experiment. *Nature*, *563*(7729), 59–64. https://doi.org/10.1038/s41586-018-0637-6

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of cognitive neuroscience*, *7*(1), 66–80. https://doi.org/10.1162/jocn.1995.7.1.66

Ballard, D. H., Kit, D., Rothkopf, C. A., & Sullivan, B. (2013). A hierarchical modular architecture for embodied cognition. *Multisensory research*, *26*(1-2), 177–204.

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*, 328. https://doi.org/10.3389/fpsyg.2013.00328

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bashiri, A., Ghazisaeedi, M., & Shahmoradi, L. (2017). The opportunities of virtual reality in the rehabilitation of children with attention deficit hyperactivity disorder: a literature review. *Korean Journal of Pediatrics*, *60*(11), 337. https://doi.org/10.3345/kjp.2017.60.11.337

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baumard, J., Osiurak, F., Lesourd, M., & Le Gall, D. (2014). Tool use disorders after left brain damage. *Frontiers in psychology*, *5*, 473. https://doi.org/10.3389/fpsyg.2014.00473

Belanche, D., Casaló, L. V., & Flavián, C. (2012). Integrating trust and personal values into the Technology Acceptance Model: The case of e-government services adoption. *Cuadernos de Economía y Dirección de la Empresa*, *15*(4), 192–204. https://doi.org/10.1016/j.cede.2012.04.004

Belardinelli, A., Barabas, M., Himmelbach, M., & Butz, M. V. (2016). Anticipatory eye fixations reveal tool knowledge for tool interaction. *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale*, *234*(8), 2415–2431. https://doi.org/10.1007/s00221-016-4646-0

Belardinelli, A., Herbort, O., & Butz, M. V. (2015). Goal-oriented gaze strategies afforded by object interaction. *Vision research*, *106*, 47–57. https://doi.org/10.1016/j.visres.2014.11.003

Belardinelli, A., Lohmann, J., Farnè, A., & Butz, M. V. (2018). Mental space maps into the future. *Cognition*, *176*, 65–73. https://doi.org/10.1016/j.cognition.2018.03.007

Belardinelli, A., Stepper, M. Y., & Butz, M. V. (2016). It's in the eyes: planning precise manual actions before execution. *Journal of vision*, *16*(1), 18. https://doi.org/10.1167/16.1.18

Bengler, K., Rettenmaier, M., Fritz, N., & Feierle, A. (2020). From HMI to HMIs: towards an HMI framework for automated driving. *Information*, *11*(2), 61. https://doi.org/10.3390/info11020061

Benleulmi, A. Z., & Blecker, T. (2017). Investigating the factors influencing the acceptance of fully autonomous cars. *Digitalization in Supply Chain Management and Logistics: Smart and Digital Solutions for an Industry 4.0 Environment. Proceedings of the Hamburg International Conference of Logistics (HICL), Vol. 23*, 99–115.

Bergmann, L. T., Schlicht, L., Meixner, C., König, P., Pipa, G., Boshammer, S., & Stephan, A. (2018a). Autonomous vehicles require socio-political acceptance—an empirical and philosophical perspective on the problem of moral decision making. *Frontiers in behavioral neuroscience*, *12*, 31.

Bergmann, L. T., Schlicht, L., Meixner, C., König, P., Pipa, G., Boshammer, S., & Stephan, A. (2018b). Autonomous vehicles require socio-political acceptance—an empirical and philosophical perspective on the problem of moral decision making. *Frontiers in Behavioral Neuroscience*, *12*. https://doi.org/10.3389/fnbeh.2018.00031

Berkman, M. I. (2018). History of virtual reality. In N. Lee (Ed.), *Encyclopedia of computer graphics and games* (pp. 1–9). Springer International Publishing. https://doi.org/10.1007/978-3-319-08234-9_169-1

Berti, A., & Frassinetti, F. (2000). When far becomes near: remapping of space by tool use. *Journal of cognitive neuroscience*, *12*(3), 415–420. https://doi.org/10.1162/089892900562237

Bischof, W. F., & Boulanger, P. (2003). Spatial navigation in virtual reality environments: an EEG analysis. *CyberPsychology & Behavior*, *6*(5), 487–495. https://doi.org/10.1089/109493103769710514

Blender Online Community. (2018). *Blender – A 3D Modelling and Rendering Package*. Blender Foundation. Amsterdam, The Netherlands.

Bonnefon, J., Shariff, A., & Rahwan, I. (2019). The trolley, the bull bar, and why engineers should care about the ethics of autonomous cars [point of view].

Proceedings of the IEEE, *107*(3), 502–504. https://doi.org/10.1109/jproc.2019.2897447

Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2015). Autonomous vehicles need experimental ethics: are we ready for utilitarian cars. *arXiv preprint arXiv:1510.03346*.

Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, *352*(6293), 1573–1576. https://doi.org/10.1126/science.aaf2654

Borenstein, J., Herkert, J., & Miller, K. W. (2019). Autonomous vehicles and the ethical tension between occupant and non-occupant safety. https://doi.org/10.25884/2vx8-3c55

Botica, N., Martins, M., Ribeiro, M. d. C. F., & Magalhães, F. (2015). *3d representation of the urban evolution of braga using the cityengine tool*. Vest-Agder-museet.

Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: the costs and benefits of shared gaze during collaborative search. *Cognition*, *106*(3), 1465–1477. https://doi.org/10.1016/j.cognition.2007.05.012

Brooks, F. (1999). What's real about virtual reality? *IEEE Computer Graphics and Applications*, *19*(6), 16–27. https://doi.org/10.1109/38.799723

Brunswick, E. (1956). *Perception and the representative design of psychological experiments*. University of California Press.

Brunswik, E. (1943). Organismic achievement and environmental probability. *Psychological Review*, *50*(3), 255–272. https://doi.org/10.1037/h0060889

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*(3), 193–217. https://doi.org/10.1037/h0047470

Burke, R. R. (2018). Virtual reality for marketing research. In L. Moutinho & M. Sokele (Eds.), *Innovative research methodologies in management* (pp. 63–82). Springer International Publishing. https://doi.org/10.1007/978-3-319-64400-4_3

Buswell, G. T. (1935). How people look at pictures: a study of the psychology and perception in art. *Chicago University Press*, *198*.

Carranza-García, M., Torres-Mateo, J., Lara-Benítez, P., & García-Gutiérrez, J. (2021). On the performance of one-stage and two-stage object detectors in autonomous vehicles using camera data. *Remote Sensing*, *13*(1), 89.

Caserman, P., Garcia-Agundez, A., Konrad, R., Göbel, S., & Steinmetz, R. (2019). Real-time body tracking in virtual reality using a vive tracker. *Virtual Reality*, *23*(2), 155–168. https://doi.org/10.1007/s10055-018-0374-z

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of vision*, *9*(3), 6.1–15. https://doi.org/10.1167/9.3.6

Castelli, L., Latini Corazzini, L., & Geminiani, G. C. (2008). Spatial navigation in large-scale virtual environments: gender differences in survey tasks. *Computers in Human Behavior*, *24*(4), 1643–1667. https://doi.org/10.1016/j.chb.2007.06.005

Chehri, A., & Mouftah, H. T. (2019). Autonomous vehicles in the sustainable cities, the beginning of a green adventure. *Sustainable Cities and Society*, *51*, 101751.

Chen, K., & Chan, A. H. S. (2011). A review of technology acceptance by older adults. *Gerontechnology*, *10*(1), 1–12.

Chicchi Giglioli, I. A., Pravettoni, G., Sutil Martín, D. L., Parra, E., & Raya, M. A. (2017). A novel integrating virtual reality approach for the assessment of the attachment behavioral system. *Frontiers in psychology*, *8*, 959.

Choi, B. C., & Pak, A. W. (2005). Peer reviewed: a catalog of biases in questionnaires. *Preventing chronic disease*, *2*(1).

CityEngine, E. (2013). 3d modeling software for urban environments. *Esri*.

