Technische Universität Darmstadt

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Fachbereich Humanwissenschaften
Centre for Cognitive Science & Institut für Psychologie

zur Erlangung des Grades
Doctor rerum naturalium (Dr. rer. nat.)

Active vision as sequential decision-making under uncertainty

Dissertation von: Florian Kadner

Erstgutachter Prof. Constantin A. Rothkopf, Ph.D.
Centre for Cognitive Science & Institut für Psychologie
Technische Universität Darmstadt

Zweitgutachterin Prof. Mary M. Hayhoe, Ph.D.
Center for Perceptual Systems
The University of Texas at Austin

Darmstadt 2024

Florian Kadner
Active vision as sequential decision-making under uncertainty
Darmstadt, Technische Universität Darmstadt
Jahr der Veröffentlichung der Dissertation auf TUPrints: 2024
Tag der mündlichen Prüfung: 23.01.2024

Abstract

Interacting with our visual environment can be challenging due to its highly dynamic nature and richness in complex interrelationships. With the human visual system's constraint of having a narrow field of high resolution, we must actively shift our attention between different visual areas to acquire relevant visual information to accomplish our tasks. Extracting this task-relevant information from our environment can be challenging and further amplified by our world's inherently probabilistic nature. Sensory perception often presents ambiguities with varying results from identical measurements and vice versa. Similarly, the consequences of our actions are usually governed by uncertainty, which originates from several internal and external factors. Finally, the relevance of completing a particular task or even the definition of the task and its associated costs are highly variable across individuals. Thus, uncertainty is a fundamental factor at multiple stages while interacting with our visual environment. Sensory perception, decision-making, and actions are inseparably intertwined, and it is, therefore, all the more critical that we deal with the arising uncertainties and develop strategies to reduce them as far as possible. Computationally, this aligns with the concept of planning. In this thesis, we are investigating the active nature of visual planning as a probabilistic decision-making process under uncertainty. We designed various experimental paradigms to quantify sensory uncertainty, action variability, and the behavioral costs of human behavior in sequential visual tasks. For this purpose, we use the framework of Partially Observable Markov Decision Processes (POMDPs), which allow us to normatively model decision-making processes by incorporating different sources of uncertainty. Using three case studies, we demonstrate its use, advantages, and possibilities, starting with the most straightforward visual action - blinking. Even this simple action has to be planned since every blink briefly interrupts the visual information stream. We then move on to more complex visual actions such as saccades and gaze selection. First, we consider one-step ahead predictions in the context of free viewing and saliency models before moving on to a complex example of a gaze-contingent paradigm task where, in addition to observations, rewards are dynamic and uncertain. Last, we consider two other studies more detached from the experimental environment and devoted to more natural stimuli. We investigate how humans navigate mazes and their associated planning strategies of eye movements to find the solution. Also, we designed a reading experiment including an adaptive font system that maximizes the subjects' individual reading speed and thus reduces the underlying internal behavioral costs. Our results conclude that human visual behavior should be seen as an active sequential decision process under uncertainty where POMDPs can provide a powerful tool for modeling.

Zusammenfassung

Die Interaktion mit unserer visuellen Umgebung kann aufgrund ihrer hochdynamischen Natur und der Fülle an komplexen Zusammenhängen eine Herausforderung darstellen. Da das menschliche visuelle System nur über eine schmale Region hoher Auflösung verfügt, müssen wir unsere Aufmerksamkeit aktiv zwischen verschiedenen visuellen Bereichen hin- und herbewegen. Diese Schwierigkeit, visuelle Informationen zur Bewältigung unserer Aufgaben aus unserer Umgebung zu extrahieren, wird durch die inhärent probabilistische Natur unserer Welt noch verstärkt, da Wahrnehmung oft nicht eindeutig ist. Auch die Folgen unserer Handlungen sind in der Regel mit Unsicherheit behaftet und selbst die Relevanz der Erledigung einer bestimmten Aufgabe oder sogar deren Definition und die damit verbundenen Kosten sind von Person zu Person sehr unterschiedlich. Unsicherheit ist also ein grundlegender Faktor in der Interaktion mit unserer visuellen Umgebung. Sinneswahrnehmung, Entscheidungsfindung und Handeln sind untrennbar miteinander verbunden, und deshalb ist es umso wichtiger, dass wir uns mit den entstehenden Unsicherheiten auseinandersetzen und Strategien dagegen entwickeln. Computational gesehen entspricht dies dem Konzept der Planung. In dieser Arbeit untersuchen wir visuelle Planung als einen probabilistischen Entscheidungsprozess unter Unsicherheit. Wir haben verschiedene experimentelle Paradigmen entwickelt, um die sensorische Unsicherheit, die Handlungsvariabilität und die Verhaltenskosten des menschlichen Verhaltens in sequenziellen visuellen Aufgaben zu quantifizieren. Zu diesem Zweck verwenden wir Partially Observable Markov Decision Processes (POMDPs), die es uns ermöglichen, Entscheidungsprozesse normativ zu modellieren und verschiedene Quellen der Unsicherheit einbeziehen. Dies demonstrieren wir anhand dreier Studien beginnend mit der einfachsten visuellen Handlung - dem Blinzeln. Selbst diese einfache Handlung muss geplant werden, da jedes Blinzeln den visuellen Informationsstrom kurz unterbricht. Anschließend gehen wir zu komplexeren visuellen Handlungen der Sakkaden über. Zunächst betrachten wir Ein-Schritt-Vorhersagen im Zusammenhang mit freiem Sehen und Salienzmodellen, bevor wir zu einem komplexen Beispiel übergehen, bei dem zusätzlich zu den Beobachtungen auch die Belohnungen dynamisch und unsicher sind. Zum Schluss betrachten wir zwei weitere Studien in natürlicheren Umgebungen. Wir untersuchen das Planungsverhalten von Menschen beim Lösen von Labyrinthen. Außerdem haben wir ein Leseexperiment mit einem adaptiven Schriftsystem entworfen, das die individuelle Lesegeschwindigkeit der Versuchspersonen maximiert. Unsere Ergebnisse lassen den Schluss zu, dass das menschliche Sehverhalten als aktiver sequentieller Entscheidungsprozess unter Unsicherheit betrachtet werden sollte, für dessen Modellierung POMDPs ein leistungsfähiges Werkzeug darstellen können.

Acknowledgement

First and foremost, I would like to thank my supervisor Constantin Rothkopf. Thank you for the privilege and opportunity to pursue my PhD in your lab. You gave me the freedom to follow my interests and always supported me, academically and personally. But most importantly, you always believed in me. Thank you for everything.

Thanks to Mary Hayhoe for agreeing to be the second adviser for this thesis and Frank Jäkel and Loes van Dam to be part of my thesis committee.

Special thanks to my student assistants - Dennis, Yannik, Thi, Tobias, Hannah - as well as my thesis students - Christian, Hoa, Marius, Sebastian, Tabea, Thabo, Wassim- it has been a great honor working with you. Thank you for your support, your outstanding work, and your trust in me.

Thanks to our secretary Inge Galinski, who supported me with all formal questions and hurdles. Furthermore, thanks to all the great colleagues over the years. In particular, I want to thank David, Fabian, Inga, Niteesh, Tobias and Vildan. You have not only become colleagues, but friends. You have made PhD life much more enjoyable and I am very grateful for all the fun, your open ears, and our activities outside of university.

An extraordinary thanks goes to Nils and Susi. I cannot put my gratitude for you and our friendship into words. We laughed together, celebrated the highs, wept and overcame the lows, and always believed in each other. Without you and your support, this whole project would not have been possible. I am eternally grateful to you both for everything.

I would like to thank all of my friends for their support outside of work. Special thanks to Anna, Dominik, Florian, Johanna and Jonas for your friendship, for always standing by my side and sharing all the ups and downs with me.

Finally, I want to thank my family. Especially my grandparents, my sister and my brother-in-law who are always there for me and support me in my goals. And of course, my parents, to whom I owe everything. You made this entire journey possible, always supported me unconditionally, and backed me up in every situation in life.

Research in this thesis was supported by German Research Foundation (DFG, grant: RO 4337/3-1; 409170253) and the Excellence Program of the Hessian Ministry of Higher Education, Science, Research and Art through the project WhiteBox.

Contents

1 Introduction

> *As humans we have the compelling experience of living in a three-dimensional visual movie.*
>
> **— Dana Ballard, 1991**

Vision is by far the most studied sensory perception and is considered by many people as the most important of their senses (Hutmacher, 2019). This is not surprising when even Plato in his days called the ability to see *divine* (Jütte, 2005) and an intensive debate has been going on for many years as to whether seeing is the dominant sense (see, e.g., Gottlieb, 1971; Hirst et al., 2018; Posner et al., 1976). The problem of where to direct our gaze was already significant for the first humans to survive, e.g., locating food sources, facilitating navigation, and detecting potential threats such as predators (Gibson, 2014; Heesy, 2009). Since we live in a fast-moving and dynamic environment, this difficulty is still present and we have to assimilate and process much visual input to achieve our goals. These might be complex situations such as driving a car (Land and Lee, 1994; Sullivan et al., 2012), navigating through crowded places (Gibson, 2014) or playing team sports (Gou et al., 2022; Kredel et al., 2017). Nevertheless, even apparently simple situations like preparing tea (Land et al., 1999) or making a sandwich (Hayhoe and Ballard, 2005; Hayhoe and Rothkopf, 2011; Land, 2009) require us to process the visual data and finally convert it into action sequences. Vision can be understood as active sequential decision making (Hayhoe and Ballard, 2005), which allows us to flexibly adapt and change our behavior concerning the visual environment and internal goals (Hoppe and Rothkopf, 2019). However, if the ability to see is of such immense importance to us and has been the most studied for years, shouldn't vision be thoroughly researched and explainable?

1.1 Individual differences in visual processing

In the last decades, vision research has yielded more and more empirical results and phenomena for describing the visual process. As an example, two different effects were found relating to the question where people look in static visual scenes: Firstly, the ones that originated from the stimuli themselves (so-called *bottom up saliency*, Itti and Koch, 2000; Torralba et al., 2006) and secondly the ones that come from the behavioral goal itself (so-called *top-down cognitive control*, Buswell, 1935; Yarbus, 1967). However, many other descriptive models of eye movement behavior were investigated in numerous studies and given their names. To name just a few, for instance: *inhibition of return* (Posner, Cohen, et al., 1984), *center bias* (Buswell, 1935; Tseng et al., 2009), *change blindness* (Simons and Levin, 1997), *binocular rivalry* (Breese, 1899; Dam and Ee, 2006), *illusory contours* (Heydt et al., 1984) and *color constancy* (Foster, 2011). In addition, many sub-fields of vision have been formed that look at different aspects, such as *face perception* (Bruce and Young, 1986; Ryali et al., 2020), *motion perception* (Adelson and Bergen, 1985; Wertheimer, 1912), *visual search* (Najemnik and Geisler, 2008; Treisman and Gelade, 1980; Wolfe, 1994) and *scene perception* (Bar, 2004; Biederman, 2017; Henderson and Hollingworth, 1999). This list is far from exhaustive and can only represent a tiny fraction of what the vision literature now encompasses. However, despite this wealth of experimental results, phenomena, and individual theories, a unified framework to describe the visual process with all internal and external processes together has not yet been fully developed. Why is that?

One of the main reasons behind this question is that visual processes are highly variable between different people (and their individual internal goals) and are driven by both sensory and cognitive factors (Hayhoe and Ballard, 2005; Kowler, 2011; Land and Tatler, 2009; Tatler et al., 2011). We have listed some phenomena in the previous section that can be observed repeatedly in human subjects. These effects often occur over the entire experimental sample, but at the individual level, it is difficult to fully explain behavior and make accurate predictions. To demonstrate this, we reconsider the example of making a sandwich, illustrated in Figure 1.1. This task may seem simple, but almost everyone has done it before. However, it shows us in what details our visual behavior can differ. Moreover, the interconnected nature between perception and action becomes visible. Much past vision research has investigated visual perception as independent of actions; however, using this example, we show how inseparable these two are from each other and that we should study them simultaneously.

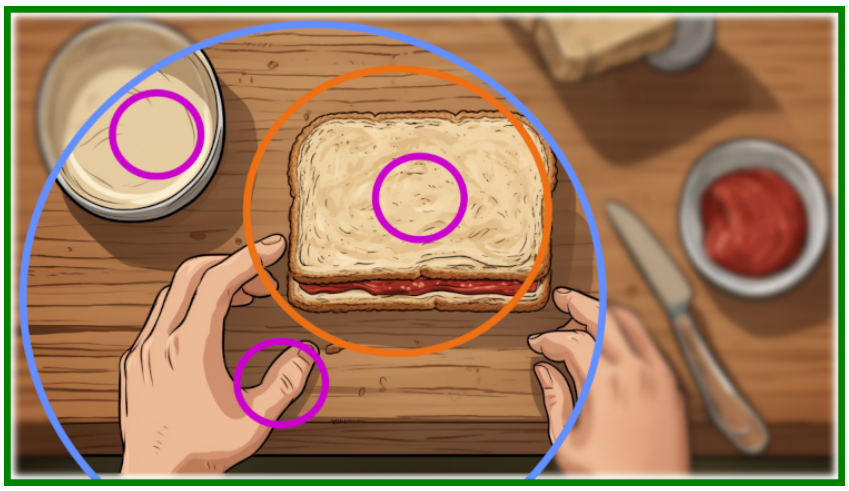


Fig. 1.1: Illustration of sensory uncertainty, action variability, behavioral cost and model uncertainty in the natural task of making a peanut butter and jelly sandwich.

We need to use visual input via our eyes to control our motor system to get to our goal (here, smearing the bread with peanut butter and jam). Even this simple action can be highly variable for different people due to multiple factors:

1. **Sensory uncertainty**: Our world is full of uncertainty, and even if perception seems deterministic to us, probabilistic noise is always present in sensory measurements of all kinds either in a thermodynamic (e.g., smell and gustation) or quantum mechanical way (e.g., seeing) (Faisal et al., 2008). For example, when we see 3D objects, they must be mapped internally to 2D images, and our brain must be able to perform this probabilistically uncertain inference so that we can plan our actions accordingly (Knill and Pouget, 2004). On the other hand, if we could represent our external environment internally in detail, there would be no visual uncertainty (O'Regan and Noë, 2001).

   Seeing is a sensory perception like tasting, hearing, smelling, and touching. All have in common that we have particular sensory organs to receive and internally represent the influences of our environment. Our eyes are responsible for seeing, and the process is, first of all, nothing else than light falling on them. It passes through the cornea and lens, concentrating it onto the retina, a light-sensitive layer of photoreceptor cells at the back of the eye (Dowling, 1987; Palmer, 1999). There, the electrical signals generated by the photoreceptors are first processed locally and integrated into visual information (Wässle, 2004), which is then sent via the optic nerve to different brain regions such as the lateral geniculate nucleus (LGN) and the primary visual cortex (V1) for further processing and interpretation (Hubel and Wiesel,

1962). The ability to see can be limited through uncertainty in the sensory measurement by internal (e.g., noise in the measurement and processing of the retina) and external circumstances (e.g., the incidence of light, individual physiological conditions like myopia or hyperopia) (Geisler, 2011; Pelli and Farell, 1999). The photoreceptors are much more concentrated in a small region on the retina, leading to higher resolution and sensitivity in parts of our visual environment in the so-called foveated field of view (Curcio et al., 1990). This is indicated in Figure 1.1 by the blue area. Outside this circle - in the peripheral vision - the resolution is lower (Reise and Waller, 2009), we can see the jar with the strawberry jam worse than the peanut butter. This can bias our sensory measurement about the position of objects and is inevitably related to our actions; for example, do we reach for the strawberry jam, or do we direct our gaze to it first to make our measurement of its position less uncertain?

2. **Action variability**: Humans are not machines, and our motor actions are highly variable (Churchland et al., 2006). When we make an eye movement, the exact trajectory cannot be predicted accurately due to noise in the neural control signals (Harris and Wolpert, 1998). This noise can occur at various points in the processing of signals and execution of actions, and our system attempts to counteract it through various strategies (Faisal et al., 2008). Consider a random fixation sequence in Figure 1.1, represented by the magenta-colored circles. Even if everyone fixates on the same objects, first the peanut butter, then the bread, and finally the thumb, we will not observe a trajectory twice due to the variability in saccade execution. This variance can also arise from factors such as fatigue, cognitive load, age, or medical conditions (Leigh and Zee, 2015). In addition to the variability that arises at the cellular level due to noise, variability can also be observed that is related to behavioral costs (which we highlight in the next section).

3. **Behavioral costs**: When accomplishing tasks, we must weigh the benefits against the costs of performing the necessary actions (e.g., time, energy, and cognitive resources) (Trommershäuser et al., 2008). We can find different behavioral patterns within the same task depending on how important the task was to us (Hoppe et al., 2018). Behavioral costs refer to individuals' trade-offs when deciding which actions to take and which objects to attend to during the task. Thus, they are critical in optimizing their strategy to achieve their goals while keeping their costs low (Hayhoe and Ballard, 2005; Hoppe et al., 2018). In our example in Figure 1.1, making a sandwich involves a series of actions that require time, effort, and attention. How important is it for us to finish the

task? Maybe we are starving and want to finish the sandwich quickly. Or am I not making the sandwich for me but for someone else? Behavioral costs can vary between individuals depending on their priorities, goals, and perceived risks during the task. They also interact with action variability since our motor system tries to reduce the noise in task-relevant dimensions (Todorov and Jordan, 2002). Moreover, even if my only priority would be to make the sandwich as fast as possible, since action variability scales with the amplitude action (Harris and Wolpert, 1998), it might be better to do it slower.

4. **Internal model uncertainty**: Even if we had perfect sensory perception, we still have to represent these measurements internally. This representation can be noisy and we refer to it as internal model uncertainty. This uncertainty can be about the internal representation of the external world itself or the outcome of our actions. The difference between an expected and actual outcome is known as prediction error and signals the brain to adjust the model (Clark, 2013). Furthermore, the robustness of our internal models can also be influenced by neuroplasticity, the ability of our brain to reorganize and adapt based on experience (Holtmaat and Svoboda, 2009). As we navigate the world, our brain continually updates its internal representations based on new experiences and information through Bayesian inference (Körding and Wolpert, 2004). The certainty of these updates can vary, leading to uncertain predictions and conclusions (Griffiths and Tenenbaum, 2009). Additionally, internal model uncertainty actively influences our decision-making process. For example, if we are more uncertain about our internal representation, we are less confident and show longer reaction times (Yeung and Summerfield, 2012). In our example in Figure 1.1 the model uncertainty is pictured as green frame since it regards the whole representation of the event.

To conclude: Like our example, many everyday tasks are sequential. There are different sub-aspects that we have to master in order to arrive at the overall solution. This also requires our eye movements to be sequential in nature in order to direct our attention specifically and actively to the relevant parts of the visual environment. In the process, our actions are subject to uncertainties, both of a sensory and motor nature. In addition, individual preferences and goals should be taken into account as well. Vision is therefore highly individual and should be modeled as an **active sequential decision process under uncertainty** (Findlay and Gilchrist, 2003; Gottlieb, 2012; Hoppe and Rothkopf, 2019) including **sensory uncertainty, action variability, behavioral costs** and **internal model uncertainty** (Hoppe and Rothkopf, 2019; Kessler et al., 2022; Neupärtl and Rothkopf, 2021; Straub and Rothkopf, 2022).

1.2 From passive to active vision

The term *active vision* first originated in 1988 and refers to the ability of a system (whether alive, such as humans or animals or artificial, such as robots) to actively control the direction and focus of its gaze in order to gather visual information about the surrounding environment (Aloimonos et al., 1988; Bajcsy, 1988). This is in contrast to the framework of passive vision, which until then was the prevailing vision theory. There, the system has no active control over its gaze and must solely rely on the visual information available at any given time through processing the static input images (Findlay and Gilchrist, 2003; Marr, 2010).

In the following 90s, the concept of active vision became increasingly popular, and arguments against the framework of passive vision were found. If the human visual system were indeed passive, then our brain would have to process the vast amount of visual input in parallel, which it is biologically incapable of doing with the number of neurons it has (Lennie, 2003). In addition, after years of investigating low-level vision, high-level features became increasingly popular. The concept of passive vision as static image processing reached its limits of explainability, as it could not close the gap and explain the relation between low and high-level features (Nakayama, 1990). These problems were solved and could now be explained by the theory of active vision, where the directed and foveated gaze serves as a solution to computationally efficiently obtain a set of visual inputs that can be processed (Ballard, 1991; Ballard et al., 1997). Vision was understood from there as sampling relevant parts in our visual environment, leading to more efficient processing and reduced cognitive load (Findlay and Gilchrist, 2003; Land, 1995; Land and Nilsson, 2012; Ma et al., 2021).

Active vision enables humans to engage with their environment, gathering relevant information to guide their actions (Aloimonos et al., 1988; Ballard, 1991). Therefore, it is often used in everyday life and provides a natural model for incorporating sensory uncertainty. It explains the mechanism to balance it through active exploration (Cai et al., 1997) and is suitable as a natural modeling approach, including action variability since there is experimental evidence that our visual apparatus can make online corrections during a motor action (Körding and Wolpert, 2006; Todorov, 2004). Eye movements are not just passive reactions to visual stimuli; we need them to actively process visual information to guide our actions and plan future eye movements (Land et al., 1999). We want to give some examples of everyday problems and tasks in which we actively use our gaze to show the importance of this property:

Navigating in crowded spaces Especially in dynamic situations with many other people, studying our visual environment is essential to avoid obstacles and maintain a safe distance from others (Gibson, 2014). With the help of peripheral vision and saccadic eye movements, we must maintain situational awareness to reach our goal (Land and Tatler, 2009). Thus, we need active anticipation of our current planning behavior and the path in real-time (Ognibene and Demiris, 2013). Evidence for those gaze strategies was found, for instance, by Rothkopf et al., 2016.

Driving a car Driving requires continuous monitoring and processing of various visual cues inside and outside the car (Land and Lee, 1994; Sullivan et al., 2012). Since the advent of interactive technologies such as navigation and entertainment systems, there are potential visual distractors that require a short period of time looking away from the road and trade-off the visual locations (Jokinen et al., 2021; Noy et al., 2004). Active vision is applied in anticipating potential hazards and making decisions under uncertainty (Lappi, 2014). Empirical evidence was collected in the form of various eye movement patterns in drivers during different driving scenarios, for instance, by (Underwood et al., 2003).

Searching for items Searching for objects in cluttered places is an excellent example of why we need to use our gaze actively. It is not enough to perceive the whole scene as a single image; instead, we have to search specifically for the item, scanning the environment based on the history of our previous attempts. This task is highly variable and can result in different strategies (Wolfe, 1998), the active vision framework is suitable for capturing these individual differences (Kieras and Hornof, 2014).

Participating in sports In sports, players or objects often need to be tracked or their trajectory predicted (Farrow and Abernethy, 2003). Eye movement patterns vary significantly between people, given their expertise and abilities (Mann et al., 2007). In the dynamic situation of a sports game, people must be able to synchronize and constantly adapt their sensory percepts and actions.

Reading and interpreting visual information While reading, we use rapid eye movements to move through the text (Rayner, 1998). The positions we fixate on are not random but based on various factors. Sensory uncertainty may play a role in this, depending on how well we can perceive the font or the text as a whole (Arditi and Cho, 2005; Blommaert and Timmers, 1987; Legge and Bigelow, 2011;

McLean, 1965). Nevertheless, the different variances in the motor execution of eye movements and the behavioral costs based on the goals of the reading process influence the movement strategies (Rayner, 1998).

Active vision has also made significant technological strides, especially in robotics and artificial intelligence (compare, e.g. Tsotsos, 2011). By employing algorithms that mimic the human eye's ability to focus on certain parts of a visual scene selectively (e.g. Itti et al., 1998; Sun, Xiaoshuai et al., 2012; Wang et al., 2011; Zanca et al., 2019), robots can now interact with their environment more effectively and with greater computational efficiency. These principles have been utilized in autonomous vehicles (Chen et al., 2015), where the system must dynamically adapt to changing road conditions, and in medical imaging (Litjens et al., 2017), where precise focus on relevant areas can aid diagnosis and treatment planning. The application of active vision in technology underscores its potential beyond human cognition, bridging the gap between biological inspiration and practical utility (e.g. Pfeifer and Bongard, 2006).

In conclusion, active vision represents a profound shift in our understanding of visual perception, bridging the gap between low-level sensory input and high-level cognitive processing. Moving beyond the constraints of passive vision offers a more nuanced and dynamic model of how humans, animals, and even artificial systems interact with their surroundings. Active vision is crucial for humans, providing an interface for our motor actions. It enables us to process the almost infinite masses of visual input in a targeted manner and thus make cognitive processing manageable. We can direct our gaze to the places in the visual environment where we get the information we need to achieve our goals. At the same time, it is a dynamic mechanism that we can use adaptively through feedback after our actions to adjust for the new information we need. Vision can thus be understood as an individual and active process that follows the same principles but differs for each person. As research in this field continues to evolve, the insights gained from studying active vision deepen our comprehension of cognitive functions and pave the way for technological innovations and practical applications that enhance our interactions with the world.

1.3 Overview of the thesis

This thesis will focus on modeling vision as an active decision-making process under uncertainty. In order to do this, we introduced the active vision framework. We looked at the developments of how vision went from being a passive framework in the late 80s/early 90s to being viewed more and more as an active process. In addition, we also referenced some examples where active vision occurs and what consequences it has for us. Chapter 2 is dedicated to the theory of computationally modeling sequential decisions under uncertainty. Through the simple model of direct inference, linear dynamical systems, and statistical decision theory, we approach the framework of Partially Observable Markov Decision Processes (POMDP), which allow us to normatively model decision-making processes by incorporating different sources of uncertainty. Using these modeling frameworks, we look at different visual action levels. We have designed rich experiments, analyzed their empirical results, and modeled them normatively to gain insight into the subjects' internal uncertainties and costs. Chapter 3 is then devoted to the most rudimentary visual action for studying sequential decision-making - blinking. We develop a probabilistic planning model combining POMDPs and non-homogeneous counting processes, allowing us to effectively incorporate temporal event rates' influence in the blink behavior analysis. We then turn to more complex eye movement statistics and first develop an algorithm to improve saliency models' predictions of the subsequent fixation by including humans' intrinsic cost of gaze shifts (Chapter 4). After that, we extend the horizon from one-step predictions to investigate the planning behavior for a sequence of fixations (Chapter 5). We used a gaze-contingent experiment where human participants needed to switch between a reward collection task involving different dynamic, uncertain reward rates and a visual detection task involving dynamic, uncertain observations. We demonstrate the advantages of the suggested POMDP framework in all these cases. We model the inter- and intraindividual differences of the subjects and infer their cost and uncertainty parameters. Finally, we examine two additional studies less connected to the experimental environment, focusing instead on more natural stimuli. We explore humans' techniques to navigate mazes, focusing on the planning strategies linked to eye movements that lead to solutions (Chapter 6). Finally, we demonstrate our visual system's nature and preferences using the example of an adaptive font system for faster processing and, thus, faster reading speed in Chapter 7.

1.4 Contributions

The chapters in this thesis are based on the following articles/publications and may contain previously published content:

Chapter 3: Blinking as probabilistic planning under uncertainty

This work is currently under preparation for submission:

- Kadner, F. & Rothkopf, C. A. (2023). A probabilistic planning model to predict human blinking behavior given temporal event statistics.

Chapter 4: Improving one-step ahead predictions of subsequent fixations using individual cost maps (Kadner et al., 2023a)

This work was published in:

- Kadner, F., Thomas, T., Hoppe, D., & Rothkopf, C. A. (2023). Improving saliency models' predictions of the next fixation with humans' intrinsic cost of gaze shifts. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 2104-2114).

  https://doi.org/10.1109/WACV56688.2023.00214

[1] Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Chapter 5: Humans trade-off reward collection and information gathering according to probabilistic planning (Kadner et al., 2022)

This work is currently under preparation for submission and was presented with a selected talk at the annual meeting of the Vision Science Society 2022:

- Kadner, F., Wilke, T. A., Vo, T. D., Hoppe, D., & Rothkopf, C. A. (2023). Humans adaptively balance gathering information and rewards in a dynamic uncertain environment.
- Kadner, F., Wilke, T. A., Vo, T. D., Hoppe, D., & Rothkopf, C. A. (2022). Trade-off between uncertainty reduction and reward collection reveals intrinsic cost of gaze switches. Journal of Vision, 22(14), 3400-3400.

  https://doi.org/10.1167/jov.22.14.3400

Chapter 6: Human eye movements and their planning strategies in maze navigation (Kadner et al., 2023b)

This work was published in:

- Kadner, F., Willkomm, H., Ibs, I., & Rothkopf, C. (2023). Finding your Way Out: Planning Strategies in Human Maze-Solving Behavior. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 45, No. 45).

  https://escholarship.org/uc/item/94t1t8kw

Chapter 7: Increasing individuals' reading speed with an adaptive font model and Bayesian optimization (Kadner et al., 2021)

This work was published in:

- Kadner, F., Keller, Y., & Rothkopf, C. (2021, May). Adaptifont: Increasing individuals' reading speed with a generative font model and bayesian optimization. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-11)*.

  https://doi.org/10.1145/3411764.3445140

2

Computational modeling of decision-making under uncertainty

> *We don't see things as they are, we see them as we are.*
>
> **— Anaïs Nin, 1961**

Our world is full of uncertainty in many different forms. There may be uncertainty about the state of the environment, the outcome of our actions, or the interaction of different agents (Kochenderfer et al., 2022). Earlier in this thesis, we explained why we should include sensory uncertainty, action variability, behavioral costs and model uncertainty in vision frameworks. We also introduced the theory of active vision, which can be understood as a sequential decision-making process. In this chapter, we will now turn to the computational modeling of these processes.

Decision-making under uncertainty refers to making choices in situations where the outcomes are not confident or without all the needed information for the decision-maker. A sophisticated framework that can implement these problems is the Partially Observable Markov Decision Process (POMDPs). Here, we will derive the idea of POMDPs piecemeal by first looking at simpler probabilistic models of perception and action, namely *direct inference*, *sequential inference* and *single decisions* and *Markov Decision Processes*. To better understand these models, we briefly overview probabilities and Bayesian networks at the beginning.

All the methods presented here are about deducing actual states from observations of our world. Each section begins with a visualization of the respective method. On the left-hand side is the respective Bayesian network of the relationship, and the right-hand side shows the interaction between humans and the world.

2.1 Basic facts about probability

We do not intend to give a complete overview of probability theory in this short section. Instead, we define the basic facts about probabilities that we need to understand in the later chapters (for a detailed introduction, see, for example, Bishop and Nasrabadi, 2006; Ma et al., 2023; Murphy, 2022).

A random variable X is a function (either discrete or continuous) that maps the outcomes of a random process to numerical values (specifically their associated probability). The probability that X takes a specific value x_i is denoted as $p(x_i)$. For instance, if we take the example of a six-sided fair die, x_i can represent the probability of rolling a 6, thus $p(x_i) = 1/6$. We are often interested in more than just a single outcome of a random process, so we define two other important concepts:

- Joint probability $p(x_i, x_j)$: the likelihood of both x_i and x_j occurring together.
- Conditional probability $p(x_i|x_j)$: the likelihood that x_i is happening, given that x_j has already occured.

With these concepts we can define some fundamental rules:

$$p(x_i, x_j) = p(x_i)p(x_j|x_i) = p(x_j)p(x_i|x_j) \quad \text{(Chain rule)}$$

$$(x_i \perp\!\!\!\perp x_j) := p(x_i, x_j) = p(x_i)p(x_j) \quad \text{(Independence I)}$$

$$(x_i \perp\!\!\!\perp x_j) := p(x_i|x_j) = p(x_i) \wedge p(x_j|x_i) = p(x_j) \quad \text{(Independence II)}$$

$$\text{[disc.]}\ \ p(x_i) = \sum_{x_j} p(x_i, x_j)\ \ \text{[cont.]}\ \ p(x_i) = \int_{x_j} p(x_i, x_j) \quad \text{(Sum rule)}$$

These are all the basic facts about probabilities we will need. With the help of this knowledge, we can derive one of the most essential and fundamental theorems of statistics - Bayes' theorem (Bayes, 1763). Bayesian statistics plays a crucial role within cognitive science and especially in the modeling of perception and action, so the following connection is also fundamental:

$$p(x_i|x_j) = \frac{p(x_i, x_j)}{p(x_j)} = \frac{\frac{p(x_i, x_j)}{p(x_i)} p(x_i)}{p(x_j)} = \frac{p(x_j|x_i)p(x_i)}{p(x_j)} \quad \text{(Bayes' Theorem)}$$

2.2 Bayesian networks

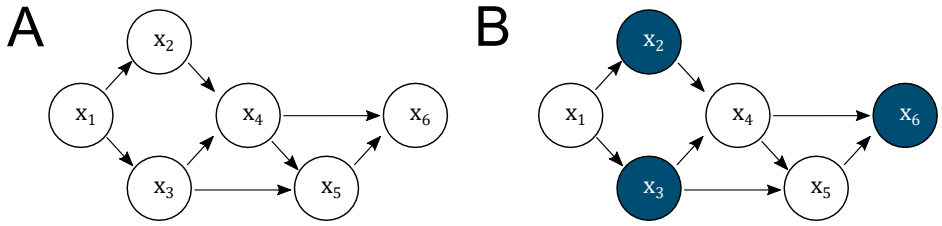


Fig. 2.1: Examples of a Bayesian network. (A) Every node is unobserved. (B) The nodes x_2,x_3 and x_6 are observed.

A Bayesian network is a set of nodes and their conditional dependencies on each other represented by a directed acyclic graph. An example for such a graph can be seen in Figure 2.1A. Nodes are random variables that can be either latent or observed (often indicated through shading, see 2.1B). Edges represent dependencies between these variables. The structure of a Bayesian network allows for the factorization of the joint probability distribution of its variables. For a set of variables $\{x_1, \ldots, x_n\}$, the joint distribution can be expressed as:

$$p(x_1, \ldots, x_n) = \sum_{i=1}^{n} p(x_i|\text{Parents}(x_i))$$

This factorization is a direct result of the local conditional independence assumptions in the network. For Figure 2.1A this leads us to a joint probability of:

$$p(x_1, \ldots, x_n) = p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_2, x_3)p(x_5|x_3, x_4)p(x_6|x_4, x_5)$$

One of the key concepts in Bayesian networks is that of conditional independence. If a node is conditionally independent of its non-descendants, given its parents, it does not provide any additional information about them once you observe them. Similar two the independence statement from the previous chapter, we can define conditional independence. If two random variables x_i and x_j are independent given a third variable x_k we write $(x_i \perp\!\!\!\perp x_j|x_k)$. In Bayesian networks d-seperation (dependence-seperation) is a useful criterion to check this assumption. Indeed, the observation of variables has a direct influence on the dependency relations in the graphical model. Two nodes x_i and x_j, are said to be d-separated by a third note x_k if every path between x_i and x_j is blocked by x_k. If x_i and x_j are d-separated by x_k, then $(x_i \perp\!\!\!\perp x_j|x_k)$. We determine how conditional independence changes using three potential cases for path blocking:

1. Causal Chain (Head-to-Tail connection)
    - in Figure 2.1A: $(x_1 \not\perp\!\!\!\perp x_4 | x_3)$
    - in Figure 2.1B: $(x_1 \perp\!\!\!\perp x_4 | x_3)$
2. Common Cause (Tail-to-Tail connection)
    - in Figure 2.1A: $(x_4 \not\perp\!\!\!\perp x_5 | x_3)$
    - in Figure 2.1B: $(x_4 \perp\!\!\!\perp x_5 | x_3)$
3. V-structure (Head-to-Head connection)
    - in Figure 2.1A: $(x_4 \perp\!\!\!\perp x_5 | x_6)$
    - in Figure 2.1B: $(x_4 \not\perp\!\!\!\perp x_5 | x_6)$

In what follows, we now consider models for modeling decision-making under uncertainty. First, however, let us take a look at Figure 2.2 to clarify why such modeling is needed. If we as researchers design an experiment, we know its dynamics and processes, i.e., the states s it can pass through. The subject goes through this experiment and takes care with actions a to get a reward r. Since it does not know s, it must make observations o. In doing so, the subject has to think about what the state probably looks like right now and which action it should best choose. As researchers, on the other hand, we are interested in inferring these latent individual costs Φ and beliefs Ψ (compare Section 1.1). Therefore, modeling is necessary to break open this black box and infer the individual properties to understand and interpret them. We are now developing piecemeal models that are capable of doing this. It started with simple observations, over sequential observations and single actions, to an all-encompassing framework.

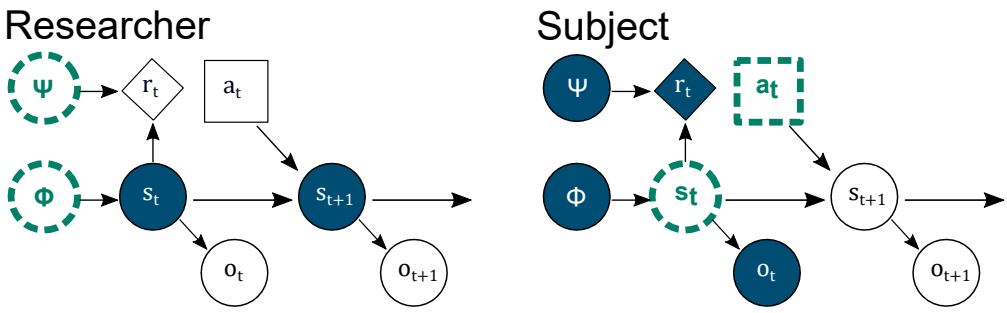


Fig. 2.2: Model from the perspective of the researcher and the subject respectively.

2.3 Direct inference

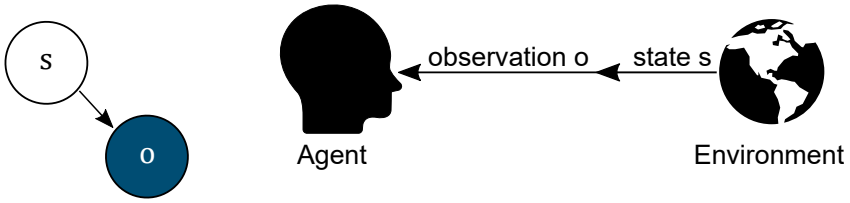


Fig. 2.3: The direct inference model.

Our perception is subject to uncertainty; each one of us perceives things in his or her unique way. To infer the true state s of real-world objects (e.g., positions or velocities), we need to combine noisy observation o (e.g., visual or auditory cues) from sensory modalities to form a coherent and robust percept (see Figure 2.3). We have to use probabilistic models to account for these sensory uncertainties to represent the relationship between actual states and observations. For the simplest case, this is the direct inference model in which we can model sensory uncertainty. However, since it does not include actions, we cannot yet add behavioral costs or action variability here. Mathematically, we express this model as a probability $p(s|o)$ using Bayes' theorem:

$$p(s|o) = \frac{p(o|s) \cdot p(s)}{p(o)} = \frac{p(o|s) \cdot p(s)}{\sum_s p(o|s) \cdot p(s)}$$

This equation contains the probability $p(s)$, which is called prior, representing the initial belief before any observations were made. The conditional probability $p(o|s)$, i.e., the probability to make the observation o, given the actual state s is called likelihood, and $p(o)$ is the marginal probability, which can be calculated as the sum of the product of the likelihood and the prior over all possible states. Meanwhile, with probabilistic programming languages, simple ways exist to approximately solve the underlying Bayesian inference (Bingham et al., 2019; Carpenter et al., 2017; Ge et al., 2018).

Let us look at a two-dimensional example to visualize the model. The aim is to estimate the position of an object in the horizontal and vertical directions, i.e., its x- and y-coordinate. For this, we assume an observation to get the observed position $(\hat{x},\hat{y})$ of the object. The prior is a bivariate Gaussian distribution with a specified mean and covariance. The likelihood is also Gaussian, representing the probability of observing a particular location given the actual location and observation noise in

both the vertical and horizontal directions. The resulting posterior distribution and the interaction of prior and likelihood can be seen in Figure 2.4.

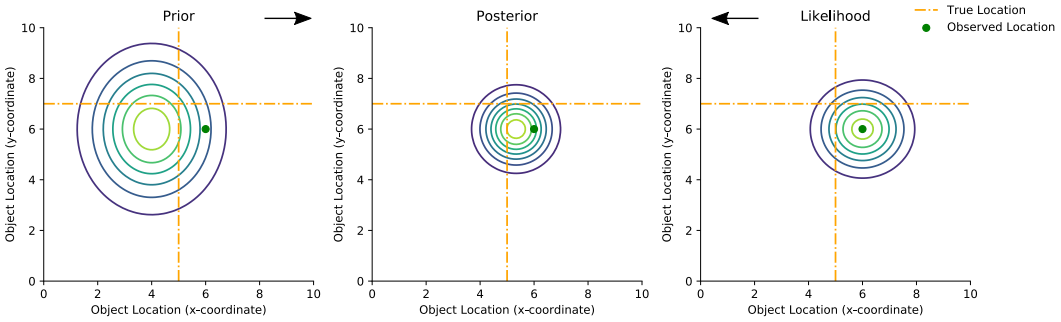


Fig. 2.4: Example of position perception using the direct inference model.

This model was widely used in the early 2000s to describe the sensory integration of multimodal stimuli using Bayesian inference. Some prominent examples are the estimation of robust percepts in multisensory integration (Ernst and Bülthoff, 2004), modeling object perception by using Bayesian inference (Kersten et al., 2004), the Bayesian integration in sensorimotor learning (Körding and Wolpert, 2004) or the question how the brain represents sensory information probabilistically (Knill and Pouget, 2004).

2.4 Sequential inference

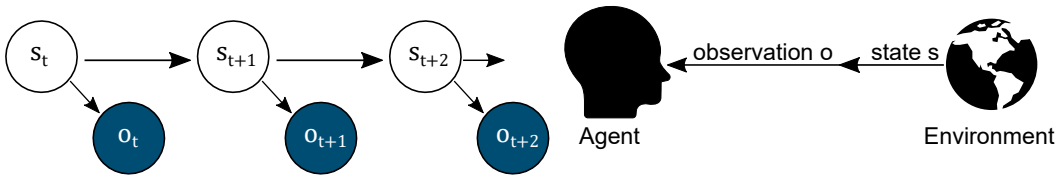


Fig. 2.5: The sequential inference model.

The real world is much more complex than just a specific observation at only one point in time. We can extend the idea of direct inference by including temporal dynamics and handling multiple observations over time. Again, the challenge is to estimate the true state s given the observation o. However, the true states $(s_t, s_{t+1}, s_{t+2}, \dots)$ vary over time and so does the sequence of observations $(o_t, o_{t+1}, o_{t+2}, \dots)$. In other words: Given past observations, the future state is now to be estimated. Like in the direct inference model, we can use Bayes Theorem to incorporate that temporal dynamic and formulate our model as probability:

$$p(s_{t+n}|o_{t:t+n}) = \frac{p(o_{t+n}|s_{t+n}) \cdot p(s_{t+n}|o_{t:t+n-1})}{p(o_{t+n}|o_{t:t+n-1})}$$

The advantage of sequential inference is that the model incorporates information from multiple past observations and updates its estimates as new ones are made. This sequential updating process is a crucial feature that distinguishes it from the previous direct inference model, which does not consider the temporal dynamics.

Sequential inference has multiple applications in the context of active vision. One famous example is the work of Najemnik and Geisler, 2008, where the authors propose a human visual search behavior model and predict an optimal sequence of eye movements, minimizing the number of fixations required to locate a target in noisy visual environments. The main idea behind their sequential inference model is that the human visual system utilizes its knowledge about the target's characteristics, the statistical properties of the background (noise), and the spatial properties of the observer's visual system (such as the decrease in acuity with eccentricity) to plan and execute a sequence of fixations that maximize the likelihood of finding the target. Other applications include models to explain the combination of prior knowledge with attentional cues to guide saccadic eye movements (Eckstein et al., 2006) or decision-making through drift-diffusion models (Bitzer et al., 2014; Fard et al., 2017).

2.5 Single decisions

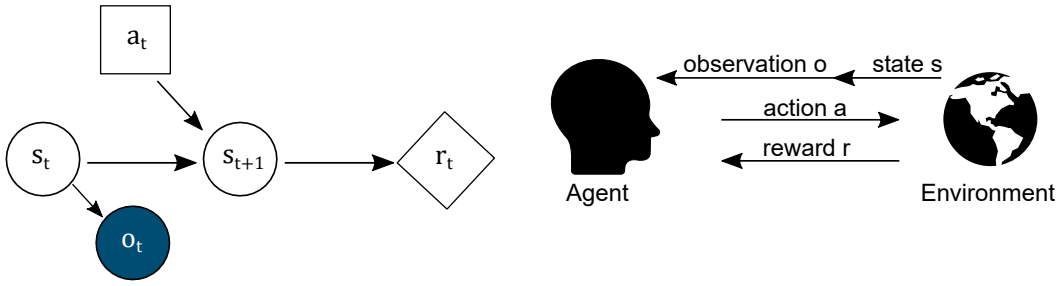


Fig. 2.6: The single decision model.

Our models, up to now, have only very passively taken observations and tried to estimate the actual states of the world from them. However, this is insufficient for modeling many events since the agent is often active and in exchange with the world (see the discussion of passive vs. active vision from the previous chapters). Therefore, we introduce an action a that the agent performs while it is still in the state s_t, which

leads it into a new state s_{t+1}. So the actions can now have positive or negative consequences; this is what we specify through the reward r_t. Therefore, the new task introduced here is not only to estimate the actual state based on observation but additionally to choose to maximize the expected reward. More importantly, this allows us to include action variability, behavioral costs, and sensory uncertainty.

Again, we want to model this mathematically and introduce:

- $r(s_t, a_t, s_{t+1})$, the reward function (either deterministic or probabilistic) which assigns a real-valued reward for taking action a_t in s_t leading to s_{t+1}.
- the probability $p(s_{t+1}|s_t, a_t)$ of transitioning to state s_{t+1} when taking action a_t in state s_t.
- the probability $p(o_t|s_t)$ of observing o_t while being in state s_t.

If we recall the direct inference model from section 2.3, we know that we can now determine the distribution $p(s|o)$ using Bayes' theorem. We can define the framework with these basics, also known as Bayesian decision theory. The expected reward for each action a, taking into account its current estimate about the state s through the observation o, can be defined as:

$$r(a_t) = \sum_{s_t, s_{t+1}} p(s_{t+1}|s_t, a_t) r(s_t, a_t, s_{t+1}) p(s_t|o_t)$$

Thus, the optimal action a_t^* at time step t can be determined as $a_t^* = \text{argmax}_{a_t}\, r(a_t)$.

Hence, we can extend our direct inference model from the beginning to make an optimal decision under uncertainty. For example, the theory of statistical decisions was applied to study how humans select rapid, goal-directed movements under uncertainty (Trommershäuser et al., 2008; Trommershäuser et al., 2003). Other applications include perceptual tasks involving transfer between stimuli (Maloney and Mamassian, 2009), neuroeconomics (Glimcher and Rustichini, 2004) and signal detection psychophysics (Dayan and Daw, 2008).

2.6 Markov Decision Processes

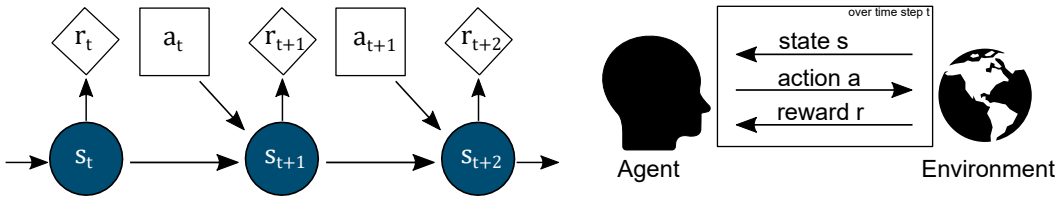


Fig. 2.7: The Markov Decision Process.

We can now combine the concepts we have seen in the previous sections. We have already looked at how we can make an optimal decision. However, what does it look like when we have to make a single decision and a whole sequence of decisions? The Partially Observable Markov Decision Processes framework is beneficial for computational modeling of human sequential behavior under uncertainty. Before we look at this in detail, we start one step earlier. For the sake of simplicity, we assume here that the true states do not have to be estimated via observations but are directly fully observable. This leads us to the underlying Markov Decision Process (MDP) (Bellmann, 1957; Kaelbling et al., 1998), which can be defined as a tuple $(\mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{T}, \gamma)$ with the following components:

- $\mathcal{S}$ is a finite set of states, representing different situations or configurations of the world.
- $\mathcal{A}$ is a finite set of actions, representing the possible decisions that can be made at each state.
- $\mathcal{R}$ is the reward function $r(s_t, a_t, s_{t+1})$, which assigns a real-valued reward to each state-action-state triplet.
- $\mathcal{T}$ is the state-transition function, which represents the probability $p(s_{t+1}|s_t, a_t)$ of transitioning to state s_{t+1} when taking action a_t in state s_t.
- γ is called a discount factor which is a scalar value between 0 and 1. It determines the relative importance of immediate versus future rewards. A value near to 1 means that the agent weighs future rewards heavier, while values towards 0 emphasize immediate rewards.

The goal is now to find an optimal policy π^*, i.e. a function that returns the appropriate action a_t for each state s_t (i.e. $\pi^*(s_t) = a_t^*$) in order to maximize the expected reward over the potentially infinitely large time horizon, in mathematical terms to maximize:

$$\pi^* = \underset{\pi}{\operatorname{argmax}} \quad \mathbb{E}_{p(s_{t+1}|s_t,a_t)} \left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t, s_{t+1}) \right] \quad \text{with } a_t = \pi(s_t)$$

Several methods exists to solve the MDP for an optimal policy. We want to give a short introduction to three of the most popular here:

- Value Iteration (Bellmann, 1957): The idea is to iteratively update the so-called value function $V(s)$ until it converges to the optimal value function $V^*(s)$. Formally, the value of a state is given by

  $$V(s) = \underset{a_t}{\operatorname{argmax}} \sum_{s_{t+1}} p(s_{t+1}|s_t, a_t) \left[r(s_t, a_t, s_{t+1}) + \gamma V(s_{t+1}) \right]$$

  At each iteration, the value of each state is updated based on the expected future rewards of all possible actions. This process is repeated until the convergence. Once the optimal value function is known, the optimal policy π^* can be easily derived by selecting the action that maximizes the expected value in each state.

- Policy Iteration (Bellmann, 1957): In the beginning, an arbitrary policy is chosen and evaluated. Then the algorithm tries to improve this policy with the help of greedy optimization. If the policy changes, the process is repeated until it finally converges and finds an optimal policy π^*. This approach usually requires fewer iterations than the previously presented value iteration. Still, they are also computationally more expensive since they perform a complete policy evaluation at each step.

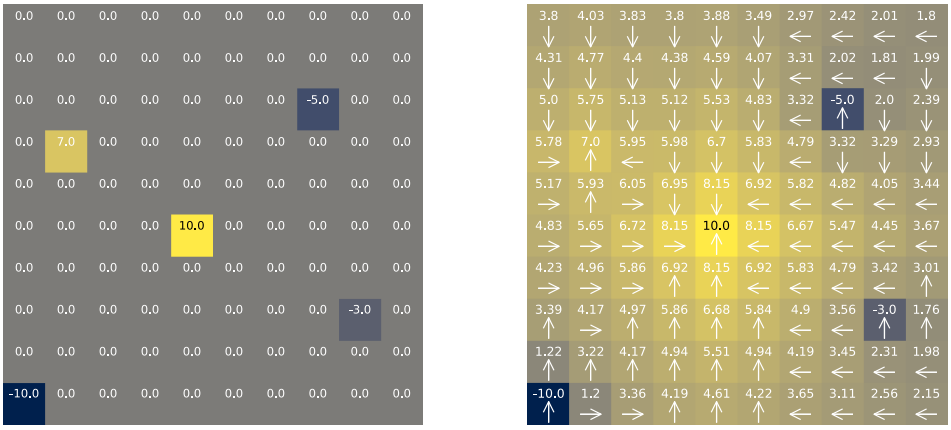
- Q-learning (Watkins and Dayan, 1992): Instead of directly estimating the value function $V(s)$, Q-learning estimates the a state-action value function $Q(s, a)$, which represents the expected reward of taking action a in state s and then following the optimal policy thereafter. The algorithm can be formulated as an update rule:

  $$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[r(s_t, a_t, s_{t+1}) + \gamma \underset{a_{t+1}}{\operatorname{argmax}} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t) \right]$$

  where α is the learning rate. By interacting with the environment and updating the Q-values iteratively using this rule, the algorithm eventually converges to the optimal Q-value function Q^* under certain conditions, from which the

optimal policy can be derived by selecting the action with the highest Q-value in each state.

We consider the famous example of Grid World for better understanding (compare, for example, Sutton and Barto, 2018). The world consists of a 10x10 grid with an agent on it. On each field, there is the possibility to visit adjacent fields; for this, four actions are available in the form of movements in one of the four cardinal directions. Each field has a reward, which can be either positive or negative. Figure 2.8a visualizes the setting. The agent aims to collect as much reward as possible by running to the proper squares, moving with each step. He receives a negative reward of -1 if he runs into a wall. The difficulty now lies in the action variability because only in 70% of the cases he runs in the direction he had in mind. In 30% of the cases, he runs in one of the other three directions. This problem can be implemented in Julia (Bezanson et al., 2017), for example, using the POMDPs package (Egorov et al., 2017) and solved there with the value iteration algorithm. The optimal policy for this sample world can be seen in Figure 2.8b, including the action selected by the policy in each state and the associated value.



(a) Definition of the grid world and its reward structure.

(b) Visualization for the optimal policy learned through value iteration.

Fig. 2.8: The grid world example for MDPs.

MDPs also have various applications in cognitive science, such as modeling the interaction between brain regions in behavioral control (Daw et al., 2005) or formalizing the framework of human decision-making to link them to neural and behavioral findings (Huys et al., 2012; Solway and Botvinick, 2012).

2.7 Partially Observable Markov Decision Process

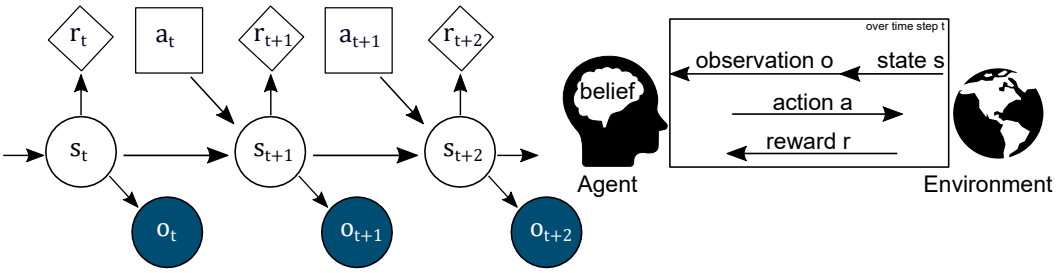


Fig. 2.9: The Partially Observable Markov Decision Process.

A Partially Observable Markov Decision Process (POMDP) is an extension of MDPs to handle problems where the state of the environment is not or only partially observable. With this, we can unite all the concepts previously seen in this chapter. To formulate our POMDP, we extend the MDP tuple with two more components Ω and $\mathcal{O}$ to $(\mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{T}, \Omega, \mathcal{O}, \gamma)$ with:

- Ω is a finite set of observations, representing the (partial) information the model can receives about the current environmental states.
- $\mathcal{O}$ is the observation function, which represents the probability $p(o_{t+1}|s_{t+1})$.