Clay, V., König, P., & König, S. U. (2019). Eye tracking in virtual reality. https://doi.org/10.16910/jemr.12.1.3

Coeckelbergh, M. (2016). Responsibility and the moral phenomenology of using self-driving cars. *Applied Artificial Intelligence*, *30*(8), 748–757. https://doi.org/10.1080/08839514.2016.1229759

Corbeil, R. R., & Searle, S. R. (1976). Restricted maximum likelihood (REML) estimation of variance components in the mixed model. *Technometrics: a journal of statistics for the physical, chemical, and engineering sciences*, *18*(1), 31–38. https://doi.org/10.1080/00401706.1976.10489397

Cowan, W. M., Harter, D. H., & Kandel, E. R. (2000). The emergence of modern neuroscience: some implications for neurology and psychiatry. *Annual Review of Neuroscience*, *23*(1), 343–391. https://doi.org/10.1146/annurev.neuro.23.1.343

Cruden, S. (n.d.). *Automotive driving simulators* [Cruden].

Czaja, S. J., Sharit, J., Charness, N., Fisk, A. D., & Rogers, W. (2001). The center for research and education on aging and technology enhancement (CREATE): a program to enhance technology for older adults. *Gerontechnology*, *1*(1), 50–59.

Czeszumski, A., Ehinger, B. V., Wahn, B., & König, P. (2019). The social situation affects how we process feedback about our actions. *Frontiers in Psychology*, *10*, 361. https://doi.org/10.3389/fpsyg.2019.00361

Czeszumski, A., Eustergerling, S., Lang, A., Menrath, D., Gerstenberger, M., Schuberth, S., Schreiber, F., Rendon, Z. Z., & König, P. (2020). Hyperscanning: a valid method to study neural inter-brain underpinnings of social interaction. *Frontiers in Human Neuroscience*, *14*, 39. https://doi.org/10.3389/fnhum.2020.00039

Dajsuren, Y., & van den Brand, M. (2019). Automotive software engineering: past, present, and future. In Y. Dajsuren & M. van den Brand (Eds.), *Automotive systems and software engineering: state of the art and future trends* (pp. 3–8). Springer International Publishing.

Davis, F. D., & Venkatesh, V. (1996). A critical assessment of potential measurement biases in the technology acceptance model: three experiments. *Int. J. Hum. Comput. Stud.*, *45*(1), 19–45.

De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, *14*(10), 441–447. https://doi.org/10.1016/j.tics.2010.06.009

de la Rosa, S., & Breidt, M. (2018). Virtual reality: a new track in psychological research. *British Journal of Psychology*, *109*(3), 427–430. https://doi.org/10.1111/bjop.12302

Department for Transport. (2013). Contributory factors for reported road accidents (RAS50).

Dietrich, M., & Weisswange, T. H. (2019). Distributive justice as an ethical principle for autonomous vehicle behavior beyond hazard scenarios. *Ethics and Information Technology*, *21*(3), 227–239. https://doi.org/10.1007/s10676-019-09504-3

Dogan, E., Honnêt, V., Masfrand, S., & Guillaume, A. (2019). Effects of non-driving-related tasks on takeover performance in different takeover situations in conditionally automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *62*, 494–504. https://doi.org/10.1016/j.trf.2019.02.010

Dong, C., Shaopeng, D., Ming, C., & Chenyue, S. (2019). The simulation and research of steel-work construction based on fuzor. *E3S Web of Conferences*, *136*, 03023. https://doi.org/10.1051/e3sconf/201913603023

Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., & Koltun, V. (2017). CARLA: an open urban driving simulator. *Conference on Robot Learning*, 1–16.

Droll, J. A., & Hayhoe, M. M. (2007). Trade-offs between gaze and working memory use. *Journal of experimental psychology. Human perception and performance*, *33*(6), 1352–1365. https://doi.org/10.1037/0096-1523.33.6.1352

Du, N., Haspiel, J., Zhang, Q., Tilbury, D., Pradhan, A. K., Yang, X. J., & Robert, L. P., Jr. (2019). Look who's talking now: implications of AV's explanations on driver's trust, AV preference, anxiety and mental workload. *Transp. Res. Part C: Emerg. Technol.*, *104*, 428–442.

Ehinger, B. V., Groß, K., Ibs, I., & König, P. (2019). A new comprehensive eye-tracking test battery concurrently evaluating the pupil labs glasses and the eyelink 1000. *PeerJ*, *7*, e7086.

Einhäuser, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of vision*, *8*(2), 2.1–19. https://doi.org/10.1167/8.2.2

Ekstrom, A. D., & Isham, E. A. (2017). Human spatial navigation: representations across dimensions and scales. *Current Opinion in Behavioral Sciences*, *17*, 84–89. https://doi.org/10.1016/j.cobeha.2017.06.005

End, A., & Gamer, M. (2017). Preferential processing of social features and their interplay with physical saliency in complex naturalistic scenes. *Frontiers in Psychology*, *8*, 418. https://doi.org/10.3389/fpsyg.2017.00418

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(2), 381–394. https://doi.org/10.1518/001872095779064555

Engel, A. K., Maye, A., Kurthen, M., & König, P. (2013). Where's the action? the pragmatic turn in cognitive science. *Trends in cognitive sciences*, *17*(5), 202–209. https://doi.org/10.1016/j.tics.2013.03.006

Epstein, R. A., Patai, E. Z., Julian, J. B., & Spiers, H. J. (2017). The cognitive map in humans: spatial navigation and beyond. *Nature Neuroscience*, *20*(11), 1504–1513. https://doi.org/10.1038/nn.4656

Faisal, A. (2017). Computer science: visionary of virtual reality. *Nature*, *551*(7680), 298–299. https://doi.org/10.1038/551298a

Fan, S., Dal Monte, O., & Chang, S. W. C. (2021). Levels of naturalism in social neuroscience research. *iScience*, *24*(7), 102702. https://doi.org/10.1016/j.isci.2021.102702

Fang, Y., Nakashima, R., Matsumiya, K., Kuriki, I., & Shioiri, S. (2015). Eye-Head Coordination for Visual Cognitive Processing (Z. Kapoula, Ed.). *PLOS ONE*, *10*(3), e0121035. https://doi.org/10.1371/journal.pone.0121035

Farnè, A., Iriki, A., & Làdavas, E. (2005). Shaping multisensory action-space with tools: evidence from patients with cross-modal extinction. *Neuropsychologia*, *43*(2), 238–248. https://doi.org/10.1016/j.neuropsychologia.2004.11.010

Faulhaber, A. K., Dittmer, A., Blind, F., Wächter, M. A., Timm, S., Sütfeld, L. R., Stephan, A., Pipa, G., & König, P. (2019). Human decisions in moral dilemmas are largely described by utilitarianism: virtual car driving study provides guidelines for autonomous driving vehicles. *Science and Engineering Ethics*, *25*(2), 399–418. https://doi.org/10.1007/s11948-018-0020-x

Favarò, F. M., Nader, N., Eurich, S. O., Tripp, M., & Varadaraju, N. (2017). Examining accident reports involving autonomous vehicles in California (X. Hu, Ed.). *Plos One*, *12*(9), e0184952. https://doi.org/10.1371/journal.pone.0184952

Federal Ministry of Transport and Digital Infrastructure. (2017). *Ethics commission report: automated and connected driving* (tech. rep.).

FeldmanHall, O., Mobbs, D., Evans, D., Hiscox, L., Navrady, L., & Dalgleish, T. (2012). What we say and what we do: the relationship between real and hypothetical moral choices. *Cognition*, *123*(3), 434–441. https://doi.org/10.1016/j.cognition.2012.02.001

Finch, H., & French, B. (2013). A monte carlo comparison of robust MANOVA test statistics. *J. Mod. Appl. Stat. Methods*, *12*(2), 35–81.

Finger, H., Goeke, C., Diekamp, D., Standvoß, K., & König, P. (2017). LabVanced: A Unified JavaScript Framework for Online Studies. *2017 International Conference on Computational Social Science IC2S2*.

Flanagan, J. R., Bowman, M. C., & Johansson, R. S. (2006). Control strategies in object manipulation tasks. *Current opinion in neurobiology*, *16*(6), 650–659. https://doi.org/10.1016/j.conb.2006.10.005

Forster, Y., Naujoks, F., & Neukum, A. (2017). Increasing anthropomorphism and trust in automated driving functions by adding speech output. *2017 IEEE Intelligent Vehicles Symposium (IV)*, 365–372.

Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research*, *51*(17), 1920–1931. https://doi.org/10.1016/j.visres.2011.07.002

Foundation, B. (n.d.). *Blender.org - home of the blender project - free and open 3d creation software* [Blender.org].

Francis, K. B., Howard, C., Howard, I. S., Gummerum, M., Ganis, G., Anderson, G., & Terbeck, S. (2016). Virtual morality: transitioning from moral judgment to moral action? *PLoS ONE*, *11*(10), e0164374. https://doi.org/10.1371/journal.pone.0164374

Freeman, D., Haselton, P., Freeman, J., Spanlang, B., Kishore, S., Albery, E., Denne, M., Brown, P., Slater, M., & Nickless, A. (2018). Automated psychological therapy using immersive virtual reality for treatment of fear of heights: a single-blind, parallel-group, randomised controlled trial. *The Lancet Psychiatry*, *5*(8), 625–632. https://doi.org/10.1016/s2215-0366(18)30226-8

Frith, C. D. (2007). The social brain? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1480), 671–678. https://doi.org/10.1098/rstb.2006.2003

Gerdes, J. C., Thornton, S. M., & Millar, J. (2019). Designing automated vehicles around human values. In *Road vehicle automation 6* (pp. 39–48). Springer International Publishing. https://doi.org/10.1007/978-3-030-22933-7_5

Gibson, J. J. (1977). The theory of affordances. *Hilldale, USA*, *1*(2), 67–82.

Gkartzonikas, C., & Gkritza, K. (2019). What have we learned? a review of stated preference and choice studies on autonomous vehicles. *Transportation Research Part C: Emerging Technologies*, *98*, 323–337. https://doi.org/10.1016/j.trc.2018.12.003

Gogoll, J., & Müller, J. F. (2016). Autonomous cars: in favor of a mandatory ethics setting. *Science and Engineering Ethics*. https://doi.org/10.1007/s11948-016-9806-x

Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). "take over!" how long does it take to get the driver back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*(1), 1938–1942. https://doi.org/10.1177/1541931213571433

Gold, N., Pulford, B. D., & Colman, A. M. (2015). Do as I say, don't do as I do: differences in moral judgments do not translate into differences in decisions in real-life trolley problems. *Journal of Economic Psychology*, *47*, 50–61. https://doi.org/10.1016/j.joep.2015.01.001

Gourlay, M. J., & Held, R. T. (2017). Head-mounted-display tracking for augmented and virtual reality. *Information Display*, *33*(1), 6–10.

Grafsgaard, J. F., Boyer, K. E., Wiebe, E. N., & Lester, J. C. (2012). Analyzing posture and affect in task-oriented tutoring. *Twenty-Fifth International FLAIRS Conference*.

Gray, D. E. (2021, November 27). *Doing research in the real world*. Sage.

Green, M. (2000). "how long does it take to stop?" methodological analysis of driver perception-brake times. *Transportation Human Factors*, *2*(3), 195–216. https://doi.org/10.1207/sthf0203_1

Griffiths, T. L. (2015). Manifesto for a new (computational) cognitive revolution. *Cognition*, *135*, 21–23. https://doi.org/10.1016/j.cognition.2014.11.026

Gunes, H., & Pantic, M. (2010). Dimensional emotion prediction from spontaneous head gestures for interaction with sensitive artificial listeners. *International conference on intelligent virtual agents*, 371–377.

Hammal, Z., Cohn, J. F., Heike, C., & Speltz, M. L. (2015). What can head and facial movements convey about positive and negative affect? *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, 281–287. https://doi.org/10.1109/ACII.2015.7344584

Hammond, K. R. (1998). Ecological validity: then and now.

Harrington, M. (2010). *The design of experiments in neuroscience*. Sage.

Harris, J. M. (2004). Binocular vision: moving closer to reality (J. M. T. Thompson, Ed.). *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, *362*(1825), 2721–2739. https://doi.org/10.1098/rsta.2004.1464

Hars, A. (2016). Top misconceptions of autonomous cars and selfdriving vehicles. thinking outside the box. *Innovation Briefs*.

Hayhoe, M. M. (2004). Advances in relating eye movements and cognition. *Infancy: the official journal of the International Society on Infant Studies*, *6*(2), 267–274. https://doi.org/10.1207/s15327078in0602_7

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of vision*, *3*(1), 49–63. https://doi.org/10.1167/3.1.6

Heitink, G. (1999). *Practical theology: history, theory, action domains: manual for practical theology*. Wm. B. Eerdmans Publishing.

Henderson, J. M. (2017). Gaze control as prediction. *Trends in cognitive sciences*, *21*(1), 15–23. https://doi.org/10.1016/j.tics.2016.11.003

Henderson, J. M., & Hayes, T. R. (2017). Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nature human behaviour*, *1*(10), 743–747. https://doi.org/10.1038/s41562-017-0208-0

Herbort, O., & Butz, M. V. (2011). Habitual and goal-directed factors in (everyday) object handling. *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale*, *213*(4), 371–382. https://doi.org/10.1007/s00221-011-2787-8

Herbort, O., & Butz, M. V. (2012). The continuous end-state comfort effect: weighted integration of multiple biases. *Psychological research*, *76*(3), 345–363. https://doi.org/10.1007/s00426-011-0334-7

Hermsdörfer, J., Li, Y., Randerath, J., Goldenberg, G., & Johannsen, L. (2012). Tool use without a tool: kinematic characteristics of pantomiming as compared to actual use and the effect of brain damage. *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale*, *218*(2), 201–214. https://doi.org/10.1007/s00221-012-3021-z

Hess, B. J. M. (2019, January 1). Chapter 10 - on the retinal correspondences across the binocular visual field. In S. Ramat & A. G. Shaikh (Eds.), *Progress in brain research* (pp. 139–156). Elsevier. https://doi.org/10.1016/bs.pbr.2019.04.006

Hoff, K. A., & Bashir, M. (2015). Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Hoffmann, J. (2003). Anticipatory behavioral control. In M. V. Butz, O. Sigaud, & P. Gérard (Eds.), *Anticipatory behavior in adaptive learning systems: foundations, theories, and systems* (pp. 44–65). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-45002-3_4

Holleman, G. A., Hooge, I. T. C., Kemner, C., & Hessels, R. S. (2020). The 'real-world approach' and its problems: a critique of the term ecological validity. *Frontiers in Psychology*, *0*. https://doi.org/10.3389/fpsyg.2020.00721

Holmqvist, K., Nyström, M., & Mulvey, F. (2012). Eye tracker data quality: what it is and how to measure it. *Proceedings of the Symposium on Eye Tracking Research and Applications*, 45–52. https://doi.org/10.1145/2168556.2168563

Holstein, T., & Dodig-Crnkovic, G. (2018). Avoiding the intrinsic unfairness of the trolley problem. *Proceedings of the International Workshop on Software Fairness*, 32–37. https://doi.org/10.1145/3194770.3194772

Hooge, I., & Camps, G. (2013). Scan path entropy and arrow plots: capturing scanning behavior of multiple observers. *Frontiers in psychology*, *4*, 996. https://doi.org/10.3389/fpsyg.2013.00996

Howard, D., & Dai, D. (2014a). Public perceptions of self-driving cars: the case of berkeley, california. prepared for the 93rd annual meeting of the transportation research board. *University of California, Berkeley–Department of City and Regional Planning. Berkeley, CA*, *94720*.

Howard, D., & Dai, D. (2014b). Public perceptions of self-driving cars: the case of berkeley, california. *Transportation research board 93rd annual meeting*, *14*, 1–16.

Hughes, J. T. (1988). The edwin smith surgical papyrus: an analysis of the first case reports of spinal cord injuries. *Spinal Cord*, *26*(2), 71–82. https://doi.org/10.1038/sc.1988.15

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS biology*, *3*(3), e79. https://doi.org/10.1371/journal.pbio.0030079

Ishii, H. (2010). Augmented reality: fundamentals and nuclear related applications. *International Journal of NUCLEAR SAFETY AND SIMULATION*, *1*.