Since the POMDP agent does not know exactly in which state s it is, it has to form a *belief* b about it. Mathematically this can be expressed as a probability distribution over all possible states. This belief can be used to convert every POMDP into a (continuous) MDP, which is also called a *belief MDP* (Åström, 1965; Kaelbling et al., 1998). It is described by a tuple ($\mathcal{B}$,$\mathcal{A}$,r,τ,γ), where:

- $\mathcal{B}$ is a set of belief states, each representing a probability distribution over the true states S.
- r is the belief reward function, $r(b, a)$, which assigns a real-valued reward to each belief state-action pair.
- τ is a belief state transition probability function, which represents the probability $P(b_{t+1}|b_t, a_t)$ of transitioning to belief state b_{t+1} when taking action a_t in belief state b_t. The belief can be recursively based on the history of actions and observations:

$$b_{t+1}(s_{t+1}) = \eta \cdot p(o_{t+1}|s_{t+1}) \sum_s p(s_{t+1}|s_t, a_t) b(s_t) \quad \text{with normalization constant } \eta$$

Converting it into a belief MDP is one way to solve the underlying POMDP, even though it is difficult because the states are continuous (even if the POMDP was discrete). However, other solution methods exist to find an optimal policy for the POMDP. We want to introduce some of them:

- Point-based value iteration (PBVI) (Pineau et al., 2003): Since the belief state can be infinite, the PBVI algorithm selects only a reduced set of points and computes their corresponding values, trying to perform the optimization on them. Throughout the process, more points, which are reachable under the current policy, get added and evaluated. Thereby the approximation for the optimal value function gets more and more refined. This approach dramatically reduces the computational cost and makes the computation feasible.

- QMDP (Littman et al., 1995): This algorithm provides a heuristic solution for the optimal policy, ignoring the uncertainty in future observations. More specifically, QMDP computes the Q-values based on the underlying MDP, which has full observability of the current states. Therefore it can select the actions based on the corresponding belief states. This method provides a suboptimal policy since it doesn't consider the effects of future observations. Still, it is computationally less expensive and can serve as a baseline or initial policy for more complex algorithms.

- Monte Carlo planning (Silver and Veness, 2010): Monte Carlo Planning for POMDPs is a sampling-based method that uses random simulations to approximate the value function. Instead of attempting to compute exact solutions over the entire belief space, it generates many random rollouts from the current belief state. The value of an action is then approximately determined based on the average outcome of these rollouts. The algorithm is a popular variant that incrementally builds a compelling search tree in complex domains with large state and action spaces.

- DESPOT (Somani et al., 2013): This online planning algorithm is suitable for POMDPs with large or continuous state, action, and observation spaces. The idea is to compactly represent all possible future spaces using a sparse tree, where each branch represents a specific sequence of actions and observations. This tree is constructed through a deterministic sampling process, which ensures that only the most likely and impactful scenarios are considered. Focusing on this compact representation significantly reduces computational requirements while delivering near-optimal policies.

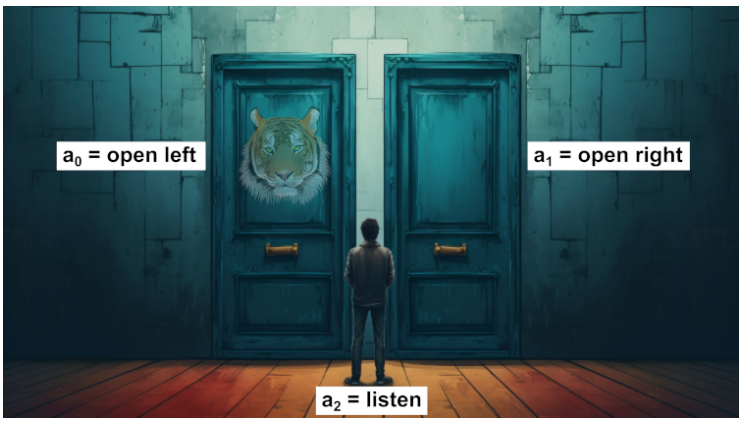


Fig. 2.10: Visualization of the Tiger POMDP.

A famous introductory example is the so-called Tiger POMDP problem (Cassandra et al., 1994), which is shown in Figure 2.10. There are two doors (left and right), and behind one of them lurks a tiger. Our task is to open the door behind which the tiger is not lurking. In this case, partial observability comes from the fact that we can listen at the doors. However, this tells us only in 85% of the cases the correct situation behind the door. Moreover, this reaction costs us a reward. So, given our beliefs about whether the tiger is behind the left or right door, how can we find an optimal policy for which action to choose? Hearing at the door gives us a -1 reward, uncovering the tiger gives us a -100 reward, and escaping gives us +10. Now our belief is initially 50/50 where the tiger is. However, by listening, we can update this. The optimal policy results using the QMPD algorithm are shown in Figure 2.11.

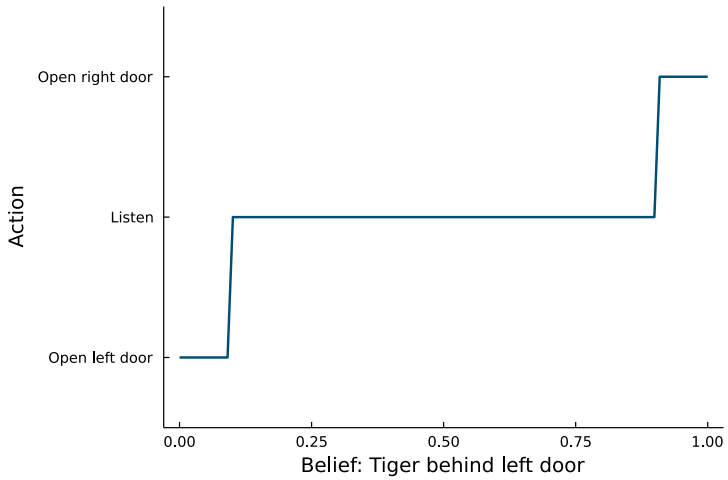
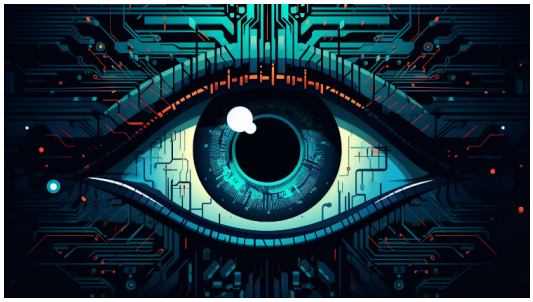


Fig. 2.11: Policy for the Tiger POMDP.

This example clarifies that we can now implement all initial conditions for individual sequential decision-making under uncertainty in the POMDP framework. We can implement **action variability** through probabilistic actions, **sensory uncertainty** through partial observability, **behavioral costs** through the structure of the reward function, and **internal model uncertainty** through belief modeling. Therefore POMDP framework has been used in previous work to model eye movements (Butko and Movellan, 2008, 2009; Stankiewicz et al., 2006), but also more complex situations where vision and action have to infer, such as catching balls (Belousov et al., 2016), navigation (Kessler et al., 2022) or optimal control in detection tasks (Chebolu et al., 2022). POMDPs are a powerful and flexible framework for modeling decision-making under uncertainty. This allows the process of active vision to be implemented and understood.

3 Blinking as probabilistic planning under uncertainty



> *We were all silent except for blink, blink, blink, blink, blink.*
>
> **— Richard Brautigan, 2014**

We blink around 15000 - 20000 times a day and often don't even notice it at all. Eye blinks are one of the most frequent visual behaviors carried out by humans and although at first they seem like a very simple and straightforward action, our visual apparatus has to plan them specifically. First of all we have to blink for maintaining and protecting healthy vision through moisturizing the eyes and clearing debris (Bron et al., 2004; Doane, 1980; Sweeney et al., 2013; Tutt et al., 2000), while at the same time providing rest for the eyes and enabling visual processing (Bristow et al., 2005). However, during blinking the stream of visual information is interrupted (Riggs et al., 1981) and we can thereby miss important information (Wiseman and Nakano, 2016). Therefore, we need to find a trade-off to keep our eyes moist and healthy while missing as little to no visual information as possible.

3.1 Related work

Previous research by Hoppe et al. (Hoppe et al., 2018) has shown that human observers are capable of this through adapting their blinking behavior to environmental event statistics by trading off the cost of information loss due to short-term eye closure with the benefit of maintaining healthy vision. Further studies provide additional empirical evidence for the link between blinking behavior and several cognitive processes. This includes a decrease in blink rate while the presentation

stimuli that are relevant for the current task (Groen et al., 2017) or especially engaging moments (Nakano et al., 2009). The same decrease is also observed in situations that subjects remember particularly well at a later time (Shin et al., 2015). These different blink rate patterns are highly individual and allow to categorize individual's engagement and interest with scene content (Nakano and Miyazaki, 2019; Ranti et al., 2020). In addition to the blink rate, the interblink interval (IBI) distribution (i.e. the distribution of times between two consecutive blinks) is also highly variable between subjects. Four different forms of these distributions (J-shaped, bimodal irregular, symmetric) were described long ago in 1927 (Ponder and Kennedy, 1927) and have been found repeatedly in empirical measurements ever since (Borges et al., 2010; Lenskiy and Paprocki, 2016; Naase et al., 2005; Nishizono et al., 2023). Hoppe et al. (Hoppe et al., 2018) showed that these differences are based on the trade-off between internal and external costs of the current task and could reproduce three of the distributions using their cost model (missing the symmetric shaped IBI).

In addition to these individual differences, there are also shared commonalities in the blinking behavior that can be found across subjects. These are often found in relation to highlights or events in various visual scenes. For example, in drivers of Formula 1 cars at important positions on the track (Nishizono et al., 2023), in key events based on the narrative while watching movies (Andreu-Sanchez et al., 2021) or in highlight scenes of sports videos (Nakano et al., 2020). These influences of the temporal rate of events seem to have a major influence on individuals' blinking behavior. The approximative computational model for relating the costs and benefits of blinking to the probability of blinking by Hoppe et. al (Hoppe et al., 2018) was able to predict the blink behavior of all subjects given the generative event model, but does not take the modeling of temporal event rates either. On promising approach to capture the dynamics of temporal event arrivals can be non-homogenous Poisson Processes (NHPPs). For example, temporal point processes were used in modeling even statistics in video stimuli of team sports such as ice hockey (Zhong et al., 2018). NHPPs have many different application areas in which event arrivals are modeled that do not follow a fixed rate. Examples include the modeling in many different areas like the distributions of plants and animals in ecology (Mugumaarhahama et al., 2022; Warton and Shepherd, 2010; Wilson and Costello, 2005), the failure and repair of components in reliability analysis (Davies et al., 2021; Saldanha et al., 2001; Shibata et al., 2006), weather and climate phenomena in metrology (Achcar and Oliveira, 2022; Lu and Garrido, 2005; Ngailo et al., 2016), noise levels in acoustics (Guarnaccia et al., 2016, 2015; Peeling et al., 2007), the occurrence of disease incidents or virus spreading in a population in epidemiology (Al-Dousari

et al., 2021; Cifuentes-Amado and Cepeda-Cuervo, 2015; Guler Dincer et al., 2022), action potentials and neurophysiological responses in neuroscience (Chaspari et al., 2014; Turcott et al., 1994), the distribution of networks in telecommunications (D'Angelo et al., 2022; Vales-Alonso et al., 2013), the spatial structure and flow in traffic analysis (Lim et al., 2016; Zhang et al., 2019) or the distribution of earthquakes in seismology (Jonsdottir et al., 2006; Shcherbakov et al., 2005).

Here we make use of the NHPP as natural modeling approach for temporal event rates and develop a probabilistic planning model using partially observable Markov decision processes (POMDPs, (Cassandra et al., 1994; Kaelbling et al., 1998)). POMDPs provide a natural and powerful computational framework of modeling uncertainty in both action outcome and state perception to model complex decision-making problems where an agent's state observation is incomplete or noisy. Using this approach allows us to to quantitatively understand the blinking behavior and inferring individuals' internal costs associated with not blinking, their benefits for detecting an event, and their perceptual beliefs implicit in behavior. Our model is able to predict key aspects of human blinking of the analyzed behavioral data. In addition to this, our model succeeds as first in predicting all four IBI distributions. We can analyze the natural functional relationship between the cost and the probability of blinking. Unlike previous work, we do not need to base or estimate a cost function, but can extract it from our normative model. Using this study, we consider the simplest visual action, blinking, as an active decision process. This framework can be extended to other visual actions such as gaze selection in the future (Kadner et al., 2023a, 2022). We conclude that human blinking behavior agrees with probabilistic planning under uncertainty.

3.2 Methods

3.2.1 Experimental design and generative model

For this chapter, we use the data collected from Hoppe et al., 2018 and their experimental setups. The experiment consists of a visual detection task, where subjects had to observe a gray circle moving on a circular path. The stimulus will move on this circular trajectory for 100 consecutive laps while the trajectory self was not visible for the subjects. In each of these laps 3-5 events will occur, consisting of the grey circle being replaced with a stylized face for 50 ms. Subjects' task is to detect as many of the events as possible by pressing a button. If they missed to detect an event, they were alerted by an acoustic signal. The positions of the appearance

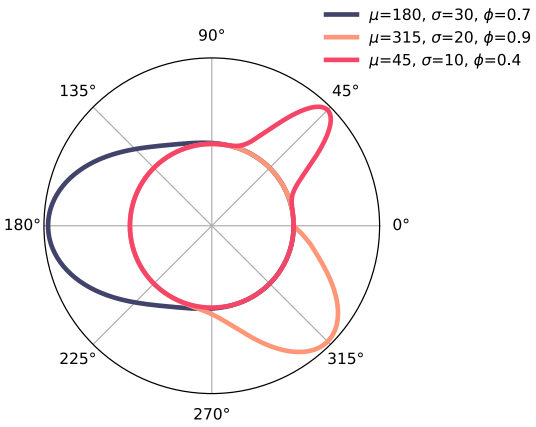


Fig. 3.1: Three exemplary mixture distributions of the generative model from Equation 3.1.

of the events were not random, but followed a generative probability distribution. This model (see Equation 3.1) for generating events is a mixture distribution of a normal distribution (with parameters μ, σ and a mixture component ϕ) and a uniform distribution (between 0° and 360° and a mixture component $(1-\phi)$) (see Figure 3.1 for a visualization).

$$p_{\text{gen}}(\theta|\mu,\sigma,\phi) = \phi \cdot \mathcal{N}(\theta|\mu,\sigma) + (1-\phi) \cdot U(0,360) \tag{3.1}$$

In the original task, mixture models with $\sigma = 30$ and $\phi = 0.8$ and varying μ were chosen. For visualization in the analysis every value for μ was translated to be $\mu = 180$.

The analysis of the available data could confirm the results found by the authors. We did additional analysis on the duration of blinks, given their relative position to the high probability regions on the circle. Besides the position of the blinkers on the circle, this could be another measure that subjects have adaptively adjusted their blinking behaviour to the temporal event statistics. Figure 3.2 shows the duration given the distance to point of highest probability. We fit a linear regression model which reveals positive correlation between distance and blink duration ($r \approx 0.67$, $p < 10^{-12}$). In fact, we find another correlation that the blink duration of subjects is also adjusted to the generative model. If they already have to blink in a region of higher event probability, they do it statistically significantly faster there.

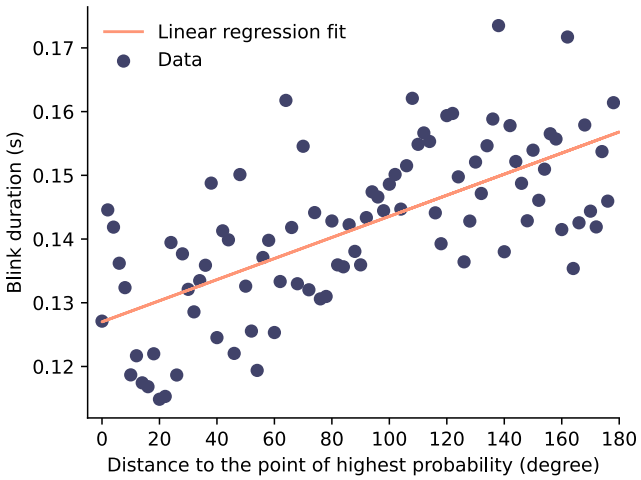


Fig. 3.2: Blink duration dependent on the relative position to the highest point of probability. Data was aggregated over all participants and trials and discretized with a window size of two degree. A linear regression was fit to the data.

3.2.2 Non-homogeneous Poisson point process

Many modeling problems involve counting the number of occurrences of a particular event within a specified period, e.g. the number of users visiting a website, the number of particles passing through a detector, the number of patients arriving at a hospital, or the number of failures in a system. This can be modeled with counting processes, which are defined as stochastic processes $\{N(t), t \geq 0\}$ which represent the number of events occurring up to a time step t. For all $t \geq 0$, the value $N(t)$ is a nonnegative integer, and the function $N(t)$ is right-continuous and non-decreasing. Furthermore the number of events in an interval $(t_n, t_{n+1}]$ can be described by $N(t_{n+1}) - N(t_n)$. Counting processes provide a powerful and flexible framework for modeling the random behavior of discrete events over time.

One specific type of counting processes are the Poisson processes, where the times between successive events are independent and identically distributed exponential random variables and the number of events in an interval follow Poisson random variables. We can distinguish two different types for the Poisson point processes: the homogeneous and non-homogeneous one. The homogenous Poisson point process has a rate parameter $\lambda \geq 0$, which is the expected number of events per unit time, i.e. the number of events in that interval is a Poisson random variable with mean $\lambda \cdot (t_{n+1} - t_n)$. The homogeneous variant of these processes has been employed widely, benefiting from its simplicity and the powerful theoretical results that surround it. However, many real-world phenomena do not adhere to the assumption of a constant event rate. To account for this the Nonhomogeneous Poisson point process

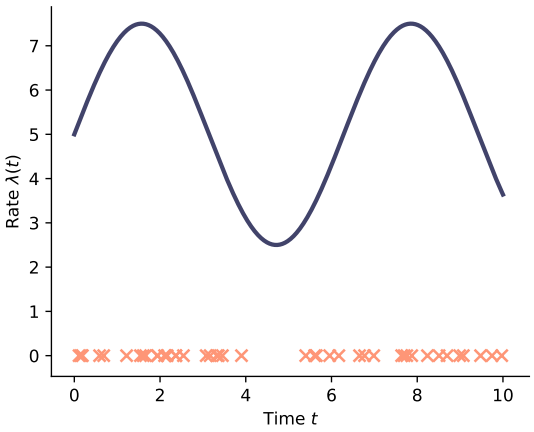


Fig. 3.3: Visualization of the Non-homogeneous Poisson point process on the real line. The exemplary rate function $\lambda(t) = 5 + 2.5 \cdot \sin(t)$ and the corresponding sampled Poisson events on the real line are shown.

(NHPP) has no longer a constant rate λ, but an intensity function $\lambda(t)$ over time. Thus, for the interval $(t_n, t_{n+1}]$, this gives the relationship that the number of events is Poisson distributed, with the following relationship:

$$N(t_{n+1}) - N(t_n) \sim \frac{\widetilde{\lambda}^k}{k!} e^{-\widetilde{\lambda}} \quad \text{with} \quad \widetilde{\lambda} = \int_{t_n}^{t_{n+1}} \lambda(u)\,\mathrm{d}u \tag{3.2}$$

This elementary feature of the NHPP is visualized in Figure 3.3. The events cluster exactly at the points where the intensity function assumes the most area under the curve. The number of events can be described by the area of the intensity function as a parameter of a Poisson distribution. The inter-arival time, i.e. the time between the individual events, can also be described using exponential distributions (Yakovlev et al., 2005).

We are interested in the distribution of the arrival times of the events (in Figure 3.3 the distribution of the positions of the crosses). To find an expression for the time step of the n-th event, we can define two different equations:

$$T(n) = \min\{t \geq 0 : N(t) = n\} \tag{3.3}$$

$$N(t) = \#\{n \in \mathbb{N} : T(n) \leq t\} \tag{3.4}$$

Using these equations and the non-homogenous Poisson distribution of $N(t)$ we can define the distribution function of the arrival times as:

$$p(T(n) \leq t) = p(N(t) \geq n) = \sum_{i=n}^{\infty} e^{-m(t)} \frac{m^k(t)}{k!} \quad \text{with} \quad m(t) = \int_0^t \lambda(u)\,\mathrm{d}u \tag{3.5}$$

Differentiating with respect to t gives our relationship, The probability for the arrival of the n-th event at point t:

$$p_{\text{arr}}(t|n) = \frac{\mathrm{d}p(T(n) \leq t)}{\mathrm{d}t} = \lambda(t)e^{-m(t)} \sum_{i=1}^{n} \left[\frac{m^{k-1}(t)}{(k-1)!} - \frac{m^k(t)}{k!} \right] = \frac{m^{n-1}(t)}{(n-1)!} \lambda(t) e^{-m(t)} \tag{3.6}$$

where $m(t)$ is often called the mean value function (Siegrist, 2021). With this, we are now able to model the statistics of temporary event arrivals.

3.2.3 Model components

Our probabilistic planning model is motivated by the framework of Partially Observable Markov Decision Processes (POMDPs) (Cassandra et al., 1994; Kaelbling et al., 1998). The implementation was done in Julia (Bezanson et al., 2017) with the *POMDPs* package (Egorov et al., 2017). The QMPD algorithm (Littman et al., 1995) was used to solve the POMDP and learn a policy to simulate our model.

State space

The state space S follows the experimental setup with the features to uniquely describe the current state. This includes the current position of the stimulus θ (discretized between 0 and 359 degrees), the remaining number of events in this lap n, and a boolean variable indicating whether an event occurs at this current position.

Action space

The action space A is limited to the blinking behavior and consists of two possible options. The model can choose either to blink or to keep the eyes open at a specific position.

State transition

The state-transition function T describes dynamics of the experiment. Therefore it is responsible for updating the environmental states according to the underlying generative model. First, it describes the position of the stimulus over the circular path and its update. If the agent does not blink, the stimuli moves on discretely by one degree. Otherwise the eyes are closed for the duration of a blink and the stimuli moves accordingly in that time. In addition, the state-transition function draws uniformly 3-5 events at the beginning of each new lap and sets their positions according to the generative model (Equation 3.1).

Observations

The true states of the experiment are only partially observable for the subjects and therefore for our model. Thus, it can perceive the current position of the stimuli only with addition of perceptual uncertainty. Also observable is whether an event has just taken place (and thus when it last took place), as well as the remaining number of events, given the initial belief.

Rewards and costs

The agent can either decide to keep his eyes open, but then gets a negative reward for suppressing it c_s. However, if he blinks and an event takes place during the blinking period, he gets a negative reward c_e for missing it. These two costs are in a relationship and we can trade-off them with $\alpha \cdot c_s$ and $(1-\alpha) \cdot c_e$.

Belief

Since the true states are not known to the model, but it can only partially observe them, it must form a belief about them. Each subject has an individual belief over the true parameters of the generative model and the number of remaining events for the current lap. We denote this belief vector as $\psi = (\tilde{\mu}, \tilde{\sigma}, \tilde{\phi}, \tilde{n})$.

Using our NHPP we can model our belief that an event will happen at position θ given the position of the last event θ_{prev} and the belief about the generative event statistics vector ψ. Therefore we can combine the probability that the event did not occur in the interval $(\theta_{\text{prev}}, \theta)$ and the probability to occur at position θ:

$$p_{\text{event}}(\theta|\theta_{\text{prev}}, \psi) = \left[1 - \int_{\theta_{\text{prev}}}^{\theta} p_{\text{arr}}(u|\tilde{n})\,\mathrm{d}u\right] \cdot p_{\text{arr}}(\theta|\tilde{n}) \tag{3.7}$$

Using this relationship, we can calculate a probability for each time point that an event will occur at the current angle θ of the stimulus. As an intensity function we can use the underlying generative model and exploit the fact that it is a probability distribution. The area under the curve equals by definition one, which corresponds exactly to a single event on the whole circle. To model the expected remaining number of events, the intensity function can be scaled by the probability function of the generative model.

In order to re-continuate the Belief, we use a Kernel Density Estimator of the form:

$$\hat{f}(\theta) = \frac{1}{n}\sum_{i=1}^{n} K_\lambda(\theta - \theta_i) \quad \text{with} \quad K_\lambda(\delta) = \frac{e^{\kappa\cos(\delta)}}{2\pi I_0(\kappa)} \tag{3.8}$$

where δ is the angular difference between evaluation and data point, κ is the concentration parameter and I_0 is the modified Bessel function of order 0.

3.3 Results

3.3.1 Model analysis

Using the definitions of the model components, we can simulate our model for different beliefs about the underlying event statistics and analyze the resulting behavior. The approximate cost model of Hoppe et al. (Hoppe et al., 2018) previously set the relationship between cost and blink behavior manually. However, the normative modeling here allows us to look at the relationships as they are naturally induced by the framework. The internal and external costs are weighted by the trade-off parameter α. The influence of this cost-trade-off on the blinking behavior are shown in Figure 3.4A. Values close to zero can be interpreted that it is particularly important for the model to maintain healthy vision. This can be seen in the top row of the first column, as the model shows constant blink behavior regardless of the exact position. As the alpha value increases, the visual detection task becomes more important, and so the blinking behavior over the laps adapts more and more inversely to the generative model.

The effects of different alpha values on the IBI distributions can be seen in the bottom row of Figure 3.4A. We can observe the recreation of the four IBI distributions first described by (Ponder and Kennedy, 1927) - namely from left to right: irregular, symmetric, bimodal, J-shaped. To our knowledge, this is the first time that a model is capable to generate all four of them.

Furthermore, we can take a look at the correlation of different trade-off values on the performance of the event detection task. Figure 3.4B visualizes the relationship, where we used the percentage of correct detected events as an measurement for performance. As conjectured by (Hoppe et al., 2018), we observe a functional relationship between both variables. Our normative model has the advantage that we now are able to also computationally verify this relationship and contrary to the linear empirical observation and assumption from the previous study, we observe logistic relationship between both. A performance of 50% is already achieved with an alpha value of around 0.2.

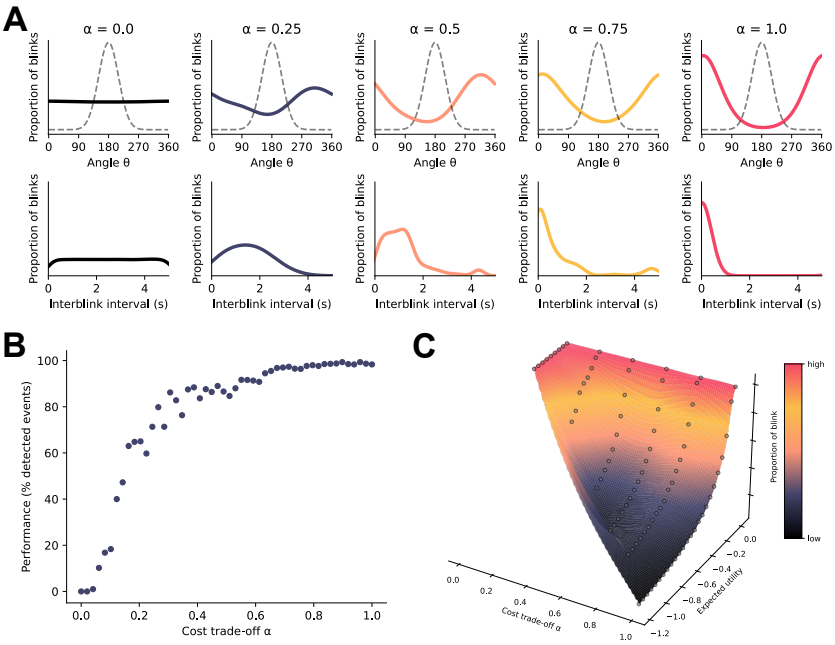


Fig. 3.4: Influence of the trade-off parameter α on blinking behavior. (A - Top row) Blinking proportion per degree. At low alpha values, the visual detection task does not matter and the model simply blinked equally distributed over the entire course to maintain healthy vision. The higher the values, the more the blinking behavior fits the underlying generative model (gray dashed). (A - Bottom row) Interblink interval distribution. At low values, an irregular blink pattern is seen. (B) The influence of different α trade-off values on the performance, here the percentage of detected events. (C) The relationship between α trade-off values and expected utility values on blinking behavior.