Jamali, A., Yousefzadeh, C., McGinty, C., Bryant, D., & Bos, P. (2018). LC lens systems to solve accommodation/convergence conflict in three-dimensional and virtual reality displays. *Optical Engineering*, *57*(10), 105101. https://doi.org/10.1117/1.oe.57.10.105101

James, W. (2007). *The principles of psychology*. Cosimo Classics.

Jarosch, O., Bellem, H., & Bengler, K. (2019). Effects of task-induced fatigue in prolonged conditional automated driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *61*(7), 1186–1199. https://doi.org/10.1177/0018720818816226

Jarvis Thomson, J. (1985). The trolley problem. *Yale Law Journal*, *94*(6), 5.

Jeannerod, M. (2006). *Motor cognition: what actions tell the self*. OUP Oxford.

Johansson, R. S., Westling, G., Bäckström, A., & Flanagan, J. R. (2001). Eye-hand coordination in object manipulation. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *21*(17), 6917–6932. https://doi.org/10.1523/jneurosci.21-17-06917.2001

Johnson, S. H., & Grafton, S. T. (2003). From 'acting on' to 'acting with': the functional anatomy of object-oriented action schemata. In *Progress in brain research* (pp. 127–139). Elsevier.

Johnson, T. (2013). Enhancing safety through automation. *Society of Automotive Engineers Gov't Industry Meeting, Automation and Connected Vehicle Safety, NHTSA*.

Johnson-Frey, S. H. (2004). The neural bases of complex tool use in humans. *Trends in cognitive sciences*, *8*(2), 71–78. https://doi.org/10.1016/j.tics.2003.12.002

Ju, U., Kang, J., & Wallraven, C. (2019). To brake or not to brake? personality traits predict decision-making in an accident situation. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00134

Juliani, A., Berges, V.-P., Teng, E., Cohen, A., Harper, J., Elion, C., Goy, C., Gao, Y., Henry, H., Mattar, M., & Lange, D. (2020). Unity: a general platform for intelligent agents. *arXiv:1809.02627 [cs, stat]*.

Jung, J. H. (2019). Accommodation and convergence. In J.-S. Lee (Ed.), *Primary eye examination: a comprehensive guide to diagnosis* (pp. 31–36). Springer. https://doi.org/10.1007/978-981-10-6940-6_3

Kallioinen, N., Pershina, M., Zeiser, J., Nosrat Nezami, F., Pipa, G., Stephan, A., & König, P. (2019). Moral judgements on the actions of Self-Driving cars and human drivers in dilemma situations from different perspectives. *Front. Psychol.*, *10*, 2415.

Kallotech, F. (n.d.). *Fuzor*.

Keeling, G. (2017). Commentary: using virtual reality to assess ethical decisions in road traffic scenarios: applicability of value-of-life-based models and influences of time pressure. *Frontiers in Behavioral Neuroscience*, *11*, 247. https://doi.org/10.3389/fnbeh.2017.00247

Keeling, G. (2019). Why trolley problems matter for the ethics of automated vehicles. *Science and Engineering Ethics*. https://doi.org/10.1007/s11948-019-00096-1

Keeling, G., Evans, K., Thornton, S. M., Mecacci, G., & de Sio, F. S. (2019). Four perspectives on what matters for the ethics of automated vehicles. In *Road vehicle automation 6* (pp. 49–60). Springer International Publishing. https://doi.org/10.1007/978-3-030-22933-7_6

Keshava, A., Aumeistere, A., Izdebski, K., & Konig, P. (2020). Decoding task from oculomotor behavior in virtual reality. *ACM Symposium on Eye Tracking Research and Applications*, (Article 30), 1–5. https://doi.org/10.1145/3379156.3391338

Keshava, A., Gottschewsky, N., Balle, S., Nezami, F. N., Schüler, T., & König, P. (2021). Action Affordance Affects Proximal And Distal Goal-oriented Planning. *bioRxiv*, 2021–2007.

Kihlstrom, J. F. (2021). Ecological validity and “ecological validity”. *Perspectives on Psychological Science*, *16*(2), 466–471. https://doi.org/10.1177/1745691620966791

Kilteni, K., Groten, R., & Slater, M. (2012). The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, *21*(4), 373–387. https://doi.org/10.1162/PRES_a_00124

Kingstone, A., Smilek, D., Ristic, J., Kelland Friesen, C., & Eastwood, J. D. (2003). Attention, researchers! it is time to take a look at the real world. *Current Directions in Psychological Science*, *12*(5), 176–180. https://doi.org/10.1111/1467-8721.01255

Knoblich, G., & Jordan, J. S. (2003). Action coordination in groups and individuals: learning anticipatory control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(5), 1006–1016. https://doi.org/10.1037/0278-7393.29.5.1006

Knuth, K. H. (2013). Optimal data-based binning for histograms.

Kollmorgen, S., Nortmann, N., Schröder, S., & König, P. (2010). Influence of low-level stimulus features, task dependent factors, and spatial biases on overt visual attention. *PLoS computational biology*, *6*(5), e1000791. https://doi.org/10.1371/journal.pcbi.1000791

König, M., & Neumayr, L. (2017). Users' resistance towards radical innovations: the case of the self-driving car. *Transp. Res. Part F Traffic Psychol. Behav.*, *44*, 42–52.

König, P., Wilming, N., Kietzmann, T. C., Ossandón, J. P., et al. (2016). Eye movements as a window to cognitive processes. *Journal of eye movement research*, *9*(5).

König, S. U., Clay, V., Nolte, D., Duesberg, L., Kuske, N., & König, P. (2019). Learning of spatial properties of a large-scale virtual city with an interactive map. *Frontiers in Human Neuroscience*, *13*, 240. https://doi.org/10.3389/fnhum.2019.00240

Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L., & Nass, C. (2015). Why did my car just do that? explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *Int. J. Interact. Des. Manuf.*, *9*(4), 269–275.

Koo, J., Shin, D., Steinert, M., & Leifer, L. (2016). Understanding driver responses to voice alerts of autonomous car operations. *Int. J. Veh. Des.*, *70*(4), 377.

Króliczak, G., Cavina-Pratesi, C., Goodman, D. A., & Culham, J. C. (2007). What does the brain do when you fake it? an FMRI study of pantomimed and real grasping. *Journal of neurophysiology*, *97*(3), 2410–2422. https://doi.org/10.1152/jn.00778.2006

Krueger, R., Rashidi, T. H., & Rose, J. M. (2016). Preferences for shared autonomous vehicles. *Transp. Res. Part C: Emerg. Technol.*, *69*, 343–355.

Ladouce, S., Donaldson, D. I., Dudchenko, P. A., & Ietswaart, M. (2017). Understanding minds in real-world environments: toward a mobile cognition approach. *Frontiers in human neuroscience*, *10*, 694.

Lague, S. (2021, June 8). *SebLague/path-creator*.

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*(11), 1311–1328. https://doi.org/10.1068/p2935

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision research*, *41*(25-26), 3559–3565. https://doi.org/10.1016/s0042-6989(01)00102-x

Land, M. F., & McLeod, P. (2000). From eye movements to actions: how batsmen hit the ball. *Nature neuroscience*, *3*(12), 1340–1345. https://doi.org/10.1038/81887

Land, M. F. (2004). The coordination of rotations of the eyes, head and trunk in saccadic turns produced in natural situations. *Experimental Brain Research*, *159*(2), 151–160. https://doi.org/10.1007/s00221-004-1951-9

Land, M. F. (2006). Eye movements and the control of actions in everyday life. *Progress in retinal and eye research*, *25*(3), 296–324. https://doi.org/10.1016/j.preteyeres.2006.01.002

Land, M. F., & Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *352*(1358), 1231–1239.

LaViola, J. J. (2000). A discussion of cybersickness in virtual environments. *ACM SIGCHI Bulletin*, *32*(1), 47–56. https://doi.org/10.1145/333329.333344

Leahey, T. (1991). A history of modern psychology. https://doi.org/10.1016/c2013-0-11479-1

Lee, C., & Coughlin, J. F. (2015). PERSPECTIVE: older adults' adoption of technology: an integrated approach to identifying determinants and barriers. *J. Prod. Innov. Manage.*, *32*(5), 747–759.

Lee, C., Seppelt, B., Reimer, B., Mehler, B., & Coughlin, J. F. (2019). Acceptance of vehicle automation: effects of demographic traits, technology experience and media exposure. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 2066–2070. https://doi.org/10.1177/1071181319631425

Lee, C., Ward, C., Raue, M., D'Ambrosio, L., & Coughlin, J. F. (2017). Age differences in acceptance of self-driving cars: a survey of perceptions and attitudes. *Human Aspects of IT for the Aged Population. Aging, Design and User Experience*, 3–13.

Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors*, 31.

Lee, M., Sul, S., & Kim, H. (2018). Social observation increases deontological judgments in moral dilemmas. *Evolution and Human Behavior*, *39*(6), 611–621. https://doi.org/10.1016/j.evolhumbehav.2018.06.004

Lenth, R. (2018). *Emmeans: estimated marginal means, aka least-squares means*.

Lewis, J. R. (2018). The system usability scale: past, present, and future. *International Journal of Human–Computer Interaction*, *34*(7), 577–590. https://doi.org/10.1080/10447318.2018.1455307

Li, A., Montaño, Z., Chen, V. J., & Gold, J. I. (2011). Virtual reality and pain management: current trends and future directions. *Pain Management*, *1*(2), 147–157. https://doi.org/10.2217/pmt.10.15

Li, J., Zhao, X., Cho, M.-J., Ju, W., & Malle, B. F. (2016). From trolley to autonomous vehicle: perceptions of responsibility and moral norms in traffic accidents with self-driving cars. *SAE Technical Paper*. https://doi.org/10.4271/2016-01-0164

Li, S., Blythe, P., Edwards, S., Goodman, P., & Hill, G. (2019, October 1). *Investigation of the influence of multitasking on drivers' takeover performance in highly automated vehicles*.

Li, S., Zhang, J., Li, P., Wang, Y., & Wang, Q. (2019). Influencing factors of driving decision-making under the moral dilemma. *IEEE Access*, *7*, 104132–104142. https://doi.org/10.1109/access.2019.2932043

Lin, P. (2015). Why Ethics Matters for Autonomous Cars. In M. Maurer, J. C. Gerdes, B. Lenz, & H. Winner (Eds.), *Autonomes Fahren: Technische, rechtliche und gesellschaftliche Aspekte* (pp. 69–85). Springer. https://doi.org/10.1007/978-3-662-45854-9_4

Lindgren, T., Fors, V., Pink, S., & Osz, K. (2020). Anticipatory experience in everyday autonomous driving. *Personal and Ubiquitous Computing*, *24*(6), 747–762. https://doi.org/10.1007/s00779-020-01410-6

Lipton, L. (2012). Brief history of electronic stereoscopic displays. *Optical Engineering*, *51*(2), 021103. https://doi.org/10.1117/1.oe.51.2.021103

Lohmann, J., Belardinelli, A., & Butz, M. V. (2019). Hands ahead in mind and motion: active inference in peripersonal hand space. *Vision (Basel, Switzerland)*, *3*(2). https://doi.org/10.3390/vision3020015

London. (1801). *Section of the rotunda, leicester square | british library - picturing places* [The british library].

Lu, Y., & Sarter, N. (2019). Eye Tracking: A Process-Oriented Method for Inferring Trust in Automation as a Function of Priming and System Reliability. *IEEE Transactions on Human-Machine Systems*, *49*(6), 560–568. https://doi.org/10.1109/THMS.2019.2930980

Lüders, M., & Brandtzæg, P. B. (2017). 'my children tell me it's so simple': a mixed-methods approach to understand older non-users' perceptions of social networking sites. *New Media & Society*, *19*(2), 181–198. https://doi.org/10.1177/1461444814554064

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior research methods*, *49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Lundstrom, M. (2003). Moore's law forever? *Science*, *299*(5604), 210–211. https://doi.org/10.1126/science.1079567

Luzuriaga, M., Heras, A., & Kunze, O. (2019). Hurting others vs. hurting myself, a dilemma for our autonomous vehicle. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3345141

M., K., R., H., & J.c.f, W. (2015). Public opinion on automated driving: results of an international questionnaire among 5000 respondents. *Transp. Res. Part F Traffic Psychol. Behav.*, *32*.

MacGregor, J. N., & Chu, Y. (2011). Human performance on the traveling salesman and related problems: a review. *The Journal of Problem Solving*, *3*(2), 2.

Mandel, D. R., & Vartanian, O. (2007). Taboo or tragic: effect of tradeoff type on moral choice, conflict, and confidence. *Mind & Society*, *7*(2), 215–226. https://doi.org/10.1007/s11299-007-0037-3

Mandler, G. (2011, January 21). *A history of modern experimental psychology: from james and wundt to cognitive science*. MIT Press.

Mann, D. L., Nakamoto, H., Logt, N., Sikkink, L., & Brenner, E. (2019). Predictive eye movements when hitting a bouncing ball. *Journal of vision*, *19*(14), 28. https://doi.org/10.1167/19.14.28

Maravita, A., Spence, C., Kennett, S., & Driver, J. (2002). Tool-use changes multi-modal spatial interactions between vision and touch in normal humans.

Cognition, *83*(2), B25–34. https://doi.org/10.1016/s0010-0277(02)00003-3

Marberger, C., Mielenz, H., Naujoks, F., Radlmayr, J., Bengler, K., & Wandtner, B. (2018). Understanding and applying the concept of "driver availability" in automated driving. In N. A. Stanton (Ed.), *Advances in human aspects of transportation* (pp. 595–605). Springer International Publishing. https://doi.org/10.1007/978-3-319-60441-1_58

Marín-Morales, J., Higuera-Trujillo, J. L., Greco, A., Guixeres, J., Llinares, C., Scilingo, E. P., Alcañiz, M., & Valenza, G. (2018). Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors. *Scientific Reports*, *8*(1), 13657. https://doi.org/10.1038/s41598-018-32063-4

Mars, F., & Navarro, J. (2012). Where we look when we drive with or without active steering wheel control. *PloS one*, *7*(8), e43858. https://doi.org/10.1371/journal.pone.0043858

Marshall, A. (2018). Tesla's favorite autopilot safety statistic doesn't hold up. *Wired*.

Martin, R., Kusev, I., Cooke, A. J., Baranova, V., Schaik, P. V., & Kusev, P. (2017). Commentary: the social dilemma of autonomous vehicles. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00808

Masalonis, A. J., Duley, J. A., & Parasuraman, R. (1999). Effects of manual and autopilot control on mental workload and vigilance during simulated general aviation flight. *Transportation Human Factors*, *1*(2), 187–200. https://doi.org/10.1207/sthf0102_7

Matthis, J. S., Yates, J. L., & Hayhoe, M. M. (2018). Gaze and the control of foot placement when walking in natural terrain. *Current biology: CB*, *28*(8), 1224–1233.e5. https://doi.org/10.1016/j.cub.2018.03.008

Maurer, M. (2015). Einleitung. In M. Maurer, J. C. Gerdes, B. Lenz, & H. Winner (Eds.), *Autonomes Fahren: Technische, rechtliche und gesellschaftliche Aspekte* (pp. 1–8). Springer. https://doi.org/10.1007/978-3-662-45854-9_1

McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences*, *98*(20), 11832–11835. https://doi.org/10.1073/pnas.211415698

McLachlan, G. J. (1992). *Discriminant analysis and statistical pattern recognition*. John Wiley & Sons, Inc., New York. https://doi.org/10.1002/0471725293 A Wiley-Interscience Publication

Meder, B., Fleischhut, N., Krumnau, N.-C., & Waldmann, M. R. (2018). How should autonomous cars drive? A preference for defaults in moral judgments under risk and uncertainty. *Risk Analysis*, *39*(2), 295–314. https://doi.org/10.1111/risa.13178

Mehrabian, A. (2017). Communication without words. In *Communication theory* (pp. 193–200). Routledge.