Further, it was assumed by (Hoppe et al., 2018) that the costs are inversely proportional to the probability of blinking at an angle θ. More specifically for costs of blink suppression c_s and blink execution c_e they proposed

$$p(\text{blink at } \theta|\alpha, \phi) = \frac{1}{(1-\alpha)c_s + \alpha c_e(\theta, \phi)}$$

We can also check this relationship using our model by looking at the learned policy for problem solving. The QMDP algorithm for solving the underlying POMDP returns an alpha vector policy. This policy consists of a set of alpha vectors, where every alpha vector correspond to a specific action. It provides the expected cumulative reward for each state if that specific action is taken while the agent is in that belief state. With the dot product of the belief vector and an alpha vector, we are effectively averaging the expected rewards across all states, weighted by the agent's belief in each state. This results in the expected reward of taking a certain action, given the agent's current belief state. This expected reward serves as a crucial piece of information for decision-making in POMDPs: given its belief about the world's state, the agent will typically choose to take the action that maximizes this expected reward. Using this measure, we test the previously presented assumption between the relationship of the blink probability and the underlying cost. Figure 3.4C and Figure 3.5 show the relationship between the cost-trade-offs and the expected utility on the blinking behavior. We can see two different influences in the dimensions and an interaction effect of both of them. Our model now makes it possible to extract and visualize this complex relationship at all.

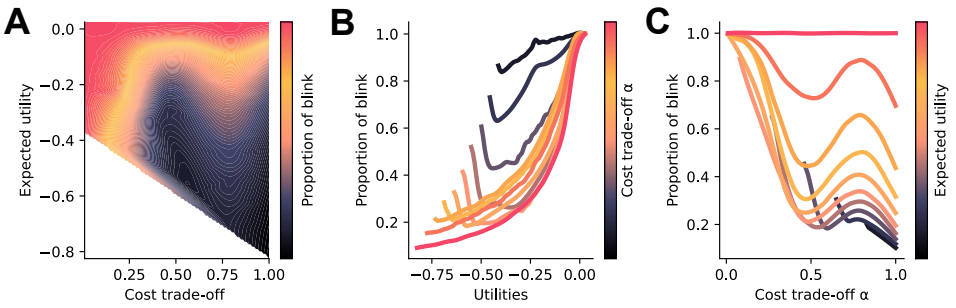


Fig. 3.5: The relationship between α trade-off values and expected utility values on blinking behavior. (A) Contour line plot of Figure 3.4C. (B) Marginal plots for utilities against probabilities. (C) Marginal plots for cost-trade-offs against probabilities.

3.3.2 Model results for blinking behavior

The model was fitted to the behavioral data provided by (Hoppe et al., 2018). The results are shown in Figure 3.6 and the estimated parameters for each individual subject can be found in Table 3.1. Single plots for every participant fit are shown in Figures 3.7 and Figure 3.8.

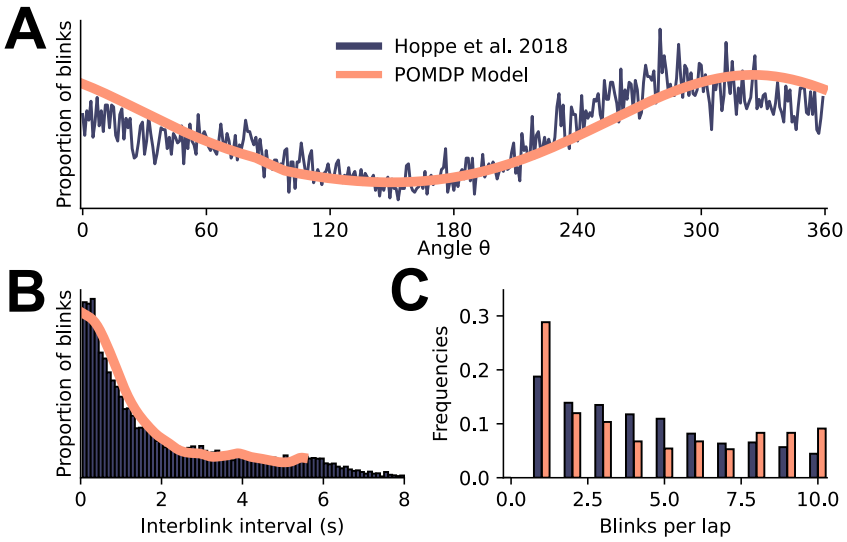


Fig. 3.6: Results of our probabilistic planning model. (A) Blinking proportion per degree (B) Interblink interval distribution (C) Blink rate in terms of blinks per lap.

Like the original model, our model is able to reproduce the blinking behavior of the subjects given the generative model and the current position of the stimulus (see Figure 3.6A. In addition to the adaption to the event probabilities of the generative event statistics, the model also shows the two phases of suppression and compensation. Furthermore, the model predictions are consistent with the empirical distribution of IBI times (see Figure 3.6B and blink rate (see Figure 3.6C). Thus, in addition to the advantages due to the better interpretability and naturalness of the model, it also provides the possibility of explaining real, empirically measured data.

3.4 Discussion

This chapter motivated and implemented a probabilistic planning model to predict human blinking behavior given temporal event statistics. Few computational models exist to predict blink behavior. Previous approaches assume voluntary blink suppression and do not address underlying task-related circumstances (Berman et al., 2012; Moraitis and Ghosh, 2014) or merely use previously empirically measured probability distributions as look-up-table (Ford et al., 2013). Other deep learning approaches can predict temporal points of blink, but do not provide any information about internal costs or underlying individual differences (Nakano et al., 2020). The approximate computational model by Hoppe et. al (Hoppe et al., 2018) related the costs and benefits of blinking to the probability of blinking, but they have not investigated the influence of the temporal rate of events. In order to quantitatively understand the blinking behavior and link it to the temporal component of event arrival, we are using partially observable Markov decision processes combining them with non-homogenous Poisson processes. Using this computational planning framework, we can estimate individual behavioral costs associated with blink suppression and the reward for detecting event, as well as the the perceptual beliefs about the underlying statistics. This approach results in further advantages that were not possible with previous models. We could simulate the relationship between the cost-trade-off parameter and task performance revealing a logistic relationship. Furthermore we developed a normative model to describe human behavior where we have not assumed or fit any functional form of the cost functions. This is common practice in many models in the field (compare e.g. (Petitet et al., 2021)), so our framework offers long-term opportunities to provide new computational modeling capabilities in other fields of cognitive science. Furthermore, we performed an additional analysis using the available data from Hoppe et al. We found a statistically significant correlation, between the blink duration and the position on the circle, given the generative model. These results suggest that, in addition to the local component, a temporal component also plays a role in blinking. Even when balancing physiological and task performance, once they have to blink in a region with higher event probability, they do so there for a significantly shorter time. Previous models have simulated the duration of a blink either constant or randomized given a predefined probability distribution. Future work should consider this important component and include the temporal information of the blinker.

Tab. 3.1: Individual parameter estimates for all subjects.

Subject	$\tilde{\mu}$	$\tilde{\sigma}$	$\tilde{\phi}$	$\tilde{\alpha}$
1	180	20	0.80	0.68
2	180	40	0.80	0.65
3	180	50	0.60	0.66
4	180	20	0.80	0.59
5	150	20	0.60	0.44
6	180	30	0.95	0.77
7	170	20	0.90	0.85
8	180	10	0.80	0.77
9	180	15	0.80	0.91
10	150	20	0.80	0.62
11	180	40	0.80	0.41
12	150	20	0.80	0.65
13	180	40	0.80	0.67
14	180	30	0.80	0.64
15	180	10	0.80	0.73
16	180	40	0.80	0.63
17	180	30	0.90	0.77
18	180	15	0.80	0.83
19	180	15	0.90	0.73
20	180	30	0.90	0.83
21	180	30	0.80	0.53
22	180	30	0.95	0.68
23	150	20	0.80	0.57
24	180	40	0.80	0.51
25	180	30	0.80	0.73

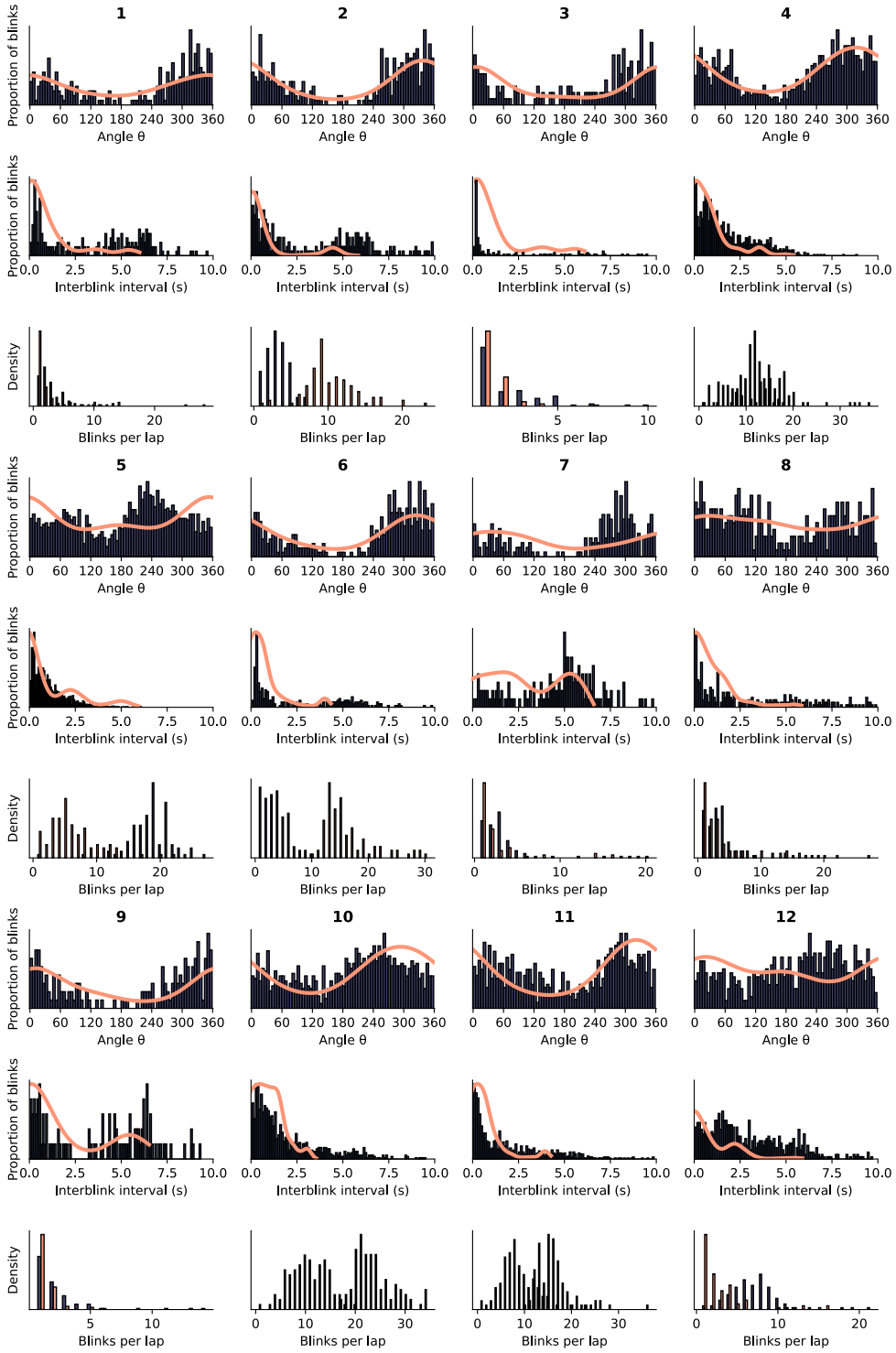


Fig. 3.7: Individual data and model fits for all participants (1–12). For each subject the blinking positions over the circle, the distribution of IBIs , and the distribution of blinks per lap are shown.

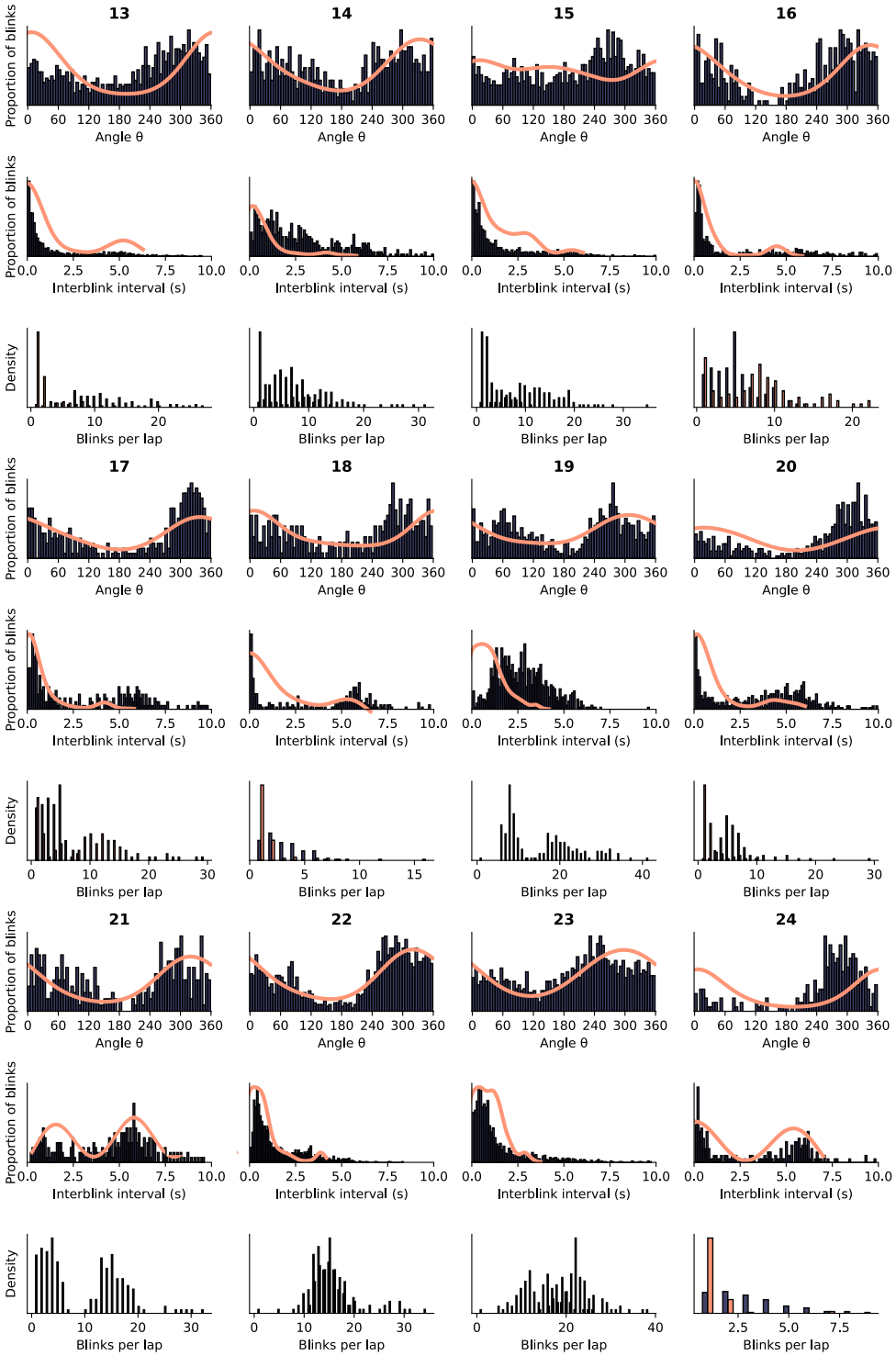
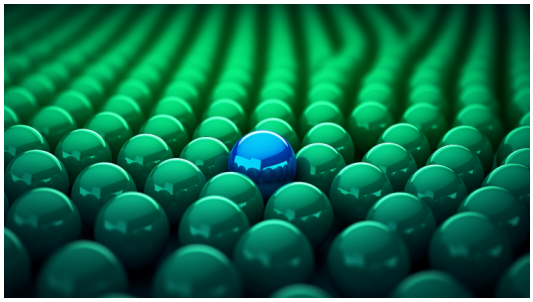


Fig. 3.8: Individual data and model fits for all participants (13-24). For each subject the blinking positions over the circle, the distribution of IBIs , and the distribution of blinks per lap are shown.

4

Improving one-step ahead predictions of subsequent fixations using individual cost maps



> *It's not what you look at that matters, it's what you see.*
>
> **— Henry David Thoreau, 2006**

Because of the inhomogeneous spatial acuity of the visual system, humans shift their gaze sequentially across visual scenes using saccadic gaze movements (Findlay and Gilchrist, 2003). Four main factors have been shown to influence observers' eye movements: the ongoing task, image features such as contrast and intensity, semantic features such as faces and scene context, but also factors that arise from the sequential interaction of an observer with the scene including center bias, proximity preference, inhibition of return (Tatler et al., 2011). Empirically, gaze targets of multiple human observers while inspecting an image given different task instructions can be collected. Assuming that gaze prioritization is image-computable, the task of predicting gaze prioritization given an image is referred to as visual saliency modeling resulting in a time invariant saliency map, whereas scanpath models generate a sequence of gaze targets. While originally developed to account for the phenomenon of pop-out (Treisman and Gelade, 1980), visual saliency modeling has been generalized to predicting the likelihood of human observers overtly shifting their gaze to image regions for arbitrary images.

4.1 Related work

Initially, saliency models used handcrafted features inspired by neurophysiological properties of the visual system (Itti et al., 1998; Sun, Xiaoshuai et al., 2012), but more recently data driven approaches (Cornia et al., 2018; Fan et al., 2018; Huang et al., 2015; Jia and Bruce, 2020; Kruthiventi et al., 2015; Kümmerer et al., 2017; Pan et al., 2016; Wang et al., 2018) have commonly used features from DNNs pretrained on large image datasets, thereby improving performance on various benchmarks (Borji, 2019; Borji and Itti, 2013). These improvements are due to the rich image structure learned by DNNs when trained on large image datasets, e.g. the VGG19 (Simonyan and Zisserman, 2015) underlying Deep Gaze II (Kümmerer et al., 2017) is trained on object recognition of one Million images before tuning to saliency problems using the SALICON dataset (Jiang et al., 2015) containing 10000 images. Scanpath models, by contrast, take an image as input and generate a full scanpath, i.e. a sequence of individual fixation locations as output (Assens et al., 2017, 2019; Boccignone and Ferraro, 2004; Liu et al., 2013; Wang et al., 2011; Xia et al., 2019).

Progress on a variety of benchmarks and image databases has been made, but, a fundamental difficulty with scanpath models compared to saliency models is the well known variability of individual gaze sequences between observers, which poses particular challenges for evaluating the quality of predictions. A recent comprehensive theoretical and empirical evaluation and comparison of scanpath models (Kümmerer and Bethge, 2021) has revealed that scanpath similarity metrics can score wrong models better than the generating model. The resulting analysis in (Kümmerer and Bethge, 2021) establishes that a more consistent and meaningful task consists in the prediction of the next fixation target conditional on past fixations within an image, which is the task we adopt in this study.

Here, we leverage two recent developments to improve the prediction of the next fixation of human observers given an arbitrary saliency map and the sequence of preceding fixations: the theoretical analysis of scanpath models (Kümmerer and Bethge, 2021) and the recently measured human cost function for gaze shifts (Thomas et al., 2022). We adopt a computational account of the scanpath as a sequential decision process in the spirit of previous approaches (Hoppe and Rothkopf, 2019; Jiang et al., 2016; Mathe and Sminchisescu, 2013), but differently from these approaches, our algorithm can utilize arbitrary saliency maps as input instead of estimating rewards for image features from scratch. First, we reason that saliency corresponds to the reward associated with the free-viewing task which is approximated by marginalizing over all visual tasks. The reason is, that the

free-viewing task is maximally ambiguous regarding its task goal. Second, our formulation allows incorporating a map representing the human preferences for gaze shifts, which have recently been estimated for the first time independently of image content through a human psychophysical experiment (Thomas et al., 2022). This gives a computational explanation for commonly used heuristics including the proximity preference. Third, we account for past fixations through a temporally changing exploration map and present the resulting predictions of subsequent gaze targets. The relative contributions of these three components were optimized on the MIT1003 dataset for the NSS score and are sufficient to significantly outperform predictions of the next gaze target on NSS and AUC scores for five state of the art saliency models on three image data sets.

The concept of saliency lies at the intersection of cognitive science, neuroscience, and computer vision (Tatler et al., 2011). Empirically, human gaze targets depend strongly on the ongoing task (Hayhoe and Ballard, 2005) but humans tend to look preferentially at certain areas even when free-viewing images (Henderson and Hollingworth, 1998). These observations have been complemented with the discoveries of multiple retinotopic maps in the visual system (Treue, 2003). While the exact relationships between attention, gaze sequences, and their neuronal underpinnings are still heavily debated, visual saliency modeling has become a canonical computer vision task.

Initially, saliency models used handcrafted lower level features like intensity, color, and orientation (Itti et al., 1998), whereas current DNN based algorithms determine salient regions in a data driven fashion by reusing learnt features e.g from CNNs (Kruthiventi et al., 2015). Other studies have emphasized the importance of higher level information in images, such as text and faces (Bylinskii et al., 2019; Cerf et al., 2008) or general semantic content (Henderson et al., 2019; Pedziwiatr et al., 2019). Relevance might be biased, e.g. towards text, see (Assens et al., 2019) for a discussion. Some approaches have incorporated task goals into models of gaze selection (Borji et al., 2012; Navalpakkam and Itti, 2005), albeit with a small number of tasks with respect to the broad range of human visual and visuomotor tasks. Semantic information has been incorporated by neural network approaches, for example by pretraining on object recognition (Cornia et al., 2018; Huang et al., 2015; Kruthiventi et al., 2015; Kümmerer et al., 2017; Pan et al., 2016; Wang et al., 2018). The work on attention in DNNs, e.g (Mnih et al., 2014; Welleck et al., 2017) is somewhat complimentary (Adebayo et al., 2018), as it is not necessarily modeling overt shifts of attention by gaze shifts but sequential processing of internal representations. Overall, visual saliency modeling is an established field with

canonical datasets and benchmarks and progress on these benchmarks has been steady (Borji, 2019; Borji and Itti, 2013).

Scanpath models have received less attention compared to saliency models but recent approaches include models based on biological and cognitive facts (Itti et al., 1998; Wang et al., 2011; Zanca et al., 2019), statistically motivated models (Boccignone and Ferraro, 2004; Liu et al., 2013; Xia et al., 2019), and models, which leverage machine learning techniques for prediction without reference to underlying mechanisms of vision (Assens et al., 2017, 2019). While some algorithms require an image as input (Cornia et al., 2018; Droste et al., 2020; Fan et al., 2018; Itti et al., 1998; Jia and Bruce, 2020; Kümmerer et al., 2017; Sun, Xiaoshuai et al., 2012), other models use a saliency map as input for generating a scanpath (Assens et al., 2017, 2019; Boccignone et al., 2020; Boccignone and Ferraro, 2004; Itti et al., 1998; Sun, Xiaoshuai et al., 2012; Wang et al., 2011; Xia et al., 2019; Zanca et al., 2019). In (Boccignone et al., 2020; Boccignone and Ferraro, 2004) the authors investigated the properties of scanpaths as function of parameters in random walks on saliency maps. (Wang et al., 2011) proposed a model incorporating an image representation map based on filter responses, foveation, and a memory module to generate sequential saliency maps. (Sun, Xiaoshuai et al., 2012) used an algorithm based on projection pursuit to select image targets for simulating scanpath in order to mimic the sparsity of human gaze selection. PathGAN (Assens et al., 2019) extracts image features using available DNNs and trains recurrent layers to generate and discriminate scanpaths in a training set. While PathGAN learned scanpaths end-to-end and outperformed several other models, qualitative results suggest persistent deviation to scanpaths of human observers.

Of particular relevance in this context is recent work on characterizing and evaluating the prediction accuracy of scanpath models relative to human gaze (Kümmerer and Bethge, 2021). The authors' in depth analyses show that some scapath similarity metrics such as ScanMatch (Cristino et al., 2010) or MultiMatch (Jarodzka et al., 2010) can score wrong models better than the generating model given ground truth. Note also, that some of the current scanpath models employ statistics of scanpaths as a means to capture behavioral biases of gaze shifts, but these have so far never been measured independently of image content. The in depth analyses in (Kümmerer and Bethge, 2021) convincingly lead to the conclusion, that instead of comparing entire scanpaths it is more adequate to evaluate models regarding their prediction of the next fixation within a given scanpath.

Finally, scanpaths have also been conceptualized as sequential decision problems, which is particularly successful in situations where observers' goals are known and

can therefore be formalized as rewards (Hoppe and Rothkopf, 2016; Hoppe and Rothkopf, 2019; Mnih et al., 2014). Very much related to the present approach, (Mathe and Sminchisescu, 2013) used inverse RL to estimate implicit rewards from human gaze sequences. While this extracts reward functions in terms of image features, it is agnostic in relation to internal, behavioral costs and benefits. Other studies have used reinforcement learning to predict scanpaths (Jiang et al., 2016) with a state consisting of low-level features, semantic features, center bias, spatial distribution of eye-fixation shifts as well as a measure indicating previous gaze visits. However, these studies did not utilize the human cost for eye movements measured independently of image content and predicted 'fixation stages' in an experiment and not individual fixations. But empirical studies have shown, that oculomotor biases are not independent of image content, e.g. by simply rotating images (Foulsham et al., 2008). Here, our goal is to leverage the recently measured human cost for making a gaze shift independently of image content (Thomas et al., 2022) for arbitrary saliency models so that we do not infer the rewards of image features from scratch. This allows arbitrary saliency models to improve their predictions of the next gaze target by incorporating the intrinsic costs of a gaze shift in human observers, which interact with the prioritization of image content.

4.2 Methods

Our general model is based on statistical decision theory and gaze sequences are viewed as reward-driven behavioral sequences, that can be described using a Markov Decision Process (MDP), similar to (Hoppe and Rothkopf, 2019; Jiang et al., 2016; Mathe and Sminchisescu, 2013). A scanpath is a sequence of gaze locations $\mathbf{x_0}, \mathbf{x_1}, \ldots, \mathbf{x_t}$ visited on an Image I through movement of the visual apparatus. Each of the visited gaze locations is the result of a decision for that particular location, following a policy $\pi(s) = \arg\max_{\mathbf{x_{t+1}}} Q(s, \mathbf{x_{t+1}})$, where $Q(s, \mathbf{x_{t+1}}) = \mathbb{E}\left[G_t \mid s, \mathbf{x_{t+1}}\right]$ are the Q-values, i.e. the expected discounted total future rewards $G_t = \sum_{i=1}^{N} \gamma^{i-1} r_{t+i}$ when switching gaze to a location $\mathbf{x_{t+1}}$ while being in state s. The state s summarizes relevant factors that contribute to the selection of the next action $\mathbf{x_{t+1}}$, γ is the discount factor, and N is the number of gaze shifts until the end of looking at a particular image. When exploring an image I, action selection is affected by past eye movements as well as the image, therefore $s = (I, \mathbf{x_0}, \ldots, \mathbf{x_t})$. For some tasks, s might also include further task-relevant features or it could represent a belief state. We argue that this is allowable here because saliency can be thought of as an

average over all possible states within all possible tasks as we will formulate in the following.

As has been shown repeatedly in the past, human action selection, in particular the generation of eye movements, is driven by multi-dimensional reward structures. However, the precise composition of the sources of rewards is usually unknown or not easy to measure. Here, we consider three components that have been shown to drive action selection: task-related reward, behavioral costs, and sequential effects related to the history of previous actions. In order to compute the state-action values $Q_{\text{task}}(s, \mathbf{x_{t+1}}) = \mathbb{E}\left[G \mid s, \mathbf{x_{t+1}}\right]$ in a specific task, we need to specify the rewards:

$$r(s, \mathbf{x_{t+1}}) = w_0 r_{\text{task}}(s, \mathbf{x_{t+1}}) + w_1 r_{\text{internal}}(s, \mathbf{x_{t+1}}) + w_2 r_{\text{fixation history}}(s, \mathbf{x_{t+1}}) \quad (4.1)$$

where $\mathbf{x_{t+1}}$ is a potential next eye movement location and r_{internal}, r_{task} and $r_{\text{fixation history}}$ are components contributing to the state-action value.

4.2.1 Saliency in the context of rewards

One dimension contributing to action selection is task-related reward. Eye movements have been shown to be carried out to lead to high rewards in their respective tasks, such as visual search (Najemnik and Geisler, 2005), image classification (Peterson and Eckstein, 2012), and can even be planned (Hoppe and Rothkopf, 2019). For free viewing of natural images, however, the reward function is difficult to obtain theoretically because the task instructions are highly ambiguous: "Just look around.". Here, we conjecture that saliency can be thought of as an average reward over all possible states within all possible tasks as we will formulate in the following. One possible approach is to formulate the task-related reward structure of free viewing as the result of marginalizing over all possible tasks:

$$\begin{aligned} r_{\text{free view}}(s, \mathbf{x_{t+1}}) &= \mathbb{E}_{\text{task}}\left[\mathbb{E}_{s_{\text{task}}}\left[r_{\text{task}}(I, s_{\text{task}}, \mathbf{x_{t+1}})\right]\right] \\ &= \mathbb{E}_{\text{task}}\left[\int_{s_{\text{task}}} r_{\text{task}}(I, s_{\text{task}}, \mathbf{x_{t+1}}) p(s_{\text{task}}) ds_{\text{task}}\right] \\ &= \sum_{\text{task}} \int_{s_{\text{task}}} r_{\text{task}}(I, s_{\text{task}}, \mathbf{x_{t+1}}) p(s_{\text{task}})\ ds_{\text{task}}\ p(\text{task}) \\ &\approx S(I, \mathbf{x_{t+1}}) \end{aligned} \quad (4.2)$$

where $r_{\text{task}}(I, s_{\text{task}}, \mathbf{x_{t+1}})$ denotes the reward when performing eye movement $\mathbf{x_{t+1}}$ in image I under a specific task while being in state s_{task}. The state s_{task} summarizes

all relevant information about the actions performed prior to the current decision for a specific task. The probability distribution over potential tasks $p(\text{task})$ weights the task-dependent reward according to how likely the task is. For example, information that is relevant for many visual tasks, e.g., faces, receives higher weights. Finally, $p(s_{\text{task}})$ is the probability distribution of the current state within a task, i.e. the action sequence (scanpath) prior to the current fixation and $S(I, \mathbf{x})$ is the saliency score.

In conclusion, we view the saliency of an image location during free viewing as the approximate average reward of that location across all possible tasks and all possible previous gaze shifts in that task.

4.2.2 Influence of past fixations

According to Equation 4.2 we can approximate the task related component to the reward using the predictions of a saliency model. However saliency models are time-invariant, depending only on the image.

Here, we propose an extension to overcome this problem and compute saliency models taking into account past actions. Our approach is based on the fact that the next fixation depends on all prior fixations. Since the exact nature of this relationship is unknown, we developed a model that quantifies the influence of a fixation within a gaze sequence on the selection of future fixation choices:

$$r_{\text{fixation history}}(s, \mathbf{x_{t+1}}) = r(\mathbf{x_0}, \ldots, \mathbf{x_{t-1}}, \mathbf{x_t}, \mathbf{x_{t+1}}) = \sum_{i=0}^{t} \phi_i \, \mathcal{N}(\mathbf{x_{t+1}}; \mathbf{x_i}, \mathbf{\Sigma}) \quad (4.3)$$

Positive values ($\phi_i > 0$) indicate that having visited location $\mathbf{x_i}$ at timestep i during the same scanpath increases the probability of targeting the next fixation to location $\mathbf{x_i}$. Negative values lead to reduced probabilities, therefore corresponding to an effect such as a spatial version of inhibition of return.

This reward can be conceptualized as the trade-off between exploration and exploitation, i.e. a reward for either parts of the state-space that have never been explored, or, if the environment can change over time, have not been explored recently (Sutton, 1990). Equivalently, this reward can be formulated as an exploration bonus. Therefore, this part of the reward structure encourages an agent to try long-ignored actions, i.e. visit locations that have not been visited yet or have not been visited in a long time. Since the exact nature of this relationship is yet to be understood, we estimated the parameters ϕ_i from the eye movement data. Note that we did not constrain the parameters to sum up to one, to allow both positive and negative values for already visited or not recently visited regions, in principle.

4.2.3 Oculomotor preference map

Saliency models commonly neglect the agent's effort expended in the actual action to gain visual information, although such internal costs influence gaze shifts (Hoppe and Rothkopf, 2016; Hoppe and Rothkopf, 2019). These costs and benefits have their origin in the effort to produce the movement, which includes cognitive costs such as deciding upon where to move next (Hoppe and Rothkopf, 2019) and when (Hoppe and Rothkopf, 2016). The oculomotor preferences were recently measured independently of image content in a psychophysical experiment involving a preference elicitation paradigm (Thomas et al., 2022). Subjects repeatedly chose between two visual targets by directing gaze to the preferred target. Eye movements were recorded using eye tracking. For each choice, three properties were manipulated for both targets: the distance to the current fixation location, the absolute direction to the target (e.g., left), and the angle relative to the last saccade. Using the decisions we inferred the value of each component and integrated them in an oculo-motor preference map. This map assigns behavioral costs to each possible gaze location dependent on the last two fixations:

$$\begin{aligned} r_{\text{internal}}(s, \mathbf{x_{t+1}}) &= r_{\text{internal}}(\mathbf{x_{t-1}}, \mathbf{x_t}, \mathbf{x_{t+1}}) \\ &= \psi_0 \left(\|\mathbf{x_{t+1}} - \mathbf{x_t}\| \right) \\ &+ \psi_1 \arccos \left(\frac{(\mathbf{x_{t+1}} - \mathbf{x_t}) \cdot (\mathbf{x_t} - \mathbf{x_{t-1}})}{\|\mathbf{x_{t+1}} - \mathbf{x_t}\| \|\mathbf{x_t} - \mathbf{x_{t-1}}\|} \right) \\ &+ \psi_2 \arccos \left(\frac{(\mathbf{x_{t+1}} - \mathbf{x_t}) \cdot [1 \quad 0]}{\|\mathbf{x_{t+1}} - \mathbf{x_t}\|} \right) \end{aligned} \tag{4.4}$$

Note that it is crucially important to estimate the human preferences for gaze shifts independently from image content instead of estimating probabilities of gaze shift parameters from free view data, because the observed gaze shifts depend both on image content and the preferences in such datasets.

4.2.4 Approximating the value map

We proposed three factors contributing to the final reward of an image location: task-related reward (Equation 4.2; approximated through saliency), fixation history (Equation 4.3) and the oculomotor costs (Equation 4.4). To account for the sequential nature of visual scanpaths we extend static saliency approaches using an additional reward component, which is an exploration part based on past fixations. Note however, that only the reward component associated with the free viewing task

is dependent on the image content whereas both the internal costs and the fixation history dependent part are independent of the image content.

By consistently formulating the components as rewards we can combine them to yield the desired reward function:

$$\begin{aligned} r(s, \mathbf{x_{t+1}}) &= w_0 r_{\text{free view}}(s, \mathbf{x_{t+1}}) + w_1 r_{\text{internal}}(\mathbf{x_{t-1}}, \mathbf{x_t}, \mathbf{x_{t+1}}) \\ &\quad + w_2 r_{\text{fixation history}}(\mathbf{x_0}, \ldots, \mathbf{x_{t-1}}, \mathbf{x_t}, \mathbf{x_{t+1}}) \\ &\approx w_0 S(I, \mathbf{x_{t+1}}) + w_1 \sum_{i \in \{0,1,2\}} \psi_i(\mathbf{x_{t-1}}, \mathbf{x_t}, \mathbf{x_{t+1}}) \\ &\quad + w_2 \sum_{i=0}^{t} \phi_i \mathcal{N}(\mathbf{x_{t+1}}; \mathbf{x_i}, \mathbf{\Sigma}) \end{aligned} \tag{4.5}$$

The parameters w_0, w_1, w_2 are linear weights and control the trade-off between task-related rewards, fixation history dependant rewards and internal costs and were estimated from the data. We set w_0 equal to 1, since the scale of our final value map does not matter and for the purpose of interpretability of the other parameters.

Computing the optimal policy in the MDP framework according to the reward function specified in Equation 4.5 would now require knowledge of the transition function, i.e. the state dependent gaze dynamics and their associated stochasticity. Similarly, knowledge of sensory uncertainties would be needed across all possible tasks in order to find the optimal gaze shift policy within the POMDP framework. Unfortunately, both these approaches are unfeasible. Instead, we use

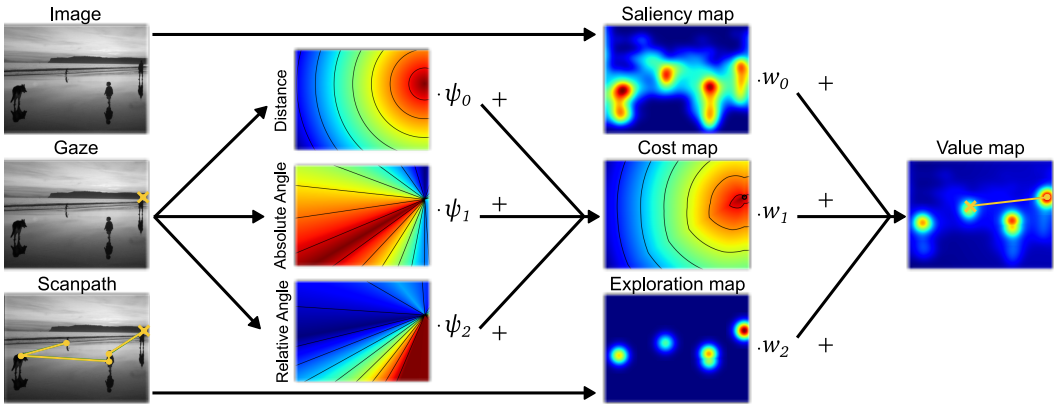


Fig. 4.1: Schematic of the algorithm. An arbitrary saliency map and the scanpath with the current gaze position are the input. Output is a value map, which integrates the saliency map, the recomputed map for the cost of gaze shifts, and the sequential history dependant map. Note that the original image is not an input to the algorithm.

Algorithm 1 Compute history dependent value map V at timestep t

Input: Arbitrary saliency map S from Image I, human scanpath $\mathbf{X} = \{\mathbf{x_0}, \mathbf{x_1}, ..., \mathbf{x_t}\}$
for all possible fixations $\mathbf{x}$ **do**
$C[\mathbf{x}] = \sum_{i \in \{0,1,2\}} \psi_i(\mathbf{x_{t-1}}, \mathbf{x_t}, \mathbf{x})$
$E[\mathbf{x}] = \sum_{i=0}^{t} \phi_i \mathcal{N}(\mathbf{x}; \mathbf{x_i}, \mathbf{\Sigma})$
$V[\mathbf{x}] = w_0 S[\mathbf{x}] + w_1 C[\mathbf{x}] + w_2 E[\mathbf{x}]$
end for
return V

the common approximation of selecting the optimal one-step look-ahead action, i.e. greedy approximation by selecting the action that maximizes the reward for a single subsequent gaze shift.

The approximate value map depends on the image (through S), on the location of the last fixation (through the internal costs) and on the entire sequence of past fixations (through the history dependant part). Crucially, as a consequence, the value map changes with every new fixation. The procedure of the computation of Q is illustrated in Figure 4.1 and examples of the respective maps for a succession of fixations is shown in Figure 4.2. Based on the approximate value map we can predict the future fixation locations from the policy π based on the value map $Q(s, \mathbf{x_{t+1}})$, see Algorithm 1.

First, to demonstrate the utility of our algorithm in improving the prediction of the next fixation of human observers for arbitrary saliency models, our model was implemented with four different underlying saliency models, which are currently among the ten best on the MIT/Tuebingen saliency benchmark (Kümmerer et al., n.d.) with respect to several evaluation metrics: DeepGaze II (Kümmerer et al., 2017), SAM-ResNet (Cornia et al., 2018), EML-NET (Jia and Bruce, 2020) and CASNet II (Fan et al., 2018).

The parameters describing the three components of the behavioral costs for gaze switches corresponding to internal motor and cognitive costs were recently estimated in a psychophysical experiment from eye movement data collected in a preference elicitation paradigm (Thomas et al., 2022). We collected a total of 70643 gaze shifts across 14 subjects following the experimental paradigm described in (Thomas et al., 2022). Values for the cost dimensions saccade amplitude, relative angle, and absolute angle were estimated using a random utility model (Train, 2009). The utility function was computed as the weighted sum of the individual dimensions.

To include the exploration map and calculate the resulting value map, the corresponding free parameters had to be estimated. Since we want to evaluate the

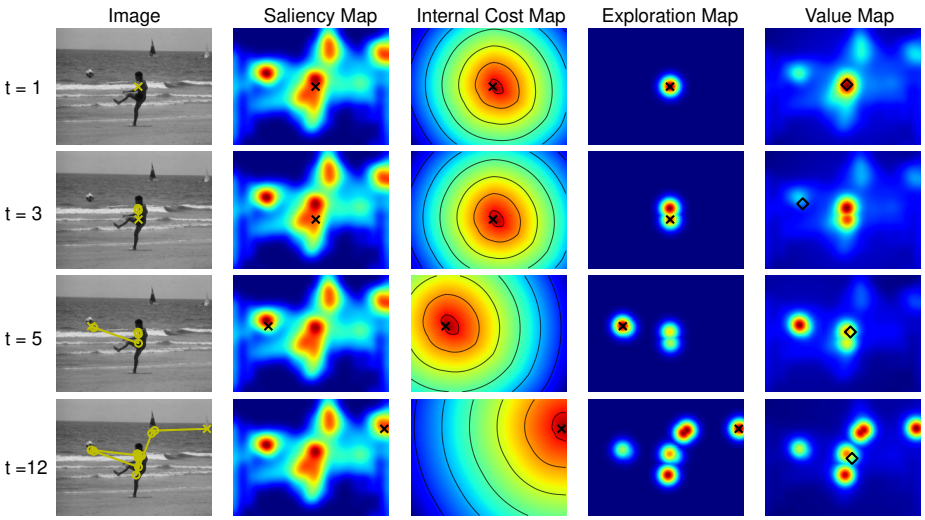


Fig. 4.2: Example predictions of the next fixation. Each row shows the original image together with the respective preceding scanpath together with the current i-th fixation marked with a cross. The corresponding saliency, cost, and exploration maps as well as the final value map are shown from left to right. The predicted fixation is shown together with the ground truth next fixation of the human observer marked with a diamond.

prediction of the next n fixations, we need to find a metric suitable for comparing individual fixations. We chose the Normalized Scanpath Saliency metric (Peters et al., 2005), which is definied as:

$$\mathrm{NSS}(S, \mathbf{x_0}, \ldots, \mathbf{x_T}) = \frac{1}{T} \sum_{i=0}^{T} S_Z(\mathbf{x_i}). \tag{4.6}$$

where T is the total amount of fixations for the current image. Here S_Z is the saliency map standardized by its mean μ_S and its standard deviation σ_S

$$S_Z = \frac{S - \mu_S}{\sigma_S} \tag{4.7}$$

Thus, the metric can be viewed as an average of the standardized saliency scores at the corresponding fixation locations. For more details on the metric score see e.g. (Bylinskii et al., 2019; Judd et al., 2012; Kümmerer et al., 2018). Since this method does not compare two continuous maps, but also considers the actual set of fixations in addition to the saliency map (Le Meur and Baccino, 2013), the metric is also suitable for our case of one-step or n-step ahead prediction. In this case, we do not average over the entire gaze sequence, but optimize the value map of our model so that there is as much mass as possible at the location of the next fixation.

More specific, a random subset of 10000 real human fixations was selected and the parameters of our model were optimized so that the value map could predict the single subsequent fixation as well as possible and thus maximizes the NSS score. All optimizations were done on the MIT1003 dataset (Judd et al., 2009). All selected fixations were between the third and eleventh gaze target in their respective sequence. This fixation interval was chosen so that at least two fixations had already been carried out by the human observers to be able to compute the cost map and because only about one percent of all fixation sequences contain more than ten fixations.

We estimated the exploration values ϕ_i, the covariance matrix $\mathbf{\Sigma}$ for the Exploration Map (Equation 4.3) and the weight parameters w_1,w_2 (Equation 4.5) through gradient based optimization. Note that the covariance matrix $\mathbf{\Sigma}$ was constrained to be a multiple of the identity matrix $\sigma^2\mathbf{I}$. We fixed the weight for the saliency map to one, so that the estimated parameters w_1,w_2 can be interpreted as quantifying the relative contributions of the costs for gaze switches and the history dependent reward relative to the saliency value. We used the Limited-memory BFGS-B algorithm (Byrd et al., 1995) given the NSS score as an objective function to be maximized. In addition, to meet the computational cost of the multidimensional problem, the images were reduced by a factor of ten in both dimensions using bilinear interpolation.

Computations were performed on a high performance computer cluster. All simulations were run on nodes with an Intel Xeon Processor E5-2680 v3 processor (2.5 GHz processor rate and 2.4 GB RAM). The results of the optimization for all parameters can be found in Table 4.1, where ϕ_i values where estimated for each model individually and Table 4.2 for the experiments with fixed ϕ_i values. The ϕ_i values for the second experiment where the weighted averages from the first one.

	w_1	w_2	σ	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8	ϕ_9	ϕ_{10}
DeepGaze II	0.35	2.89	34.16	1.74	2.09	2.02	2.46	3.32	3.38	4.74	5.22	5.22	4.37
SAM-ResNet	0.01	0.01	93.34	0.41	0.10	0.03	0.17	0.20	0.24	0.41	0.33	0.95	-2.17
EML-NET	0.10	0.48	18.30	0.16	0.79	0.43	0.75	1.08	1.10	1.45	-0.22	2.55	4.52
CASNet II	0.16	0.85	22.33	0.58	1.41	1.14	1.42	1.93	1.61	2.59	1.62	3.19	5.33

Tab. 4.1: Estimated model parameters with individual exploration values.

Additionally, the estimated exploration values, and thus the weighting of past fixations are shown for all four models over time in Figure 4.3. These intermediate results were used to derive general weights of past fixations independent of the particular model saliency. To this end, the estimated exploration values of the four models were averaged to be flexibly applied to arbitrary, new models. The resulting distribution is shown by the bold black curve in Figure 4.3. The same experiment was repeated for all saliency models and image databases, except that the weight

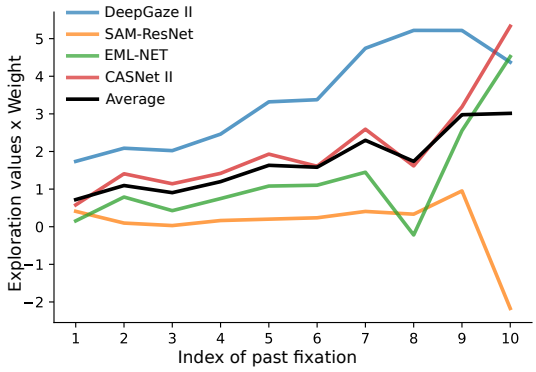


Fig. 4.3: Estimated exploration values for four different underlying saliency models (blue, orange, red, green) and the corresponding averaged curve (black). Note that the estimated ϕ_i from Table 4.1 have been multiplied here by their associated weight w_2 to visualize the actual influence of the exploration map.

parameters were no longer co-estimated, i.e. the previously determined values were used. The results of this experiment can be found in Table 4.2.

	w_1	w_2	σ	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8	ϕ_9	ϕ_{10}
DeepGaze II	0.35	1.99	33.63	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
SAM-ResNet	0.11	0.51	26.74	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
EML-NET	0.10	0.62	21.55	0.72	1.10	0.91	1.20	1.63	1.58	2.30	1.74	2.98	3.01
CASNet II	0.16	1.13	25.96	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
UNISAL	0.06	0.48	12.64	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Tab. 4.2: Estimated model parameters with fixed exploration values.

In addition, a new model was evaluated, which also belongs to the top evaluated saliency models on the MIT/Tuebingen benchmark - UNISAL (Droste et al., 2020). This was to investigate the degree to which the optimized model parameters would generalize from the four baseline models to a new model.

4.3 Results

We evaluated our method on three frequently used benchmarks, the MIT1003 (Judd et al., 2009), the OSIE (Xu et al., 2014) and the Toronto dataset (Bruce and Tsotsos, 2007). The MIT1003 and the OSIE dataset contain eye movements of 15 subjects during a three-second free viewing task on 1003 and 700 natural indoor and outdoor scenes, respectively. The Toronto dataset consits of 20 subjects during a four-second free viewing task on 120 color images of outdoor and indoor scenes.

4.3.1 One-step ahead predictions

We evaluated the one-step ahead predictions of our model with the NSS metric on the three datasets. Additionally we used a second metric, the Area under Curve (AUC) (see (Bylinskii et al., 2019; Judd et al., 2012; Kümmerer et al., 2018; Wilming et al., 2011) for details) for a second evaluation measurement, which was not considered during optimization. AUC is also a well known hybrid measure for evaluating fixation prediction and saliency models (Le Meur and Baccino, 2013), which can be understood as a binary classifier for whether pixels are fixated or not.

These two metrics can be used in evaluating the prediction of the next fixation, see (Kümmerer and Bethge, 2021). Other saliency metrics, like KL-divergence, Correlation Coefficient or Information gain are distribution based, so they assume the ground truth map to be a density and not a single fixation. Therefore, they cannot be used to evaluate models predicting the next gaze target or any other per fixation evaluation. Example images with best and worst NSS scores are provided in Figure 4.4 and 4.5.

Regarding scanpath prediction metrics (like ScanMatch or MultiMatch), we follow the evidence and argumentation from (Kümmerer and Bethge, 2021), arguing that it makes more sense to evaluate the capability of a model to predict the next fixation, which is exactly what saliency metrics do.

For evaluation, both metrics were calculated on all fixations of the three datasets above. For the first fixation, the model selects a target exclusively based on the saliency map as neither the internal cost nor an fixation history can contribute. To

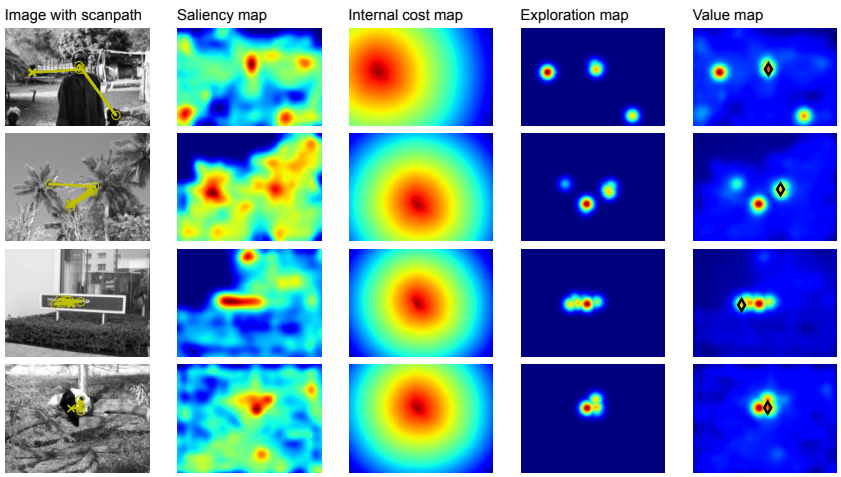


Fig. 4.4: Examples of predictions of the next fixations with highest NSS score.

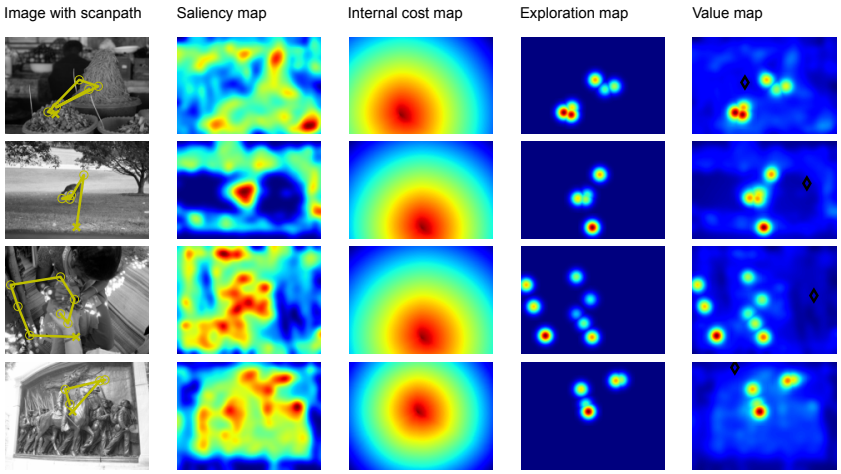


Fig. 4.5: Examples of predictions of the next fixations with lowest NSS score.

predict the second fixation, we assumed that the fixation prior to image onset was at the image's center. This is true for most experiments and this only influences the relative angle of the cost map.

The baseline saliency models were evaluated equivalently, but instead of using our dynamic value maps, the static history-independent maps were used. The results on the three different datasets with five different baseline models are shown in Table 4.3(a). We reached higher scores on all three datasets compared to all baseline models, even for the UNISAL model, which was not used in the estimation of the parameters of the exploration map. These results transferred in all cases to the AUC score, which had not been used in the optimization. Thus, subsequent fixations on the datasets are better predicted by our dynamic one-step ahead prediction maps compared to the static baseline saliency models. This provides evidence, that

Tab. 4.3: Evaluation results. AUC and NSS scores for the one-step and two-step ahead prediction of gaze targets based on sequential value maps compared to the respective saliency model's baseline

(a) One-step ahead predictions

	MIT 1003		OSIE		Toronto	
	AUC	NSS	AUC	NSS	AUC	NSS
DeepGaze II	0.844	1.506	0.906	1.867	0.497	-0.031
Our extension	**0.874**	**1.856**	**0.908**	**2.569**	**0.632**	**0.823**
SAM-ResNet	0.864	2.222	0.905	3.088	0.477	-0.105
Our extension	**0.881**	**2.323**	**0.917**	**3.315**	**0.639**	**0.706**
EML-NET	0.864	2.255	0.902	3.050	0.490	-0.073
Our extension	**0.882**	**2.329**	**0.919**	**3.330**	**0.638**	**0.656**
CASNet II	0.860	1.993	0.898	2.587	0.515	-0.059
Our extension	**0.879**	**2.155**	**0.915**	**3.003**	**0.684**	**1.033**
UNISAL	0.889	2.612	0.890	2.755	0.542	0.020
Our extension	**0.898**	**2.653**	**0.909**	**3.159**	**0.626**	**0.451**

(b) Two-step ahead predictions

MIT 1003		OSIE		Toronto	
AUC	NSS	AUC	NSS	AUC	NSS
0.844	1.506	0.906	1.867	0.497	-0.031
0.8554	**1.725**	0.888	**1.899**	**0.602**	**0.624**
0.864	2.222	0.905	3.088	0.477	-0.105
0.862	**2.301**	0.894	2.862	**0.598**	**0.535**
0.864	2.255	0.902	3.050	0.490	-0.073
0.869	**2.332**	**0.903**	2.897	**0.601**	**0.508**
0.860	1.993	0.898	2.587	0.515	-0.059
0.865	**2.098**	0.894	2.475	**0.608**	**0.598**
0.889	2.612	0.890	2.755	0.542	0.020
0.888	**2.667**	**0.893**	**2.841**	**0.604**	**0.371**

including the independently measured human cost function for carrying out eye movements improves predictions by saliency maps.

4.3.2 n-step ahead predictions

Although the free parameters of the model were optimized to maximize predictions of the single next fixation on the NSS score for the MIT1003 data set, we can test the performance of the n-step predictions. Table 4.3(b) and Table 4.4 report the results of the two-step and three-step predictions respectively. These results show, that the present model performs better consistently on the NSS score for both the MIT1003 and Toronto datasets across the second and third fixation predictions for all tested saliency models. Performance on the AUC score starts deteriorating for the prediction of the third fixation on the MIT1003 dataset but not the Toronto dataset. By comparison, both AUC and NSS scores are weaker already for the predictions of the second fixations on the OSIE dataset for all tested saliency models.