Melcher, V., Rauh, S., Diederichs, F., Widlroither, H., & Bauer, W. (2015). Take-over requests for automated driving. *Procedia Manufacturing*, *3*, 2867–2873. https://doi.org/10.1016/j.promfg.2015.07.788

Mennie, N., Hayhoe, M., & Sullivan, B. (2007). Look-ahead fixations: anticipatory eye movements in natural tasks. *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale*, *179*(3), 427–442. https://doi.org/10.1007/s00221-006-0804-0

Mercer, C., & Macaulay, T. (2019). Which companies are making driverless cars? (Idg, Ed.). *Techworld*.

Miller, A. (2018). The effect of virtual reality education tools on the retention of information. *South Carolina Junior Academy of Science*.

Miller, G. A. (2003). The cognitive revolution: a historical perspective. *Trends in Cognitive Sciences*, *7*(3), 141–144. https://doi.org/10.1016/s1364-6613(03)00029-9

Mine, M. R. (1995). *Virtual environment interaction techniques*.

Montfoort, I., Frens, M. A., Hooge, I. T. C., Haselen, G. C. L.-v., & van der Geest, J. N. (2007). Visual search deficits in Williams-Beuren syndrome. *Neuropsychologia*, *45*(5), 931–938. https://doi.org/10.1016/j.neuropsychologia.2006.08.022

Morra, L., Lamberti, F., Prattico, F. G., Rosa, S. L., & Montuschi, P. (2019). Building trust in autonomous vehicles: role of virtual reality driving simulators in HMI design. *IEEE Transactions on Vehicular Technology*, *68*(10), 9438–9450. https://doi.org/10.1109/tvt.2019.2933601

Morton Leonard Heilig. (1962, August 28). *United states patent: 3050870* (pat. No. 3050870).

National Highway Traffic Safety Administration. (2008). *Report to congress: doths811059* (tech. rep.). U.S. Department of Transportation.

Navarro, J., Hernout, E., Osiurak, F., & Reynaud, E. (2020). On the nature of eye-hand coordination in natural steering behavior. *PloS one*, *15*(11), e0242818. https://doi.org/10.1371/journal.pone.0242818

Neisser, U. (1991). A case of misplaced nostalgia. *American Psychologist*, *46*(1), 34–36. https://doi.org/10.1037/0003-066x.46.1.34

Newen, A., Bruin, L. D., & Gallagher, S. (Eds.). (2018, September 13). *The oxford handbook of 4e cognition*. Oxford University Press.

Nezami, F. N., Wächter, M. A., Pipa, G., & König, P. (2020). Project westdrive: unity city with self-driving cars and pedestrians for virtual reality studies. *Frontiers in ICT*, *7*, 1. https://doi.org/10.3389/fict.2020.00001

Niehorster, D. C., Santini, T., Hessels, R. S., Hooge, I. T. C., Kasneci, E., & Nyström, M. (2020). The impact of slippage on the data quality of head-worn eye trackers. *Behavior research methods*, *52*(3), 1140–1160. https://doi.org/10.3758/s13428-019-01307-0

Niforatos, E., Palma, A., Gluszny, R., Vourvopoulos, A., & Liarokapis, F. (2020). Would you do it?: enacting moral dilemmas in virtual reality for understanding ethical decision-making. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3313831.3376788

Norman, D. A., & Shallice, T. (1986). Attention to action. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and Self-Regulation: advances in research and theory volume 4* (pp. 1–18). Springer US. https://doi.org/10.1007/978-1-4757-0629-1_1

Norman, D. A. (1990). The ‘problem ’ with automation: inappropriate feedback and interaction, not ‘over-automation’. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *327*(1241), 585–593. https://doi.org/10.1098/rstb.1990.0101

North, M. M., North, S. M., & Coble, J. R. (2002). Virtual reality therapy: an effective treatment for psychological disorders. In *Handbook of virtual environments*. CRC Press.

Nyholm, S. (2018a). The ethics of crashes with self-driving cars: A roadmap, I. *Philosophy Compass*, *13*(7), e12507. https://doi.org/10.1111/phc3.12507

Nyholm, S. (2018b). The ethics of crashes with self-driving cars: A roadmap, II. *Philosophy Compass*, *13*(7), e12506. https://doi.org/10.1111/phc3.12506

Oculus. (2021). *Oculus* [Oculus].

Ogdon, D. C. (2019). HoloLens and VIVE pro: virtual reality headsets. *Journal of the Medical Library Association*, *107*(1). https://doi.org/10.5195/jmla.2019.602

OpenStreetMap, F. (n.d.). *OpenStreetMap* [OpenStreetMap].

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *The Behavioral and brain sciences*, *24*(5), 939–73, discussion 973–1031. https://doi.org/10.1017/s0140525x01000115

Othman, K. (2021). Public acceptance and perception of autonomous vehicles: a comprehensive review. *AI and Ethics*, *63*.

Owens, J. D., Houston, M., Luebke, D., Green, S., Stone, J. E., & Phillips, J. C. (2008). GPU computing. *Proceedings of the IEEE*, *96*(5), 879–899. https://doi.org/10.1109/jproc.2008.917757

Ozana, A., Berman, S., & Ganel, T. (2018). Grasping trajectories in a virtual environment adhere to weber's law. *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale*, *236*(6), 1775–1787. https://doi.org/10.1007/s00221-018-5265-8

Pan, X., & Hamilton, A. F. d. C. (2018). Why and how to use virtual reality to study human social interaction: the challenges of exploring a new research landscape. *British Journal of Psychology*, *109*(3), 395–417. https://doi.org/10.1111/bjop.12290

Papadoulis, A., Quddus, M., & Imprialou, M. (2019). Evaluating the safety impact of connected and autonomous vehicles on motorways. *Accident Analysis & Prevention*, *124*, 12–22.

Parada, F. J. (2018). Understanding natural cognition in everyday settings: 3 pressing challenges. *Frontiers in human neuroscience*, *12*, 386.

Parada, F. J., & Rossi, A. (2020). Perfect timing: mobile brain/body imaging scaffolds the 4e-cognition research program. *European Journal of Neuroscience*.

Parkinson, C., Sinnott-Armstrong, W., Koralus, P. E., Mendelovici, A., McGeer, V., & Wheatley, T. (2011). Is morality unified? evidence that distinct neural systems underlie moral judgments of harm, dishonesty, and disgust. *Journal of Cognitive Neuroscience*, *23*(10), 3162–3180. https://doi.org/10.1162/jocn_a_00017

Parsons, T. D., Gaggioli, A., & Riva, G. (2017). Virtual reality for research in social neuroscience. *Brain sciences*, *7*(4), 42.

Parveau, M., & Adda, M. (2020). Toward a user-centric classification scheme for extended reality paradigms. *Journal of Ambient Intelligence and Humanized Computing*, *11*(6), 2237–2249. https://doi.org/10.1007/s12652-019-01352-9

Patil, I., Cogoni, C., Zangrando, N., Chittaro, L., & Silani, G. (2014). Affective basis of judgment-behavior discrepancy in virtual experiences of moral dilemmas.

Social Neuroscience, *9*(1), 94–107. https://doi.org/10.1080/17470919.2013.870091

Peck, T. C., Seinfeld, S., Aglioti, S. M., & Slater, M. (2013). Putting yourself in the skin of a black avatar reduces implicit racial bias. *Consciousness and Cognition*, *22*(3), 779–787. https://doi.org/10.1016/j.concog.2013.04.016

Pek, J., Wong, O., & Wong, A. C. M. (2018). How to address non-normality: a taxonomy of approaches, reviewed, and illustrated. *Front. Psychol.*, *9*, 2104.

Pelz, J. B., & Canosa, R. (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision research*, *41*(25-26), 3587–3596. https://doi.org/10.1016/s0042-6989(01)00245-0

Pezzulo, G., Zorzi, M., & Corbetta, M. (2021). The secret life of predictive brains: what's spontaneous activity for? *Trends in cognitive sciences*.

Pezzulo, G., Hoffmann, J., & Falcone, R. (2007). Anticipation and anticipatory behavior. *Cognitive processing*, *8*(2), 67–70. https://doi.org/10.1007/s10339-007-0173-z

Pizlo, Z., & Li, Z. (2005). Solving combinatorial problems: the 15-puzzle. *Memory & cognition*, *33*(6), 1069–1084. https://doi.org/10.3758/bf03193214

Poggio, G. F., & Poggio, T. (1984). The analysis of stereopsis. *Annual Review of Neuroscience*, *7*(1), 379–412. https://doi.org/10.1146/annurev.ne.07.030184.002115

Poyet, L. (1900, September 1). *Illustration of the cineorama balloon simulation, at the 1900 paris exposition*.

Prautzsch, H., Boehm, W., & Paluszny, M. (2002). *Bézier and b-spline techniques*. Springer-Verlag. https://doi.org/10.1007/978-3-662-04919-8

Proudlock, F. A., Shekhar, H., & Gottlob, I. (2003). Coordination of Eye and Head Movements during Reading. *Investigative Opthalmology & Visual Science*, *44*(7), 2991. https://doi.org/10.1167/iovs.02-1315

Psotka, J. (1995). Immersive training systems: virtual reality and education and training. *Instructional Science*, *23*(5), 405–431. https://doi.org/10.1007/bf00896880

R Core Team. (2018). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria.

Raats, K., Fors, V., & Pink, S. (2020). Trusting autonomous vehicles: An interdisciplinary approach. *Transportation Research Interdisciplinary Perspectives*, *7*, 100201. https://doi.org/10.1016/j.trip.2020.100201

Redcay, E., & Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews Neuroscience*, *20*(8), 495–505. https://doi.org/10.1038/s41583-019-0179-4

Reggente, N., Essoe, J. K.-Y., Aghajan, Z. M., Tavakoli, A. V., McGuire, J. F., Suthana, N. A., & Rissman, J. (2018). Enhancing the ecological validity of fMRI memory research using virtual reality. *Frontiers in Neuroscience*, *12*, 408. https://doi.org/10.3389/fnins.2018.00408

Reichelt, S., Häussler, R., Fütterer, G., & Leister, N. (2010). Depth cues in human visual perception and their realization in 3d displays. *Three-Dimensional Imaging, Visualization, and Display 2010 and Display Technologies and Applications for Defense, Security, and Avionics IV*, *7690*, 76900b. https://doi.org/10.1117/12.850094

Rezaei, A., & Caulfield, B. (2020a). Examining public acceptance of autonomous mobility. *Travel Behaviour and Society*, *21*, 235–246. https://doi.org/https://doi.org/10.1016/j.tbs.2020.07.002

Rezaei, A., & Caulfield, B. (2020b). Examining public acceptance of autonomous mobility. *Travel Behaviour and Society*, *21*, 235–246.

Rhim, J., Lee, G.-b., & Lee, J.-H. (2020). Human moral reasoning types in autonomous vehicle moral dilemma: A cross-cultural comparison of Korea and Canada. *Computers in Human Behavior*, *102*, 39–56. https://doi.org/10.1016/j.chb.2019.08.010

Riva, G. (2005). Virtual reality in psychotherapy: review. *CyberPsychology & Behavior*, *8*(3), 220–230. https://doi.org/10.1089/cpb.2005.8.220

Rosson, L. (2014, April 15). *The virtual interface environment workstation (VIEW), 1990* [Nasa].

Royzman, E. B., Landy, J. F., & Leeman, R. F. (2014). Are thoughtful people more utilitarian? CRT as a unique predictor of moral minimalism in the dilemmatic context. *Cognitive Science*, *39*(2), 325–352. https://doi.org/10.1111/cogs.12136

Rus-Calafell, M., Garety, P., Sason, E., Craig, T. J. K., & Valmaggia, L. R. (2018). Virtual reality in the assessment and treatment of psychosis: a systematic review of its utility, acceptability and effectiveness. *Psychological Medicine*, *48*(3), 362–391. https://doi.org/10.1017/s0033291717001945

Ryan, M. (2020). The future of transportation: ethical, legal, social and economic impacts of self-driving vehicles in the year 2025. *Sci. Eng. Ethics*, *26*(3), 1185–1208.

Sachdeva, S., Iliev, R., Ekhtiari, H., & Dehghani, M. (2015). The role of self-sacrifice in moral dilemmas (T. Boraud, Ed.). *PLoS ONE*, *10*(6), e0127409. https://doi.org/10.1371/journal.pone.0127409

SAE Internation. (2014). J3016: Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems - SAE International.

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2018). How to capitalize on a priori contrasts in linear (mixed) models: a tutorial.

Schoettle, B. (2017). Sensor fusion: a comparison of sensing capabilities of human drivers and highly automated vehicles. *University of Michigan*.

Schumann, F., Einhäuser-Treyer, W., Vockeroth, J., Bartl, K., Schneider, E., & König, P. (2008). Salient features in gaze-aligned recordings of human visual input during free exploration of natural environments. *Journal of vision*, *8*(14), 12.1–17. https://doi.org/10.1167/8.14.12

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76. https://doi.org/10.1016/j.tics.2005.12.009

Seppelt, B. D., & Lee, J. D. (2019). Keeping the driver in the loop: dynamic feedback to support appropriate use of imperfect vehicle control automation. *International Journal of Human-Computer Studies*, *125*, 66–80.

Shamay-Tsoory, S. G., & Mendelsohn, A. (2019). Real-life neuroscience: an ecological approach to brain and behavior research. *Perspectives on Psychological Science*, *14*(5), 841–859. https://doi.org/10.1177/1745691619856350

Shariff, A., Bonnefon, J.-F., & Rahwan, I. (2017). Psychological roadblocks to the adoption of self-driving vehicles. *Nature Human Behaviour*, *1*(10), 694–696. https://doi.org/10.1038/s41562-017-0202-6

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2018). *Afex: analysis of factorial experiments*.

Skulmowski, A., Bunge, A., Kaspar, K., & Pipa, G. (2014a). Forced-choice decision-making in modified trolley dilemma situations: a virtual reality and eye tracking study. *Frontiers in behavioral neuroscience*, *8*, 426.

Skulmowski, A., Bunge, A., Kaspar, K., & Pipa, G. (2014b). Forced-choice decision-making in modified trolley dilemma situations: a virtual reality and eye tracking study. *Frontiers in Behavioral Neuroscience*, *8*. https://doi.org/10.3389/fnbeh.2014.00426

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal*

Society B: Biological Sciences, *364*(1535), 3549–3557. https://doi.org/10.1098/rstb.2009.0138

Slater, M., Pérez Marcos, D., Ehrsson, H., & Sanchez-Vives, M. (2009). Inducing illusory ownership of a virtual body. *Frontiers in Neuroscience*, *3*, 29. https://doi.org/10.3389/neuro.01.029.2009

Slater, M., Spanlang, B., Sanchez-Vives, M. V., & Blanke, O. (2010). First person experience of body transfer in virtual reality. *Plos One*, *5*(5), e10564. https://doi.org/10.1371/journal.pone.0010564

Slater, M., Usoh, M., & Steed, A. (1994). Depth of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, *3*(2), 130–144. https://doi.org/10.1162/pres.1994.3.2.130

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, *6*(6), 603–616. https://doi.org/10.1162/pres.1997.6.6.603

Smith, B. (2019). Personality facets and ethics positions as directives for self-driving vehicles. *Technology in Society*, *57*, 115–124. https://doi.org/10.1016/j.techsoc.2018.12.006

Statistisches Bundesamt. (2018). *Verkehr: verkehrsunfälle* (tech. rep. Reihe 7).

Stephan, A. (2019). *Trust in highly automated driving* (Doctoral dissertation). Technische Universität Berlin.

Stereoscope. (2021, August 12). In *Wikipedia*.

Streppel, B., Pantförder, D., & Vogel-Heuser, B. (2018). Interaction in virtual environments - how to control the environment by using VR-glasses in the most immersive way. In J. Y. Chen & G. Fragomeni (Eds.), *Virtual, augmented and mixed reality: interaction, navigation, visualization, embodiment, and simulation* (pp. 183–201). Springer International Publishing. https://doi.org/10.1007/978-3-319-91581-4_14

Sullivan, B., Ludwig, C. J., Damen, D., Mayol-Cuevas, W., & Gilchrist, I. D. (2021). Look-ahead fixations during visuomotor behavior: evidence from assembling a camping tent. *Journal of vision*, *21*(3), 13–13.