	MIT 1003		OSIE		Toronto	
	AUC	NSS	AUC	NSS	AUC	NSS
DeepGaze II	0.844	1.506	0.906	1.867	0.497	-0.031
Our extension	**0.850**	**1.662**	0.885	1.784	**0.565**	**0.394**
SAM-ResNet	0.864	2.222	0.905	3.088	0.477	-0.105
Our extension	0.857	**2.271**	0.886	2.813	**0.557**	**0.317**
EML-NET	0.864	2.255	0.902	3.050	0.490	-0.073
Our extension	0.864	**2.304**	0.867	2.851	**0.564**	**0.307**
CASNet II	0.860	1.993	0.898	2.587	0.515	-0.059
Our extension	**0.861**	**2.062**	0.888	2.406	**0.576**	**0.354**
UNISAL	0.889	2.612	0.890	2.755	0.542	0.020
Our extension	0.885	**2.646**	0.887	**2.786**	**0.582**	**0.243**

Tab. 4.4: Evaluation results. AUC and NSS scores for the three-step ahead prediction of gaze targets based on sequential value maps compared to the respective saliency model's baseline.

4.3.3 Influence of past fixations

In addition to predicting the next fixation in a gaze sequence, our model allows quantifying and explaining the relative influence of past fixations. Since the exploration values ϕ_i were not constrained, we are able to interpret them directly. Figure 4.3 shows the relative value of past fixations over time. Overall, the value of refixating an image location increases approximately linearly over time. This indicates that having visited location $\mathbf{x_i}$ i fixations ago during the same gaze sequence increases

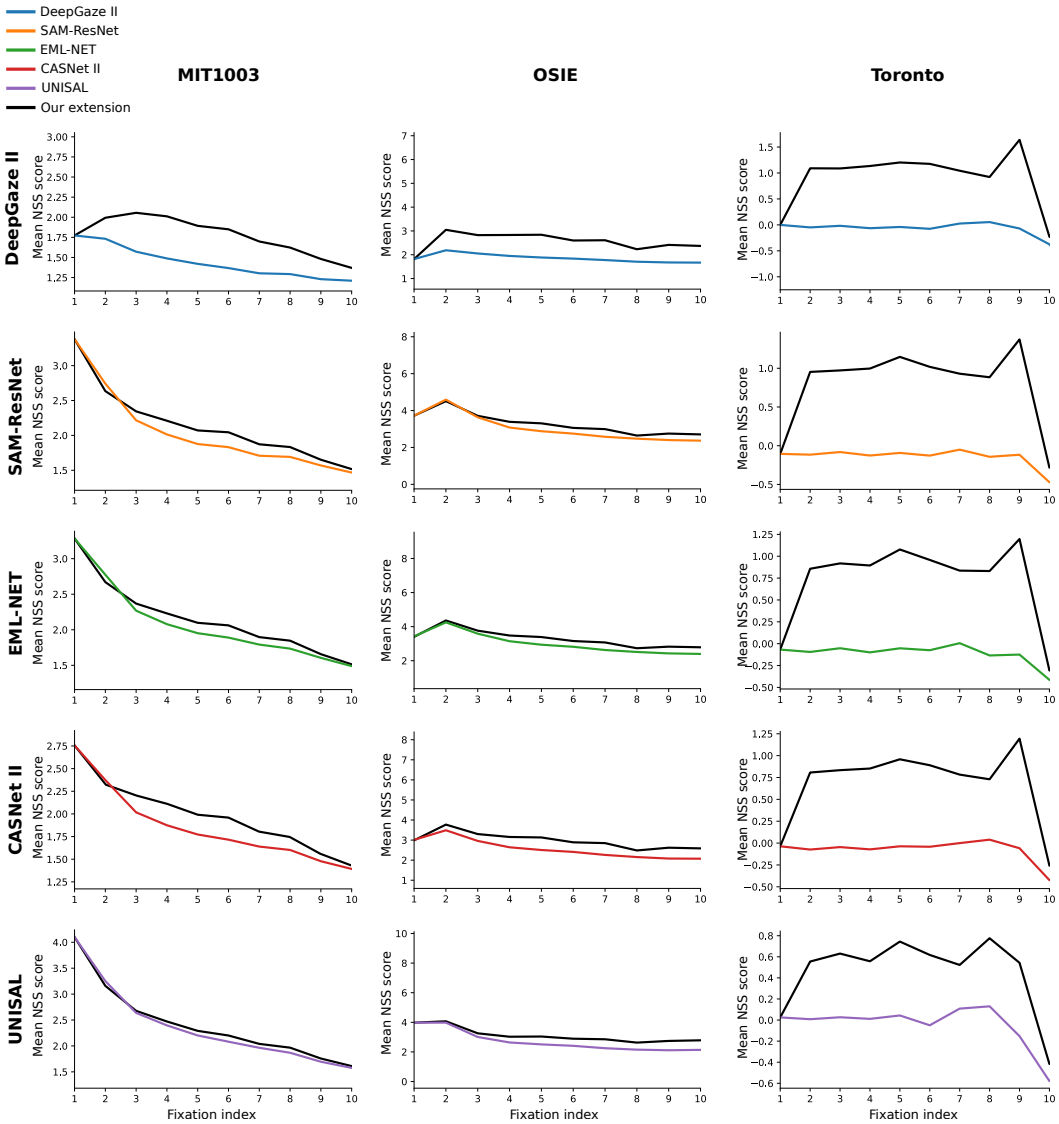


Fig. 4.6: NSS scores for the one-step ahead prediction depending on the ordinal position in the gaze sequence.

the probability of targeting the next fixation at location $\mathbf{x_i}$. This effect increases with increasing i, which means that fixation locations visited longer ago become more attractive for the observer.

For further analysis, we can quantify how well our predictions work for individual ordinal positions in the gaze sequence. For this, we selected all predictions by their ordinal position and averaged the NSS scores grouped by their fixation index. The progression of the goodness of the predictions can be seen in Figure 4.6 for all five models on all three datasets in comparison to the underlying baseline saliency models.

The differences in NSS scores can be seen in Figure 4.7. These results demonstrate, that the prediction accuracy is higher throughout the entire sequence up to the tenth gaze target, which was the last considered for almost all combinations of saliency models and image data sets. This further supports the usefulness and validity of the current approach.

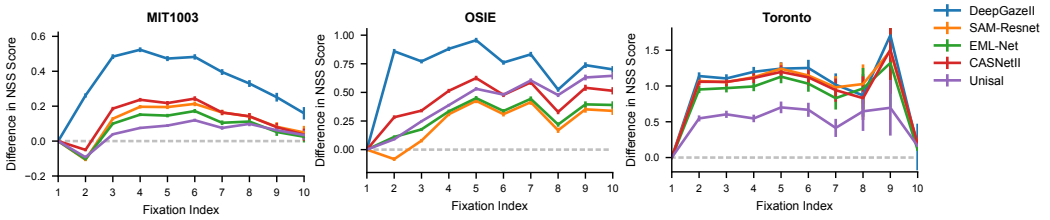


Fig. 4.7: Differences in the NSS scores between our dynamic value maps and the underlying static saliency maps. Positive values indicate that our dynamic model predicted the subsequent fixation better than the baseline model. The errorbars indicate ± standard error of the mean.

4.4 Discussion

In this chapter, we introduced a computational model utilizing arbitrary saliency maps for computing sequential value maps to predict the next gaze target in human fixation sequences (Kümmerer and Bethge, 2021). We conceptualized gaze sequences as sequential decision making within the framework of statistical decision theory, similar to previous approaches (Hoppe and Rothkopf, 2019; Jiang et al., 2016; Mathe and Sminchisescu, 2013). The model converts static saliency maps into a sequence of value maps, which incorporate saliency as a general conspicuity value across tasks, intrinsic human costs for gaze shifts, and a sequential history dependant reward. Given a saliency map of arbitrary origin and a sequence of previous gaze targets on an image, the model generates predictions of the next most likely fixation. The intrinsic preferences for gaze shifts used in the algorithm were recently estimated through a preference elicitation experiment independently of image content (Thomas et al., 2022) and the spatial and temporal parameters of the influence of fixation history were inferred based on the MIT1003 data set. Finally, the relative contributions of the three value maps were optimized on the same data set to maximize prediction of the next fixation. The algorithm can be applied to arbitrary saliency models and is available upon request from the authors.

The results demonstrate that the three components of the intrinsic costs for human gaze shifts (Thomas et al., 2022) are sufficient to improve predictions of subsequent gaze targets obtained from a saliency model.

These results are evidence that the common simplifying assumption that human scan paths are independent of behavioral preferences in gaze selection does not hold. Instead, the analysis of the distribution of preferred angles demonstrates, that image content and preferences in gaze shifts interact in non-trivial ways, a fact that has previously been demonstrated empirically (Foulsham et al., 2008). Although some previous approaches in scanpath modeling have acknowledged or implemented statistics of human gaze shifts (Boccignone and Ferraro, 2004; Le Meur et al., 2017; Tavakoli et al., 2013; Wang et al., 2011; Zanca et al., 2019), these were not measured independently of image content. The problem this gives rise to, is that the empirical statistics e.g. of saccade lengths measured in free viewing is the result of the preferences for gaze shifts and the distribution of image features. Thus, predictions of the next fixation need to be generated by taking the actual human costs of gaze shifts into account instead of the empirical distributions of gaze obtained from the databases, because the latter are the result of the interaction between image features and the costs for gaze shifts. Further research will evaluate, how frequently applied heuristics including inhibition of return, center bias, and proximity preference arise from the interactions of an observer with a visual scene building on concepts derived in this work.

5

Humans trade-off reward collection and information gathering according to probabilistic planning

We are so familiar with seeing, that it takes a leap of imagination to realise that there are problems to be solved.

— Richard Gregory, 1966

A fundamental component of active vision (Findlay and Gilchrist, 2003; Hayhoe and Ballard, 2005; Land and Tatler, 2009; Yarbus, 2013) in a dynamic and uncertain environment consists in acquiring task-relevant information to reduce uncertainty and achieve our goals. Such situations are ubiquitous in our everyday lives and include e.g. preparing tea (Land et al., 1999), making a sandwich (Hayhoe and Rothkopf, 2011), manipulating objects (Johansson et al., 2001), or catching balls (Diaz et al., 2013; Land and McLeod, 2000). Tasks become even more demanding, if one region of the visual environment needs to be looked at to achieve our long term task goal while other regions may contain information pertaining to developing threads, such as in team sports (Gou et al., 2022; Kredel et al., 2017) and when driving a car (Land and Lee, 1994; Sullivan et al., 2012). First and foremost, we want to reach our destination safely and must watch the road ahead carefully to accomplish this long term goal. However, there are potential threads developing in the context of traffic that require monitoring the rear-view mirror. Therefore we need to include this view from time to time to reduce uncertainty about potential sources of danger. Figure 5.1A illustrates this trade-off dilemma between watching the road to safely reach our destination in the future while at the same time having to monitor the surrounding traffic.

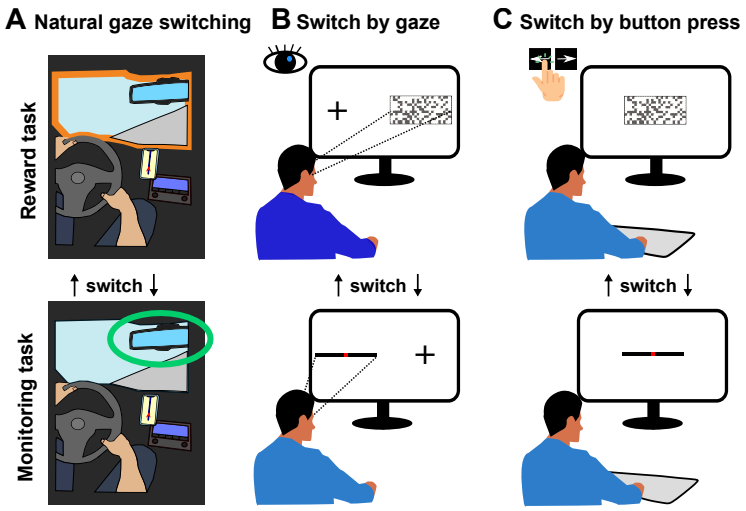


Fig. 5.1: The gaze-contingent paradigm. (A) Example for a natural gaze switching situation. The main task for the driver is to get to the destination safely and to monitor the road (monitoring task), but maybe she also wants to change the radio station or operate the navigation system (reward task). A trade-off must be found between the two tasks. (B) Eye movement condition. Subjects could switch between the tasks by gaze, while the task that was not fixated was hidden so that no information could be recorded extrafoveally. (C) Button press condition. Both tasks were in the same spatial location, but only one was visible at any one time. Subjects could switch between them by using the two arrow keys.

5.1 Related work

Adaptively balancing reward collection and uncertainty reduction involving dynamic and uncertain environments requires sequential decisions where to look next. It is therefore not sufficient to use habitual actions, i.e. always selecting the action that leads to the best immediate outcome (Dolan and Dayan, 2013; Redish, 2016). Instead, we have to plan our actions with regard to long-term consequences (Mattar and Lengyel, 2022) since our cognitive skills are computationally finite (Callaway et al., 2022; Griffiths et al., 2019). Computationally, this corresponds to planning (Kaelbling et al., 1998; Russell and Norvig, 2021; Sutton and Barto, 2018). For our visual apparatus this implies specifically that we have to actively navigate it to relevant parts of our visual environment (Findlay and Gilchrist, 2003). Therefore, vision can be understood as active sequential decision making (Hayhoe and Ballard, 2005) which allows us to flexibly adapt and change our behavior in respect to the visual environment and internal goals (Hoppe and Rothkopf, 2019).

Although planning and sequential decision making is an important ability and is involved in almost all of our actions (Mattar and Lengyel, 2022), planning has been studied in the context of problem solving (Huys et al., 2012; Opheusden

et al., 2023; Schrittwieser et al., 2020), but much less in the perceptual domain. Many studies and corresponding models do not consider long-term planning of eye movements, but are limited to maximising information or reward at the next time step, such as saliency (Borji et al., 2013; Itti and Koch, 2000; Kadner et al., 2023a), visual search (Hoppe and Rothkopf, 2019; Najemnik and Geisler, 2005), face recognition (Peterson and Eckstein, 2012), perceptual decision making (Padoa-Schioppa and Assad, 2006), economic choice (Gold, Shadlen, et al., 2007) and pattern classification (Yang et al., 2016). A first study that explored an extended planning horizon of two steps in advance and gives as a first indications that we need planning in active vision was provided by Hoppe et al. in 2019 (Hoppe and Rothkopf, 2019). However, while intrinsic factors of visual behaviors conceptualized as behavioral costs in the framework of decision-making have been investigated more carefully (Hoppe and Rothkopf, 2016; Hoppe and Rothkopf, 2019; Lisi et al., 2019; Petitet et al., 2021), the costs of saccadic gaze shifts have been captured with generic cost functions determined by the specific task under investigation. Importantly, it is not clear whether humans can plan their eye movements in dynamic and uncertain environments to balance reward collection and information gathering.

Here, we investigate the adaptability of temporal eye movement planning behavior with a gaze-contingent experiment. Like in real-world scenarios such as team sports and driving the uncertainties and rewards are unknown and the experimental setup allows to spatially separate the locations where the uncertainty of obtaining a task reward can be reduced and where the reward can be collected. To reduce uncertainty whether a reward can be collected subjects must observe a random walk. In addition to these dynamic uncertain observations we also vary the reward rates to be uncertain and dynamic with three different functional relationships (linear, quadratic and square root) to measure the adaptive behavior in response to these rates. We tested two different conditions of how to switch between the two tasks - either by an eye movement over the screen or by pressing a button on the keyboard. This design allows comparing the planning behavior to trade off reward collection and uncertainty reduction depending on the switching condition. We hypothesized, contrary to popular opinion, that human gaze switching behavior underlies high subjective internal costs, compared to simpler actions like the button press on a keyboard.

To understand these underlying mechanisms we adopt (Anderson, 1991; Gershman et al., 2015; Simon, 1955) and implement the computational model of planning. In order to implement these sequential decisions under uncertainty, we develop a probabilistic planning model using Partially Observable Markov Decision Processes (POMDPs). These have been used with eye movements in previous studies

(Butko and Movellan, 2008, 2009; Stankiewicz et al., 2006) and optimal control in detection tasks (Chebolu et al., 2022) and allow us to not only to quantitatively understand the switching behavior but inferring participants' behavioral switching costs and perceptual uncertainties. Our computational model results suggest that the subject internal costs for gaze switching are indeed higher than for the button presses. Our model implements components and concepts from the intertwined fields perception and action like perceptual uncertainty and internal costs and is able to predicts significant components of subjects' behavioral data. Therefore we conclude that temporal eye movement strategies align with probabilistic planning under uncertainty.

5.2 Methods

5.2.1 Participants and hardware

Eleven (5 females, 6 males) undergraduate or graduate students recruited at the Technical University of Darmstadt who received course credit participated in the experiment. All experimental procedures were carried out in accordance with the guidelines of the German Psychological Society and approved by the ethics committee of the Technical University of Darmstadt. Informed consent was obtained from all participants prior to carrying out the experiment. All participants had normal or corrected to normal vision and were seated approximately 62cm away from a 1920x1080 Samsung S24A300BL monitor fixated in a chinrest. Eye movement data were collected using a SMI iViewRED eye tracker with 60Hz sample rate and five point calibration.

5.2.2 Experimental design

The experimental setup was based on an gaze contingent visual reward harvesting paradigm consisting of two tasks - a reward collection and a monitoring task. Subjects could freely switch between the two different tasks and their goal was to collect as much reward as possible while not failing in the monitoring task.

The monitoring tasks consisted of a $0.25°$ long red square, which moved on a gaussian random walk ($\sigma = 0.2$) within a $10°$ long black bar. The reward collection task was a rectangle in which a certain proportion of pixels were coloured, corresponding to the current rate of reward. Figure 5.1B-C shows the experimental setup and visualizes the two different tasks.

The gaze contingent paradigm allows the subjects to only monitor one of the two tasks at a time by hiding the respective other. While subjects harvested reward by waiting on the reward collection side, the random walk continued moving and risked reaching to the bounds of the black bar, resulting in a loss of collected rewards. The task was now to collect as much reward as possible. The reward rate is visualized through the amount of pixels that are filled in the rectangle. The pixel pattern was chosen to prevent the subjects from getting an exact measurement of the current reward, but rather to get a perception of the functional relationship of the reward growth. If the reward collection task was chosen at a time step t a reward whose size is defined by a reward function $r(t)$ is added to a fictitious account. Subjects were able to secure and receive these accumulated rewards when they switched back to the random walk side. However, in the meantime of the reward harvesting the random walk has also continued moving and the subjects receive the collected rewards after the switch only if the random walk has not touched the boundaries while it was not fixated. If this was not the case and subjects stayed to long at the reward collection, they were informed of this by the screen briefly turning completely red.

We tested three different reward functions $r(t)$ to manipulate the rates subjects could harvest rewards. Besides a linear and quadratic, we also presented a square root relationship (see Figure 5.2). In addition to the different reward rates, we varied the possibility of switching between the two tasks in two different conditions. In the first condition, it was possible to switch back and forth between the two edges of the screen by making an eye movement. The task that was not fixed was hidden in order not to allow any extrafoveal conclusions. In order for the subjects to still be able to perform their fixations precisely, a fixation cross was displayed instead of the

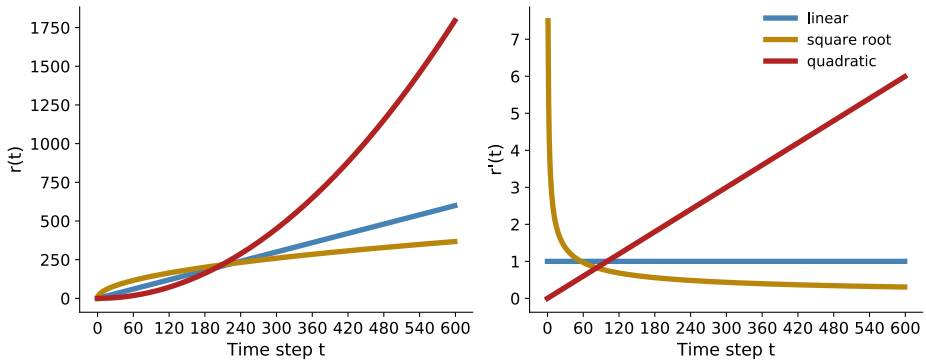


Fig. 5.2: The graphs of the three used reward functions (left side) and their corresponding derivatives (right side) plotted for the 600 time steps (10 seconds). Square Root: $r_1(t) = 15\sqrt{t}$, Linear: $r_2(t) = t$, Quadratic: $r_3(t) = 1/200t^2$

hidden task (Figure 5.1B). In the second condition, it was possible to switch with the arrow keys on the keyboard. Both tasks were in the middle and the active task was displayed. The subject therefore did not need to do any eye movement at all (Figure 5.1C).

The entire experiment for each participant consisted of 6 blocks (2 conditions and 3 reward functions). Each block in turn consisted of ten trials, all 60 seconds long. After each trial, subjects were shown their achieved score for this trial. The six blocks were displayed to each subject in a random order.

5.3 Results

Participants had to switch between two tasks, a reward collection task and a monitoring task. The two tasks were implemented through a gaze contingent experimental paradigm, i.e. subjects could not see both tasks at the same time but had to actively switch between them. While subjects could switch between the two tasks through directing their gaze to different regions on a computer monitor in the "gaze" condition (Figure 5.1B), they could instead use a button press in the "button" condition (Figure 5.1C) to display the respective other task at the center of the monitor through a button press. The reward collection task entailed looking at a display of random dots dynamically appearing and disappearing on the monitor. The rate at which dots appeared was proportional to the reward rate that could be collected. The total reward collected on a trial was determined by the time duration of looking at the reward task. Thus, the longer subjects looked at the reward task, the more reward they could collect. All subjects carried out the experiment with three different functional relationships of the reward rates. The consequence of different functional forms is that the reward rate either stays constant over the duration of a single look (linear), or the reward rate decreases with the duration of the look (square root), or increases (squared). The monitoring task involved monitoring a visual stimulus governed by a one dimensional spatial random walk. Subjects were told that if the random walk reached either one of the two symmetrically placed boundaries, the reward collected on that trial would be lost. Thus, participants had to balance collecting reward by engaging in the reward collection task and monitoring the random walk in the monitoring task so as not to loose all reward collected.

Because of this tasks design, the reward collection task and the monitoring task are both dynamic and uncertain (Figure 5.3A). Importantly, because looking longer at the reward collection task both increases the reward collected but also at the same time increases the uncertainty about the state of the random walk in the monitoring

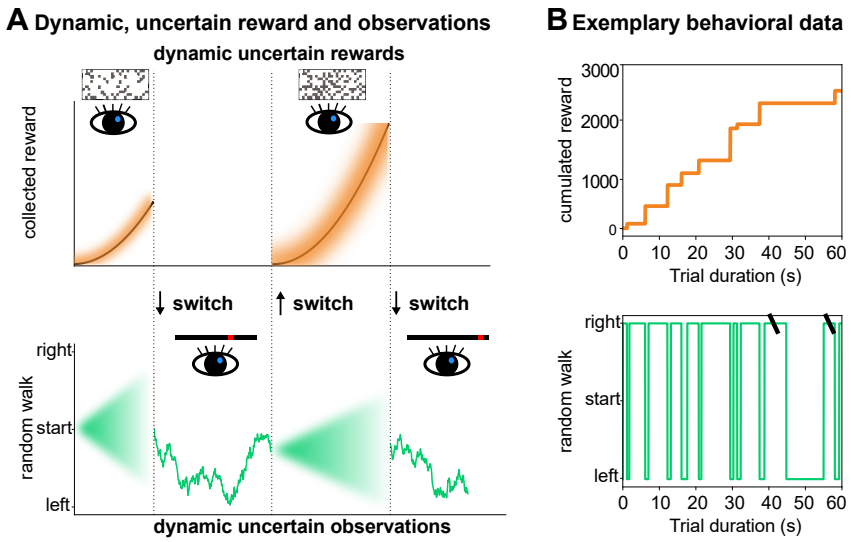


Fig. 5.3: (A) Trade-off between dynamic, uncertain observations and rewards. In addition to the uncertain observations of the dynamic motions of the random walk, the rewards are also modeled dynamically over different rates in our task. The exact reward magnitude cannot be perceptually captured in this task, so subjects have uncertainty about this as well. (B) Raw data for one trial with linear reward function in the eye movement condition of one subject. Attempts in which the subject has collected reward for too long and the random walk has reached the boundary are marked by black lines. No additional reward was credited after these switches.

task and therefore the risk of loosing all reward, an optimal policy does not consist of a fixed rate of switching between the two tasks. Instead, subjects need to maintain an internal dynamic belief about the state of the random walk and the associated risk of loosing all reward. Thus, computationally, to maximize the cumulative reward, switching decisions have to trade off subsequent uncertainty and rewards, which involves planning under uncertainty. Figure 5.3B shows the exemplary behaviour of one trial of one participant in the gaze switch condition for the linear reward function.

5.3.1 Behavioral Results

On average subjects reached 1684.44 points per trial and significantly more points in the button press (1816.56 points) than in the gaze switch condition (1552.30 points). The fewest points were collected with the quadratic reward rate, the most with the root function. The average points per trial separately for switching conditions and reward functions are shown in Figure 5.4A. A 3 by 2 repeated measurements ANOVA showed a statistically significant interaction between the switching condition and the reward functions on points per trial ($F(2, 20) = 6.42$, $p = 0.007$).

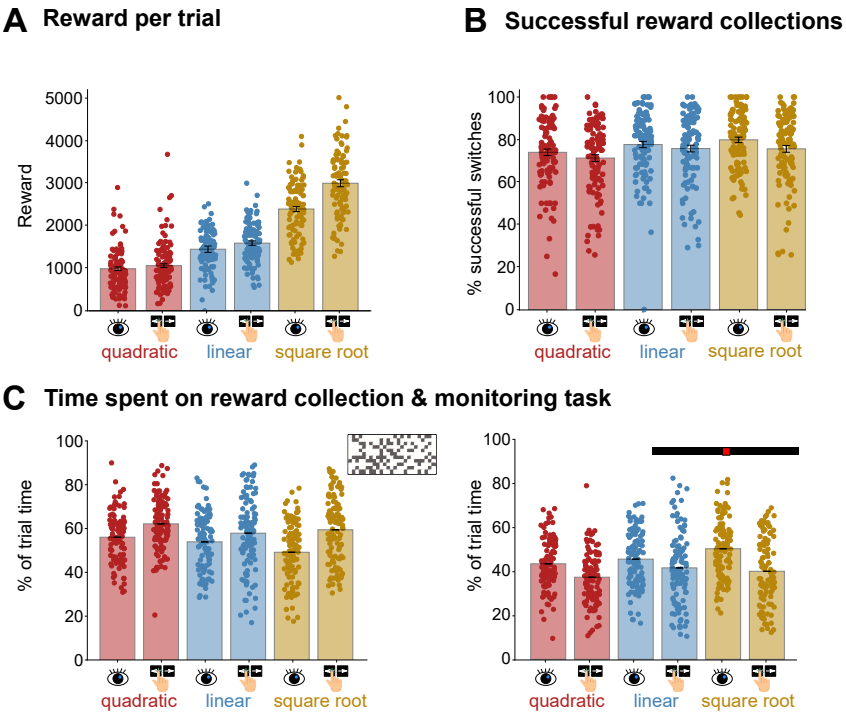


Fig. 5.4: Behavioral results for $n = 11$ human subjects. (A) Mean score collected per trial. (B) Mean amount of successful reward collections per trial. (C) Mean time spent on the reward collection and the monitoring task respectively. Results are split up by the switching conditions and the three reward rates. All errorbars correspond to the standard error of the mean.

Figure 5.4B shows the proportion of successful reward collection attempts. A total of 75.68 % percent of all reward collections were successful, i.e. the random walk did not reach the boundaries in the meantime while collecting rewards. There were no major differences between the square root (77.25 %) and the linear (76.69 %) reward function. The least successful performance was achieved with the quadratic reward function (72.63 %), and the gaze switch condition was in general more successful (77.15 %) than the key press condition (74.21 %). A linear mixed effects model of proportions as function of condition and reward function with subjects as random effects with random slopes and intercepts showed a statistically significant effect both of condition and reward function on proportion of successful reward collection attempts (see Table 5.1).