Sullivan, B. T., Johnson, L., Rothkopf, C. A., Ballard, D., & Hayhoe, M. (2012). The role of uncertainty and reward on eye movements in a virtual driving task. *Journal of vision*, *12*(13), 19. https://doi.org/10.1167/12.13.19

Summala, H. (2000). Brake reaction times and driver behavior analysis. *Transportation Human Factors*, *2*(3), 217–226. https://doi.org/10.1207/sthf0203_2

Sütfeld, L. R., Gast, R., König, P., & Pipa, G. (2017). Using virtual reality to assess ethical decisions in road traffic scenarios: applicability of value-of-life-based models and influences of time pressure. *Frontiers in behavioral neuroscience*, *11*, 122.

Szikora, P., & Madarász, N. (2017). Self-driving cars — the human side. *2017 IEEE 14th International Scientific Conference on Informatics*, 383–387. https://doi.org/10.1109/informatics.2017.8327279

Tarr, M. J., & Warren, W. H. (2002). Virtual reality in behavioral neuroscience and beyond. *Nature Neuroscience*, *5*(11), 1089–1092. https://doi.org/10.1038/nn948

Tassy, S., Oullier, O., Mancini, J., & Wicker, B. (2013). Discrepancies between judgment and choice of action in moral dilemmas. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00250

Tauscher, J.-P., Schottky, F. W., Grogorick, S., Bittner, P. M., Mustafa, M., & Magnor, M. (2019). Immersive EEG: evaluating electroencephalography in virtual reality. *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 1794–1800. https://doi.org/10.1109/vr.2019.8797858

Tesla, M. (2020). *Autopilot* [Tesla Autopilot].

The Tesla Team. (2016). All tesla cars being produced now have full self-driving hardware. *Tesla Blog*.

The Tesla Team. (2019). Introducing a more seamless navigate on autopilot. *Tesla Blog*.

Thomas, B. A. (2018). A closer inspection of tesla's autopilot safety statistics. *Wired*.

Trappl, R. (2016). Ethical systems for self-driving cars: an introduction. *Applied Artificial Intelligence*, *30*(8), 745–747. https://doi.org/10.1080/08839514.2016.1229737

ultraleap. (2021). *Digital worlds that feel human | ultraleap*.

Unity Technologies. (2018). *Unity*. Unity Technologies. San Francisco, United States.

University of Houston. (2000). *Introduction to simulation and modeling: historical perspective*.

Van der Stigchel, S. (2020). An embodied account of visual working memory. *Visual cognition*, *28*(5-8), 414–419. https://doi.org/10.1080/13506285.2020.1742827

van Emmerik, M. L., de Vries, S. C., & Bos, J. E. (2011). Internal and external fields of view affect cybersickness. *Displays*, *32*(4), 169–174.

Venkatesh, V. (2000). Determinants of perceived ease of use: integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information Systems Research*, *11*(4), 342–365.

Virtual Reality Society. (2017). *History of virtual reality* [Virtual reality society].

Vive. (2021). *VIVE hand tracking SDK - developer resources* [Vive delelopers].

Voloh, B., Watson, M. R., König, S., & Womelsdorf, T. (2019). Mad saccade: statistically robust saccade threshold estimation via the median absolute deviation. *Journal of Eye Movement Research*, *12*(8).

VR design studio | FORUM8 | 3d VR & visual interactive simulation [Forum8]. (n.d.).

Ward, M. R., Lee, C., D'Ambrosio, L., F, J., & M. (2017). Coughlin,"Acceptance of automated driving across generations: the role of risk and benefit perception, knowledge, and trust. In *HCI 2017, lecture notes in ComputerScience, vol.10271* (pp. 254–266). Springer.

Warne, R. T. (2014). A primer on multivariate analysis of variance (MANOVA) for behavioral scientists. *Practical Assessment, Research & Evaluation*, *19*.

Warne, R. T., Lazo, M., Ramos, T., & Ritter, N. (2012). Statistical methods used in gifted education journals, 2006-2010. *Gifted Child Quarterly*, *56*(3), 134–149. https://doi.org/10.1177/0016986212444122

Wheatstone, C. (1838). XVIII. contributions to the physiology of vision. —part the first. on some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London*, *128*, 371–394. https://doi.org/10.1098/rstl.1838.0019

Wienrich, C., Schindler, K., Dollinger, N., Kock, S., & Traupe, O. (2018). Social presence and cooperation in large-scale multi-user virtual reality - the relevance of social interdependence for location-based environments. *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 207–214. https://doi.org/10.1109/vr.2018.8446575

Wiesing, M., Fink, G. R., & Weidner, R. (2020). Accuracy and precision of stimulus timing and reaction times with unreal engine and SteamVR. *Plos One*, *15*(4), e0231152. https://doi.org/10.1371/journal.pone.0231152

Wikimedia. (2010). *Charles_wheatstone-mirror_stereoscope_xixc.jpg (800×441)* [Wikimedia].

Wilkinson, G. N., & Rogers, C. E. (1973). Symbolic description of factorial models for analysis of variance. *Journal of the Royal Statistical Society. Series C, Applied statistics*, *22*(3), 392. https://doi.org/10.2307/2346786

Wilson, H., Theodorou, A., & Bryson, J. J. (2019). Slam the brakes: perceptions of moral decisions in driving dilemmas. *International Workshop in Artificial Intelligence Safety (AISafety), IJCAI, Macau*.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic bulletin & review*, *9*(4), 625–636. https://doi.org/10.3758/bf03196322

Wintersberger, P., Frison, A. K., & Riener, A. (2017). The experience of ethics: evaluation of self harm risks in automated vehicles. *IEEE Intelligent Vehicles Symposium*. https://doi.org/10.1109/ivs.2017.7995749

Wintersberger, P., Frison, A.-K., Riener, A., & Sawitzky, T. v. (2019). Fostering User Acceptance and Trust in Fully Automated Vehicles: Evaluating the Potential of Augmented Reality. *PRESENCE: Virtual and Augmented Reality*, *27*(1), 46–62. https://doi.org/10.1162/pres_a_00320

Wintersberger, P., Nicklas, H., Martlbauer, T., Hammer, S., & Riener, A. (2020). Explainable automation: personalized and adaptive uis to foster trust and understanding of driving automation systems. *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 252–261.

Wintersberger, P., & Riener, A. (2016). Trust in Technology as a Safety Aspect in Highly Automated Driving. *i-com*, *15*(3), 297–310. https://doi.org/10.1515/icom-2016-0034

Wohlschläger, A., Gattis, M., & Bekkering, H. (2003). Action generation and action perception in imitation: an instance of the ideomotor principle. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *358*(1431), 501–515. https://doi.org/10.1098/rstb.2002.1257

Yarbus, A. L. (2013). *Eye movements and vision*. Springer.

Yarbus, A. L. (1967). Eye movements during perception of complex objects. In *Eye movements and vision* (pp. 171–211). Springer.

Zaki, J., & Ochsner, K. (2009). The need for a cognitive neuroscience of naturalistic social cognition. *Annals of the New York Academy of Sciences*, *1167*, 16–30. https://doi.org/10.1111/j.1749-6632.2009.04601.x

Zeeb, K., Buchner, A., & Schrauf, M. (2015). What determines the take-over time? an integrated model approach of driver take-over after automated driving. *Accident Analysis & Prevention*, *78*, 212–221. https://doi.org/10.1016/j.aap.2015.02.023

Zhao, Y., Wang, X., Goubran, M., Whalen, T., & Petriu, E. M. (2013). Human emotion and cognition recognition from body language of the head using soft computing techniques. *Journal of Ambient Intelligence and Humanized Computing*, *4*(1), 121–140. https://doi.org/10.1007/s12652-012-0107-1

Zheng, J., Chan, K., & Gibson, I. (1998). Virtual reality. *IEEE Potentials*, *17*(2), 20–23. https://doi.org/10.1109/45.666641