To exclude sequential effects across the duration of the experiment, we tested for significant effects of ordinal position of conditions. To test for effects of the sequential order of conditions and whether the switching behavior changed significantly over the duration of a condition we performed linear regressions. There was no significant effect of the position of a block (i.e. one of the six combinations of reward functions and switching conditions) and the switching perfor-

Tab. 5.1: Linear mixed effects model of proportions as function of condition and reward function with subjects as random effects with random slopes and intercepts.

Model:	MixedLM	Dependent Variable:	proportion
No. Observations:	660	Method:	REML
No. Groups:	11	Scale:	69.2133
Min. group size:	60	Log-Likelihood:	-2351.9951
Max. group size:	60	Converged:	Yes
Mean group size:	60.0		

	Coef.	Std.Err.	z	P> |z|	[0.025	0.975]
Intercept	86.066	2.223	38.714	0.000	81.708	90.423
eye tracking	3.561	0.648	5.499	0.000	2.292	4.831
func	-1.111	0.397	-2.802	0.005	-1.889	-0.334
Group Var	50.329	2.789				

mance ($R^2 = 0.00, F(1, 64) = 0.023, p = 0.88$, Figure 5.5A). To test whether the switching behavior changed significantly over the duration within condition, we performed multiple linear regressions on the ordinal position of the individual trial within a block on the percentage performance of successful switches, split by reward function and switching condition. We found no significant effects for within-condition performance to change significantly across the respective ten trials (see Figure 5.5B; eye-linear: $R^2 = 0.00, F(1, 108) = 0.039, p = 0.84$, eye-sqrt: $R^2 = 0.01, F(1, 108) = 0.73, p = 0.39$, eye-quadratic: $R^2 = 0.01, F(1, 108) = 1.86, p = 0.17$, button-linear: $R^2 = -0.02, F(1, 108) = 0.07, p = 0.80$, button-sqrt: $R^2 = 0.00, F(1, 108) = 0.40, p = 0.54$, button-quadratic: $R^2 = 0.04, F(1, 108) = 4.18, p = 0.05$).

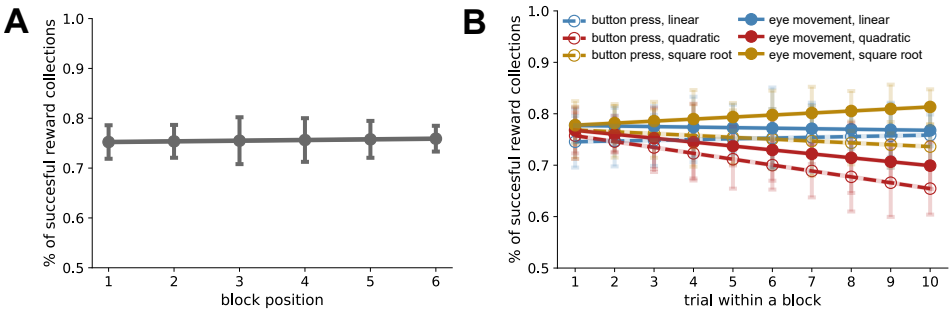


Fig. 5.5: Effects of sequential order and within-condition performance. (A) Linear regression over the ordinal position of the experimental block on the percentage performance of successful switches. (B) Linear regression on the ordinal position of the individual trial within a block on the percentage performance of successful switches, split by reward function and switching condition.

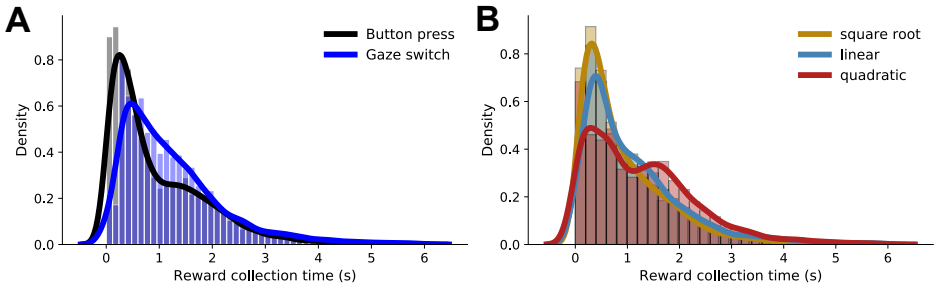


Fig. 5.6: Time spent on reward collection task. (A) Split up by reward function. (B) Split up by switching condition. Kernel Density Estimation was performed for better comparison using a Gaussian kernel.

Figure 5.6A shows the underlying distributions for the times spent on the reward collection task split up by the two switching conditions. The average time spent on the reward location is significantly longer in the eye movement condition ($\bar{t}_{\text{eye}} = 1.29s$) than in the key-press condition ($\bar{t}_{\text{key}} = 1.01s$) with $t(19373) = -18.42, p < 0.001$. This also holds for all pairwise comparisons of the reward functions between and within the two switching conditions (see Figure 5.4C). We performed independent 2-sample t-tests in order to compare the average reward collection times between conditions. In the button press condition the mean time on the reward side under the square root reward function ($\bar{t}_{\text{button/sqrt}} = 0.84$) is significantly shorter than with the linear ($\bar{t}_{\text{button/lin}} = 1.03$) with $t(8114) = -9.28, p < 10^{-20}$. Additionally the linear reward function results in a less significant reward collection time time comparing against the quadratic ($\bar{t}_{\text{button/quad}} = 1.20$) with $t(6906) = -6.47, p < 10^{-11}$. Same effects can be observed for the eye movement condition, where the square root function results lead to less significant reward collection times ($\bar{t}_{\text{eye/sqrt}} = 1.11$ against the linear function ($\bar{t}_{\text{eye/lin}} = 1.29$ with $t(5546) = -6.02, p < 10^{-10}$. The same applies to the comparison of the linear function versus the quadratic one ($\bar{t}_{\text{eye/quad}} = 1.50$) with $t(5095) = -6.16, p < 10^{-10}$. Thus, in general, our subjects linger longer on the reward collection task when they have to switch locations with eye movements. Figure 5.6B shows the comparison for the reward collection times between the different reward functions. The comparisons between the two switching conditions within a reward function X (e.g. $\bar{t}_{\text{button/X}} < \bar{t}_{\text{eye/X}}$) are also statistically significant, this holds for the linear ($t(6294) = -9.37, p < 10^{-21}$), the square root ($t(7368) = -12.58, p < 10^{-36}$) and the quadratic reward function ($t(5709) = -9.56, p < 10^{-22}$). Thus, in general, our subjects linger longer on the reward collection task when they have to switch locations with eye movements. This is a first indication of higher switching costs in the eye movement condition, as this is associated with statistically significant fewer switches overall. Moreover, they adapt their behavior given the various reward functions.

Figure 5.4A had shown the points achieved per trial divided according to the different change conditions and reward functions. Further analysis of the scores shows that across all trials and all subjects there was a minimum of 0 points and a maximum of 7157 points with a mean score of 1683.90 points (std. error = 116.54). The associated descriptive statistics are shown in Table 5.2, broken down by all subjects. The distribution of these scores are shown in Figure 5.7.

Tab. 5.2: Descriptive statistics for the trial scores over all subjects.

	Mean	Median	Std. Error	Min	Max
A	1564.22	1303.5	117.51	162	3969
B	1671.52	1512.0	123.53	248	7157
C	1473.85	1236.0	110.81	316	3442
D	1704.48	1703.5	92.80	273	3497
E	1847.57	1610.0	126.38	365	3966
F	1909.03	1680.0	175.40	172	6116
G	1527.87	1409.5	86.80	407	3125
H	1506.77	1303.5	95.61	416	3373
I	1930.58	1625.5	123.98	356	4831
J	1681.68	1473.5	107.24	518	3260
K	1711.23	1518.5	128.89	0	4289

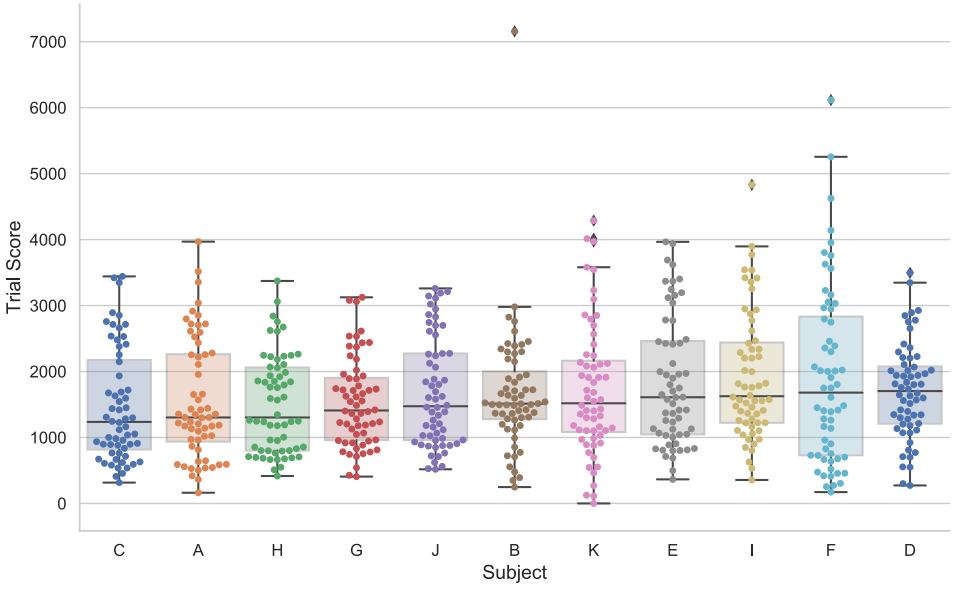


Fig. 5.7: Distribution of the trial scores per subject, in ascending order.

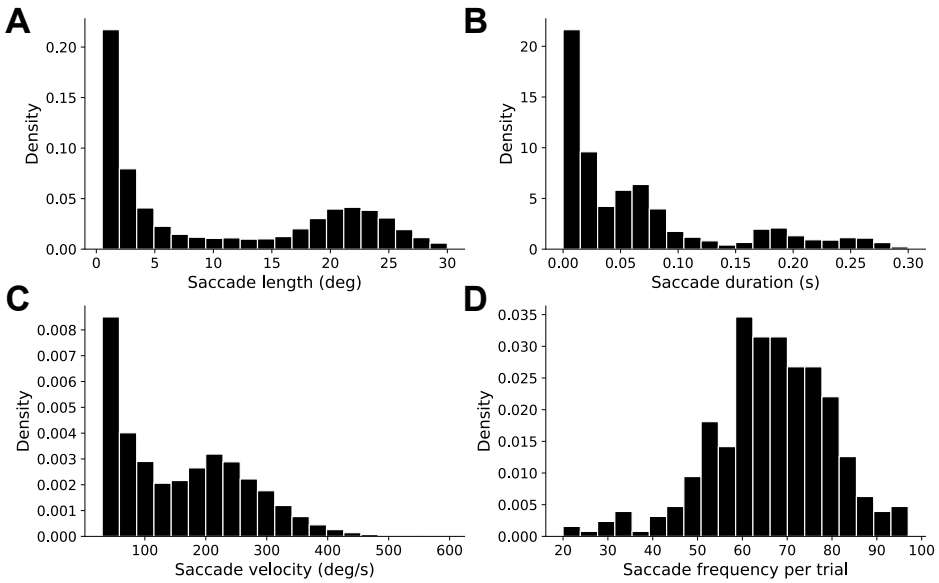


Fig. 5.8: Eye movement statistics for the gaze switch condition trials. (A) Distribution over saccade amplitudes. (B) Distribution over saccade durations. (C) Distribution over mean saccade amplitudes. (D) Distribution over saccade frequencies per trials.

Investigating eye movements, particularly saccadic behaviors, has emerged as a compelling approach to understanding various aspects of human cognitive and perceptual processes (Liversedge and Findlay, 2000; Rayner, 1998). Investigating eye movements, particularly saccadic behaviors, has emerged as a compelling approach to understanding various aspects of human cognitive and perceptual processes. Specific measures such as saccade amplitude (euclidean distance), frequency (number per unit time), duration (time), and velocity (speed) offer insights into cognitive load, decision-making processes, and skill expertise (Holmqvist et al., 2011). For example, saccade amplitude has been linked to the breadth of attentional focus (Hoffman and Subramaniam, 1995), saccade frequency and duration can explain the rate and depth of information processing (Salvucci and Goldberg, 2000). For the present data from the gaze switch condition, we analyzed these saccade statistics. Figure 5.8 shows the amplitude, duration, average velocity and frequency. No significant differences were found between the three different reward functions.

According to the hypothesis that participants plan their switching between the reward collection and monitoring tasks, the reward collection times should depend on the reward rate, i.e. the three different functional relationships, and the uncertainty in the monitoring task, which depends on the last observed position in the random walk. Figure 5.9A shows the mean reward collection times over all participants and trials, given the last observed position of the random walk for both conditions and all three reward rates, respectively. In addition to the pairwise comparisons

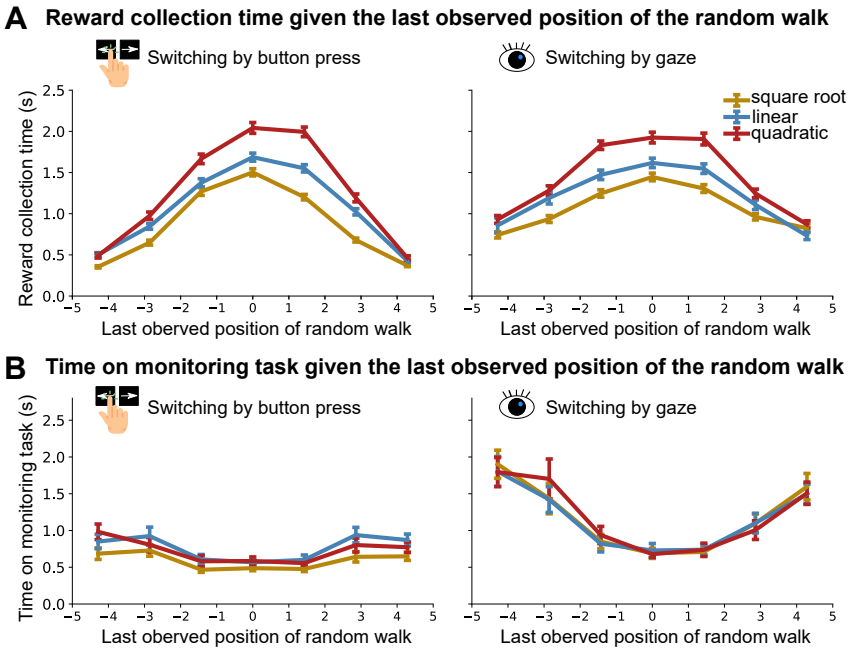


Fig. 5.9: (A) Mean reward collection time over all participants and trials, given the last observed position of the random walk split up by switching condition and reward rate and (B) Mean time on the monitoring task over all participants and trials, given the first observed position since fixating. Data was binned into seven equally spaced intervals. All errorbars correspond to the standard error of the mean.

of the average time spent on the reward task between switching condition and reward function, the last observed position of the random walk also seems to have an influence on behaviour. A three-way ANOVA revealed that there was a statistically significant interaction of reward function, switching condition and last observed of the random walk on reward collection time ($F(7, 19364) = 2.81$, $p = 0.09$). Simple main effects analysis showed that all three factors did have a statistically significant effect on the reward collection time ($p < .01$). Our hypothesis was confirmed, as the last observed position of the random walk significantly influenced the reward collection time, i.e. subjects' reward collection times were shorter the closer the random walk was to the boundary at the beginning of the reward collection phase and therefore the probability of a loss of reward was more probable.

Our hypothesis also predicts, that the reward collection times should be affected by the reward rate. In line with the hypothesis, subjects duration of reward collection was significantly shorter for the square root reward rate and significantly longer for the quadratic reward rate compared to the linear reward rate. All average collection times differ pairwise significantly based on independent two-sample t-tests. A further prediction of the hypothesis that human subjects can plan their gaze sequences is that their looking behavior in the monitoring task depends on the position of the

random walk. The reason is that, if people did not plan their switches, they should visit the location of the random walk only to reduce their uncertainty about the current position and switch back to the reward collection task. However, if they also show specific behavior for the fixation time at this location depending on the observed position of the random walk, this is a further indication of planning. Figure 5.9B shows the time spent in the monitoring task, i.e. the duration of the fixation on the random walk, as a function of where the random walk is at the beginning of the same fixation. The planning behavior is confirmed by the significant difference in times spend on the monitoring task as a function of last position of the monitoring task, which is more distinct in the eye movement condition, though we can not observe statistical significant difference between the three reward functions. A three-way ANOVA revealed that there was no statistically significant interaction of reward function, switching condition and last observed position of the random walk on the time spent on the monitoring task ($F(7, 19364) = 0.09$, $p = 0.7$). However the main effects of switching condition and last observed position and their interaction is statistically significant ($p < 0.001$).

5.3.2 A probabilistic planning model using POMDPs

The design of our experiment provides full access to the statistical structure of the task in terms of a generative model of the involved uncertainties and rewards. Crucially, this allows quantifying a particular strategy of adaptively switching between the two tasks in terms of the collected reward. This is the distinct advantage of using a controlled laboratory experiment compared to gaze measurements in team sports or driving, where it is difficult to capture uncertainties and rewards in everyday settings quantitatively.

The computational model is motivated by capturing the trade-off between the costs and benefits of (i) collecting rewards during the reward collection tasks, (ii) the costs of missing the random walk hitting either of the two boundaries in the monitoring task resulting in a loss of the accumulated rewards in a trial, and (iii) the intrinsic cost of switching between the two tasks, either using a gaze switch in the "gaze" condition or using a button press in the "button" condition. Formally, to model subjects' sequential decision-making behaviour under uncertainty, we developed a Partially Observable Markov Decision Process (POMDP) model. We can fully describe the task in such a model by formalizing it in terms of a set of possible experimental states S, a set of actions A, a state-transition function T, and a reward function R (Cassandra et al., 1994; Kaelbling et al., 1998). The implementation was done in Julia (Bezanson et al., 2017) with the *POMDPs* package (Egorov et al., 2017). To

find the optimal switching behavior for the planning under uncertainty problem formalized as a POMDP we used the so called QMDP algorithm (Littman et al., 1995). In the following we briefly describe the tupel (S, A, T, R).

State space The state space S is directly derived from the experimental setting and describes the true state of both the experiment and the subject at each moment in time. Accordingly, we can represent each situation unambiguously with four variables: (i) the current *position* of the random walk in the monitoring task, (ii) a binary variable indicating whether the random walk in the monitoring task has reached either one of the two bounds implying the loss of reward and therefore a *fail*, (iii) a binary variable defining which of the two *tasks* the subject is engaged in at the moment, and (iv) how much *time* has passed since the participant last switched between tasks.

Action space The model has an action space A comprising two different actions as the participant, i.e. whether to *stay* in the current task or alternatively to *switch* to the other task.

State transitions The state-transition function T describes the experimental dynamics and is identical to the generative model of the experiment, i.e. it describes how the variables representing the state of the experiment and the subject change over time, either due to the dynamic properties of the task design or due to the actions of the participant. Accordingly, the state-transition function describes the evolution of the *position* through the random walk of the monitoring task and sets the state to *fail* if the bound of the random walk is reached. The state transition function also updates the *time* depending on a switch according to *tasks*. Importantly, the model needs to incorporate known physiological constraints to be useful, specifically, the duration of a switching action. Based on previous research, switching by button press was modeled as having a duration of 250ms reaction time and 120ms execution time (Bjørklund, 1991; Dhakal et al., 2018; Silverman, 2010), while a gaze switch took 160ms reaction time and 90ms execution time based on a $30°$ long saccade (Baloh et al., 1975; Fischer and Ramsperger, 1984; Gezeck et al., 1997; Robinson, 1964). In addition, minimum and maximum dwell times for a task were implemented, motivated by the interquartile range of the behavioural data. Thus, the agent spends a minimum of 400 ms and a maximum of 5000 ms on one of the two tasks.

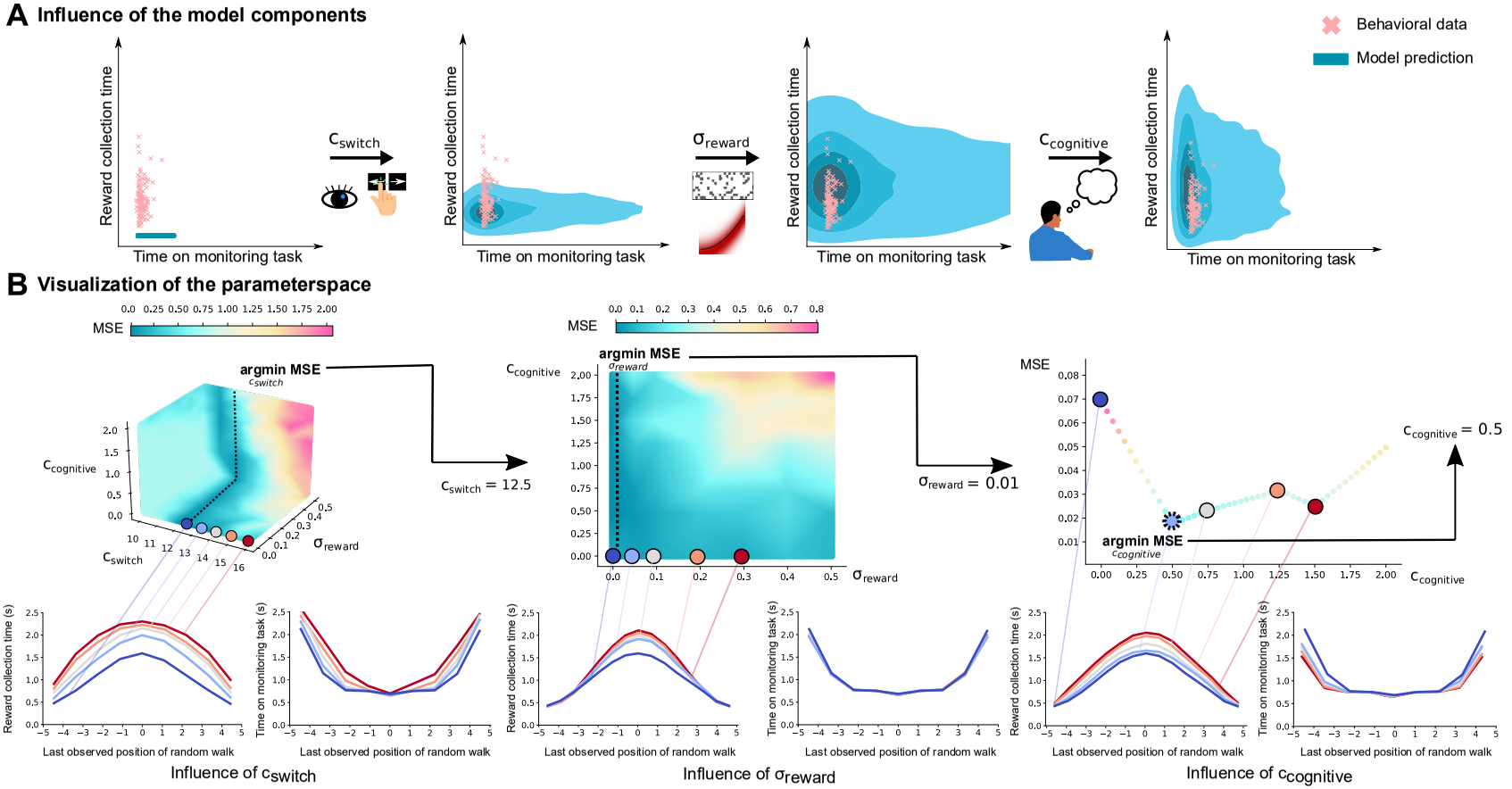


Fig. 5.10: (A) Influence of the model components on the distribution of task execution times using the square root function and the eye movement condition as an example. Crosses are the mean task execution durations of the subjects aggregated over individual trials. (B) The first illustration shows combinations of all three parameters and the MSE between model and human data. Towards the middle, the cost parameter for switching is fixed and the space is reduced by one dimension. To the right, the uncertainty about the reward rate is fixed and only the influence of the cognitive cost is considered. This gives the global minimum and the fit used of $c_{\text{switch}} = 12.5$, $\sigma_{\text{reward}} = 0.01$ and $c_{\text{cognitive}} = 0.5$ to best approximate the human data by our model. In the bottom row the influence of the three parameters on the simulated behaviour data in respect to the time spent in the reward collection and the monitoring task given the last observed position of the random walk. For each parameter five different values are simulated while the other two parameters have been kept constant.

Observations The model receives observations based on the states and its actions that are equivalent to those that are (partially) visible to the participant during the experiment. While one of the two tasks is being performed, its corresponding statistics are visible, although they are biased by perceptual uncertainty (e.g. time perception is modeled by a Weber fraction (Hoppe and Rothkopf, 2016; Mauk, Buonomano, et al., 2004)). About the unselected task nothing is available at the current time step, only the observations that were updated shortly before the last change during its execution.

Rewards and costs The reward function corresponds to the actual reward gain in the experiment, so the the model collects reward at the same rate as subjects in the experiment. In order to represent human behaviour and to take care of the uncertainty and the costs of actions, we introduce three free parameters (all reward was chosen to be scalar (Silver et al., 2021)):

First of all, costs for the two different actions. The parameter c_{switch} is introduced, which is applied to the action of changing between the two tasks. In addition, the cost factor $c_{\text{cognitive}}$, which represents the cognitive cost of processing information, should the agent decide to stay with a task and incorporate its information. The pixel pattern in the original experiment was chosen so that subjects do not have direct access to the exact amount of the current reward, but rather can see the functional relationship of the reward function over time. We also take this into account in our POMDP model, where we also introduce the uncertainty parameter σ_{reward}, which represents normally distributed noise over the reward received. Figure 5.10A visualizes the influence of the introduction of the individual components on the model behavior of task execution times.

Belief The model itself (as the subjects in the experiment) does not have knowledge about the current underlying states. The models posterior over the possible states given its observation (belief states) represent individuals' probabilistic knowledge about the environmental state, which is biased due to the perceptual uncertainty and the complexity of the visual world.

Simulating influence of costs, uncertainty and planning We introduced three free parameters in our model: c_{switch}, $c_{\text{cognitive}}$ and σ_{reward}. These can be freely chosen to simulate the model behaviour in our experimental task. Figure 5.10B visualises the parameter space. For different combinations of the three free parameters, the Mean Squared Error (MSE) to the actual human data is shown (here exemplary for

the linear reward function in the eye movement condition). For this purpose, the simulated model data was processed in the same way as the human data in Figure 5.9. We see an influence in the interaction of the three parameters and can identify regions where the minimum of the error for the fit to the human data is reached. For illustration, we fix the three parameters in Figure 5.10B one after the other from left to right in order to reduce the space by one dimension at a time and thus arrive at a total solution. We first find the argument of the minimum for $c_{\text{switch}} = 12.5$, then for $\sigma_{\text{reward}} = 0.01$ and finally for $c_{\text{cognitive}} = 0.5$ to obtain the solution that best fits the human data for the linear reward function in the eye movement condition.

The bottom row of Figure 5.10B shows the influence of the three parameters on the simulated behaviour data in respect to the time spent in the reward collection and the monitoring task given the last observed position of the random walk. For this purpose, five different values per parameter have been considered, while the other two have been kept constant. (1) Higher values for c_{switch} represent longer times both in the reward collection and monitoring task. To keep costs over the trial as low as possible, switching is only done when necessary, so longer times on tasks are preferred when costs are higher. (2) Higher values for σ_{reward} increase in times for the reward collection task (even steeper than for the first parameter) and no influence at all for the monitoring task. This can be explained by the fact that greater uncertainty about the actual reward leads to greedier behavior in the reward collection task and thus to longer execution times. However, if the random walk is selected, the uncertainty about the reward rate does not matter and thus does not lead to any change in behavior. (3) Higher parameter values for $c_{\text{cognitive}}$ lead to longer times in the reward collection task, but here contrary to shorter times in the monitoring task.

The question may arise whether planning is needed at all in the task, or whether the myopic/greedy model does not lead to the same results. POMDP models have a so-called discount factor γ which determines the relative importance of immediate versus future rewards. A higher value of γ means that the model places more importance on future rewards, while a lower value emphasizes immediate rewards. Figure 5.11A illustrates the influence of this discount factor on the model behavior. We find that the behavior can only be reproduced when the discount factor is close to one and thus includes almost the entire time horizon. However, the entire trial was 60 seconds long and one can rightly counter that humans are not able to plan and include everything over this entire period. Figure 5.11B now visualizes the behavior for different lengths of the entire trial. We cannot detect any difference in the behavior. This leads us to the conclusion that the optimal strategy (which people also use here) is an optimization over single reward-collecting actions. Therefore,

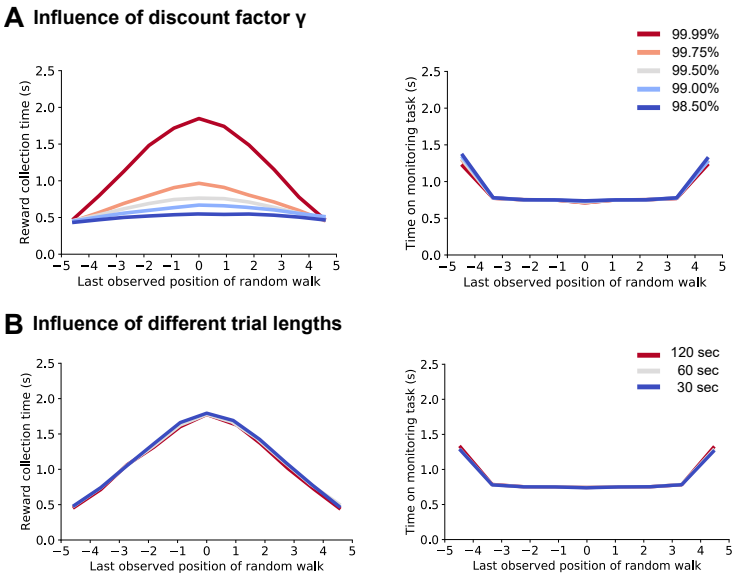


Fig. 5.11: (A) Influence of different discount factors γ on the switching behavior. (B) Influence of different trial length on the switching behavior.

people do not plan over the entire trial, but optimize the process of first looking at the random walk and then starting a single reward collection action.

Model results The results of the simulation and the inferred parameters are shown in Figure 5.12. The following points are particularly noteworthy: (i) The ratio between the costs for switching in the eye movement and the button press condition are constant for all three reward conditions ($c_{\text{switch (eye)}}/c_{\text{switch (button)}} \approx 1.2$). (ii) The ration between the costs for cognitive processing in the eye movement and the button press condition are constant for all three reward conditions ($c_{\text{cognitive (eye)}}/c_{\text{cognitive (button)}} \approx 0.2$). (iii) Within the three reward functions, the ratios of switching costs and cognitive costs are constant ($c_{\text{switch (eye)}}/c_{\text{cognitive (eye)}} \approx 25$ and $c_{\text{switch (button)}}/c_{\text{cognitive (button)}} \approx 4$). Not only can our model reproduce key aspects of people's behavioral data, but the cost parameters introduced show interpretation and are not just fitted values in a black-box model.

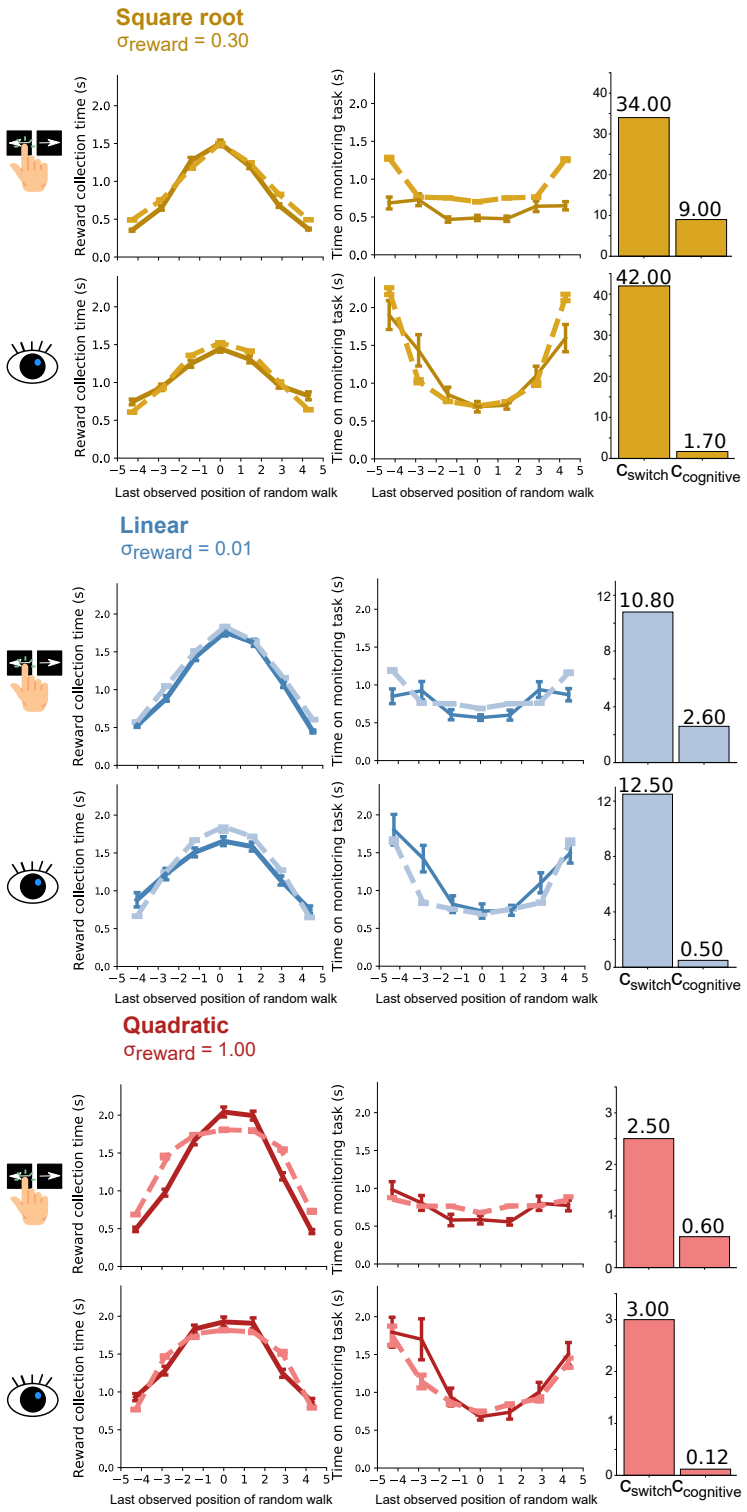


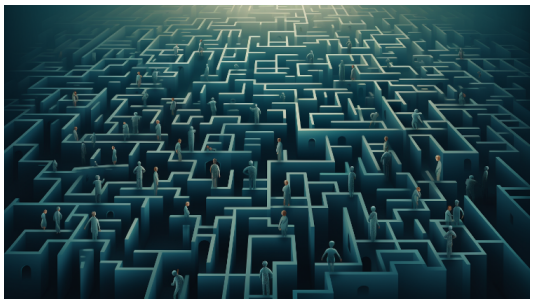
Fig. 5.12: Mean reward collection time and mean time on the monitoring task over all participants and trials, given the last observed position of the random walk split up by condition and reward rate. In addition to this, the inferred sizes for c_{switch}, $c_{\text{cognitive}}$ and σ_{reward} are shown. Dashed lines belong to the model of the eleven POMDP simulations. Errorbars correspond to the standard error of the mean. Data was binned into seven equally spaced intervals.

5.4 Discussion

The ability to plan is elementary for survival in a dynamic, uncertain environment; for instance, we can observe these abilities in birds and mammals (Doll et al., 2015; Jones et al., 2012; Schmidt et al., 2019). As a contrary example, the visual structure of an underwater environment is visually less complex and can not observe planning behavior in creatures like fish (MacIver et al., 2017; Mugan and MacIver, 2020). This study investigated human planning and how they trade off uncertainty reduction and collecting rewards in dynamic environments. In addition to dynamic uncertainties in the information gathering process, our experimental design also accounts for dynamic reward rates. This allows us to study the planning process and investigate the switching behavior by inferring participants' internal costs and uncertainties. We conceptualized the trade-off paradigm as sequential decision-making under uncertainty within the framework of Partially Observable Markov Decision Models and were able to quantify behavioral switching costs and perceptual uncertainties.

Our results suggest that humans can cope with temporal uncertainties in their observations and simultaneously adaptively balance dynamic, uncertain rewards to find trade-offs. Given that the three reward functions were tested differently, subjects could adjust the frequency of their changes and the execution time of each task to maximize their score. Given their slope rates, subjects collected reward longest on the quadratic reward function, followed by the linear and, finally, the root function. Moreover, they also adapted their strategies given the observations of the random walk. The further the walker moved towards the edges, the shorter the subsequent collection time was since the probability of an error became higher. Using our developed POMDP model, we could also investigate the planning behavior computationally. We concluded that subjects plan and optimize their strategy over the horizon of one collection attempt. We find empirical evidence that subjective internal costs for saccadic gaze shifts are higher than previous research suggested. In the context of active vision, these results have profound implications. They highlight the role of planning in the active navigation of our visual apparatus, enabling us to flexibly adapt our behavior in response to the visual environment and internal goals.

6 Human eye movements and their planning strategies in maze navigation



> *How beautiful the world would be if there were a procedure for moving through labyrinths.*
>
> **— Umberto Eco, 1980**

In order to achieve our goals and navigate the complexities of daily life, we need to plan our actions. In doing so, tasks can turn out to be difficult, as they offer several alternative courses of action or because the consequences of our actions have to be considered over a longer period of time. One way to study this planning behaviour in controlled environments is via maze-solving tasks. In contrast to simple binary decision tasks, maze-solving involves sequential multi-step decision-making, which requires planning in order to achieve long term goals.

Therefore, for analyzing decision-making and its underlying planning mechanisms, mazes have been of particular popularity in studying humans, animals and machines in various fields including Cognitive Science (Buecher et al., 2009; Kryven et al., 2022; Wu et al., 2016), Neuroscience (Alonso et al., 2020; Rosenberg et al., 2021), and Robotics (Aqel et al., 2017; Dang et al., 2010).

6.1 Related work

Real world naturalistic navigation tasks usually require the integration of internal and external cues, the execution of motor actions, and internal planning (Kessler et al., 2022). However, one major advantage of mazes as experimental environments is that they can be clearly defined and generated in terms of their topology (Kim and

Crawfis, 2018). Thus, we can use mazes to investigate planning mechanisms during navigation and are able to control the exact generative model of the environment. Elements of a maze's topology include for example, specific cell types (dead-ends, turns, crossings) and their overall distribution, but also the spatial arrangement of cells or the length of the solution path. In recent years, the possibilities for automated generation of mazes given various hyperparameters have been investigated and constantly developed further (Bellot et al., 2021; Kim and Crawfis, 2015). In human maze-solving the underlying topology characteristics have been proposed to influence behaviour in terms of solving time: Solution path length and the number of turns along the solution path have been shown to render a maze more complex and therefore increase solving time (Crowe et al., 2000). The number of alternatives influences the exploration behavior, where participants examine the task-relevant structure of the environment more thoroughly with an increasing amount of alternatives (Zhu et al., 2022).

Since planning and the underlying internal processes are not readily observable while decisions are made, eye movements have been used successfully as indication of ongoing cognitive processes (Hayhoe and Ballard, 2005; König et al., 2016; Spering, 2022), particularly in tasks in which spatial locations allow reducing uncertainty about task relevant quantities (Kaplan and Friston, 2018; Zhu et al., 2022). Furthermore, eye movements have been proposed to be closely related to the planning horizon since they can be understood as information sampling in the visual environment (Ma et al., 2021) and have indeed been shown to be planned ahead (Hoppe and Rothkopf, 2019). Furthermore, the planning horizon and strategy can differ dynamically within a task and between subjects (Carton et al., 2016; Tsividis et al., 2021) not the least by the simple fact that human scan paths are not independent of individual behavioral preferences in gaze selection (De Haas et al., 2019; Kadner et al., 2023a). Recent work suggests that humans balance depth and breadth searches (Vidal et al., 2022) and prune decision trees related to their plans when encountering large losses (Huys et al., 2012) in sequential decision making tasks different from mazes. However, whether and, if so, how humans potentially balance different strategies such as depth and breadth searches based on the availability of alternatives and the depth in search trees, is not known.

Eye movements have been investigated in previous studies involving mazes to gain insight into human maze-solving strategies, particularly related to planning. Previous studies suggest that gaze reflects a mental simulation process during maze-solving and is, therefore, reflective of the maze and its solution path structure (Crowe et al., 2000; Li et al., 2022; Zhu et al., 2022). (Zhao and Marquez, 2013) found that gaze patterns during maze-solving can be differentiated into those that subserve

exploration and those that aid in motor guidance. Although these studies indicate that gaze patterns represent planning behaviour in mazes, the exact influence of topological features on human planning strategy and the adopted planning horizon remains unknown.

In this study, we parametrically generated different mazes by controlling topological parameters, which influence the number of alternative paths and the length of the solution path. We analyzed participants' behavior by converting mazes into equivalent decision trees, allowing us to compute principled features quantifying the topology.

First, we confirm previous results showing that both the length of the solution path and the number of possible alternative routes impact performance, i.e. search time. Secondly, we look at the influence of the overall topology and the influence at the level of individual cells on solving time. Finally, by measuring subjects' eye movements, we can quantify participants' planning strategy by inferring depth and breadth features of their visual search. We find an effect of the number of alternate paths in a maze on the depth of their planning. Thus, with a larger number of alternatives, they plan less deeply, but keep the width of their planning constant, which hints at an adaptive planning strategy that adjusts the depth of planning with the number of paths that have to be considered. Such a strategy is computationally adequate, the memory resources for storing paths that have to be evaluated is limited.

6.2 Methods

6.2.1 Participants

16 subjects (9 female, 7 male; age M = 22, SD = 2.39) participated in the experiment. For seven subjects additional eye tracking data was recorded. All of the eye tracking subjects had normal or corrected to normal vision.

6.2.2 Apparatus

Mazes were presented on a 2560×1440 (559×335 mm) monitor. Participants were seated approximately 102 cm from the monitor using a chin rest. The mazes were presented centrally on an area of 1200×1200 pixels, such that the stimuli were displayed at a visual angle of ∼15°× ∼16°. Eye movement data were collected using

an Eyelink 1000 Plus eye tracker with a 35 mm lens, allowing online event parsing. We recorded the data from the participant's dominant eye determined before the start of the experiment. We performed a 9-point array calibration and validation procedure for each participant prior to the start of the experiment to ensure the accuracy of eye tracking data. The average validation accuracy was 0.3°, with all individual point measurements being under 0.94°.

6.2.3 Experimental Design

Maze Generation

We used the search-based procedural content generation (SBPCG) approach Kim and Crawfis, 2018 constrained to solution path length and the number of dead ends to generate mazes varying in their difficulty in terms of number of alternative paths and depths. This was done to elicit different human exploration patterns. All the generated mazes were *perfect*, so there is exactly one correct solution path that leads from the starting point to the goal. We opted for a 3×3 experimental design, choosing three different solution path lengths (short, medium, long) and three different levels of number of junctions (low, medium, high). The solution path lengths and their classification into the three levels are motivated by previous studies that observed variation in solution times when manipulating these variables Crowe et al., 2000. We ensured the different numbers of junctions by using different generation algorithms, which are known to generate different amounts of dead ends correlating with the number of junctions: Recursive Backtracking for low, Hunt-and-Kill for medium and Kruskal for a high amount of dead ends. Figure 6.1 shows the resulting nine mazes used and the respective subdivision given by solution path length and the number of junctions.

Procedure

Subjects were asked to solve the mazes as quickly as possible. To navigate through the maze and reach the goal, they could move from cell to cell using the arrow keys on a regular keyboard. The player position was indicated via a blue dot. Before the first move, the player position was placed at the start position. The goal was indicated via a red cross. After the first movement in the maze, the start position disappeared. Cells already visited were not marked, so participants needed to keep their path in memory in order to be able to trace back their steps.

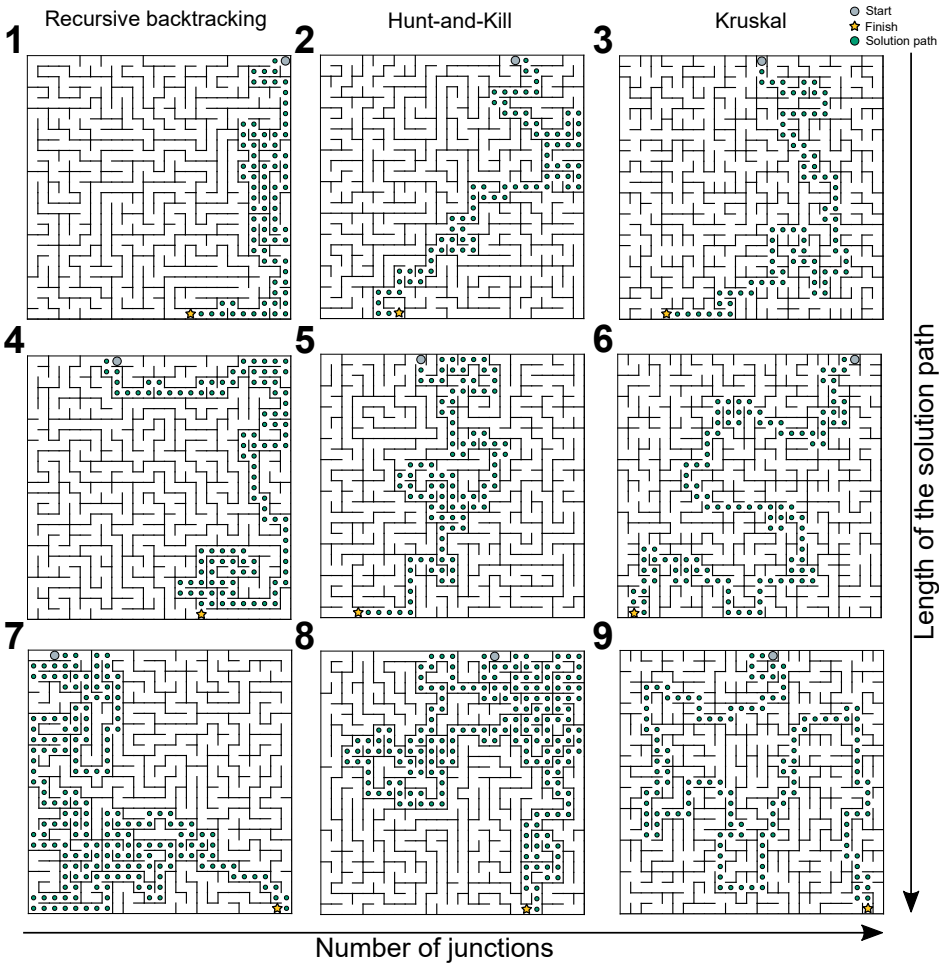


Fig. 6.1: The nine mazes used in the experiment. The columns contain different generation algorithms, which correlate with the number of junctions in the maze. In the rows, the length of the solution path increases from top to bottom. The starting point is shown as a grey point, the yellow star visualises the finish, and green dots visualise the solution path.

In order to exclude potential biases simply stemming from the geometric orientation of the solution path, each maze was shown twice, once as shown in Figure 6.1 with the solution path aligned in the horizontal direction and, a second time, rotated 90 degrees to the left.

In addition to the nine mazes, an easy-to-solve test maze was generated using the Recursive Division algorithm Reynolds, 2010. It was shown at the beginning of the experiment to familiarise participants with the control and task mechanics of the experiments. Afterward, the 9x2 experimental mazes were presented in random order. Finally, the rotated version of the test maze was shown. The test maze was excluded from further analyses.

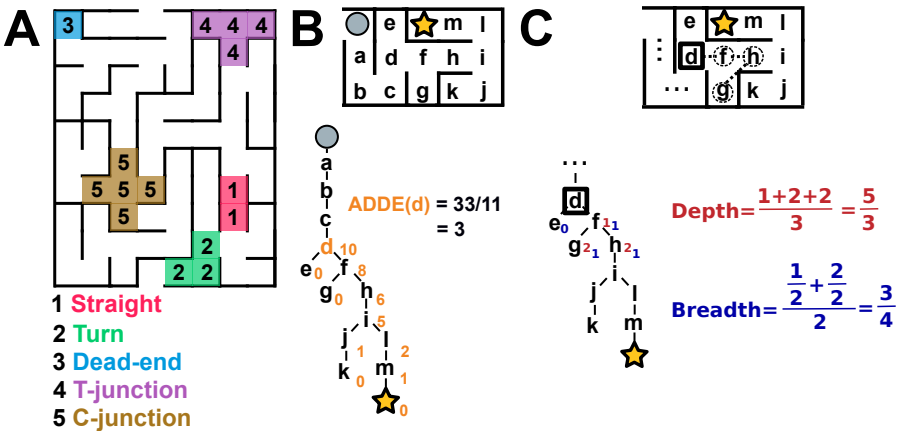


Fig. 6.2: Topological features of mazes. (A) Visualisation of the cell types (1) straights (2) turns (3) dead-ends (4) T-junctions, and (5) C-junctions. (B) The internal representation of a maze in a tree structure. For node d, the calculation of the Average Distance to Dead-Ends is exemplarily shown. (C) Exemplary calculation of depth and breadth. The player is in cell d and then fixates on cells f, h and g.

6.2.4 Metrics of topology

Following Kim and Crawfis, 2018, we define a set of metrics describing the global topology of a maze. First, this includes the different cell types that can occur within the maze, which are visualised in Figure 6.2A. The simplest cell types are dead-ends, which allow the player only to move backward, followed by straights and turns, which allow the player to move forward or backward. More complex cell types involve a higher number of choices, including so-called T-junctions for three alternatives and C-junctions for four, based on their shape. These cell types, together with their frequency and distribution especially along the solution path, build the basis for the topological properties of a labyrinth. Another important property to describe the complexity of the mazes is the length and branching of individual paths until they reach a dead end.

In order to quantify this property, we introduce the Average Distance to Dead-Ends (ADDE) measurement, computed for all cells in a maze. To calculate this property, each maze is converted into its equivalent tree structure with its unique identity. Then, starting from one cell, the mean distance to all dead ends is calculated. The value thus captures two important properties: First, it increases with the number of alternative paths given the current position since, for example, a C-junction can have one more path running into a dead end than a T-junction. Secondly, it increases with longer sub-trees after the given position, which makes it more difficult for the subjects to look at the entire set of paths, plan and remember the findings. The tree

representation of a maze and the exemplary calculation of the ADDE score for a given cell can be seen in Figure 6.2B.

6.2.5 Measuring planning behavior

We use the subjects' eye movements to gain insight into their planning behaviour. To do this, we look at the eye movements while the player stands still at a location in the maze and the participant explores the maze with gaze. For the planning process, we then compute two features of the search carried out by participants with there eyes, specifically a feature quantifying the depth of a search and a second feature quantifying the breath of a search.

The depth of a search episode is calculated as the average distance of the fixed cells to the current player position. The breadth, on the other hand, is the average coverage of possible paths per depth level up to the level of the fixation. Figure 6.2C shows an exemplary calculation of these two measures.

6.3 Results

6.3.1 Ruling out Confounders

Comparing the average solution times of the two types of rotations showed no major effect (two-sample t-test, t=0.261, p=0.795). In addition, the subjects were presented with the mazes (except for the first and last maze) in a random order. We performed a Linear Regression where the position of the maze within the experiment was compared to the corresponding solution time. The order of presentation had no significant effect on subjects' performance (F(1,16)=0.125, β=-0.49, p=0.072).

6.3.2 Performance

In order to quantify the performance of individual subjects, we consider the total time they took to solve the mazes. The mean solving time for the maze was 89.39 seconds ($\sigma_{\text{mazes}} = 59.17$). The descriptive statistics and differences for all nine mazes are shown in Table 6.1 in the Supplementary Information. On average, the fastest subject needed 41.77 seconds to solve one maze and the slowest 154.75 seconds ($\sigma_{\text{subjects}} = 54.76$). The descriptive statistics and differences for all sixteen participants are shown in Table 6.2 in the Supplementary Information.

Tab. 6.1: Descriptive statistics grouped for the nine mazes.

	Mean	Std	Min	Max
Maze 1	54.64	21.98	24.33	118.19
Maze 2	50.12	33.15	23.30	183.75
Maze 3	85.94	55.83	25.27	238.81
Maze 4	71.58	51.73	20.39	254.83
Maze 5	95.67	77.37	29.35	347.82
Maze 6	92.09	60.12	33.04	299.62
Maze 7	96.54	62.25	35.92	284.53
Maze 8	129.59	83.61	45.84	481.61
Maze 9	128.31	86.50	54.28	480.42

Tab. 6.2: Descriptive statistics grouped for the sixteen participants.

	Mean	Std	Min	Max
Participant 1	71.86	40.27	27.92	185.56
Participant 2	113.39	54.20	38.65	238.81
Participant 3	130.29	102.04	47.48	480.42
Participant 4	154.76	77.43	41.60	289.23
Participant 5	74.24	43.80	36.67	215.84
Participant 6	41.77	23.40	20.40	122.14
Participant 7	70.93	32.07	30.38	132.69
Participant 8	99.49	75.86	25.77	256.20
Participant 9	90.41	59.71	35.62	284.53
Participant 10	77.95	64.93	29.20	299.62
Participant 11	64.90	34.78	30.77	172.45
Participant 12	128.07	78.04	33.87	347.83
Participant 13	53.62	31.89	23.21	137.93
Participant 14	57.25	26.41	33.51	120.50
Participant 15	65.22	29.83	31.25	155.79
Participant 16	136.09	101.57	36.76	481.61

6.3.3 Why planning?

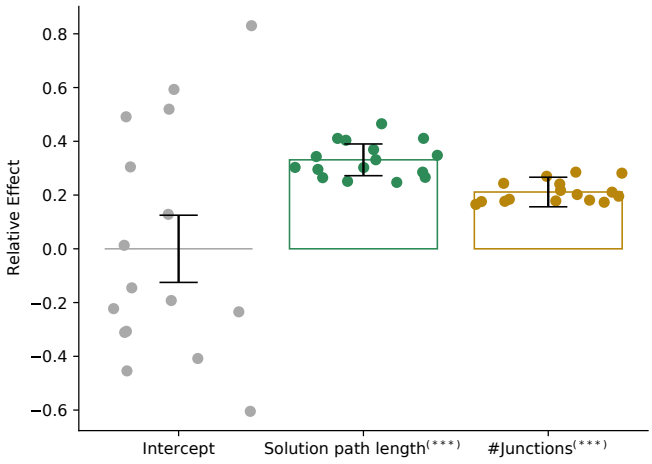


Fig. 6.3: Relative effects for the Linear Mixed Effects Model. Participants were chosen as random effects with random intercepts and slopes. All variables were z-scaled before. Errorbars correspond to the standard error of the mean.

To investigate the subjects' performance, given topological constraints at the global maze level, we used a Linear Mixed Effects Model. We chose the subjects as random effects, the solution path length and the number of junctions per path unit on the solution path as fixed effects. Increasing the solution path length should also increase the complexity of the maze since a larger planning horizon is needed to traverse down the path and rule alternatives out on the way. The number of junctions should also correlate with a higher complexity due to a higher number of alternatives that must be considered. The number of junctions is defined as alternative paths from the solution path, where C-junctions add two alternatives and T-junction one choice. We choose random slopes and intercepts for each subject to incorporate individual differences in personal performance. Equation 6.1 shows the resulting model for all the mazes.

$$\begin{aligned} \text{solving time} \sim\ & (1|\text{subject}) + \text{length} + (\text{length} - 1|\text{subject}) \\ & + \text{junctions} + (\text{junctions} - 1|\text{subject}) \end{aligned} \tag{6.1}$$

The relative effects are shown in Figure 6.3. Our results suggest that the solution path length and the density of junctions on the solution path make the mazes harder to solve for all participants and require increased planning effort.

6.3.4 Where to plan?

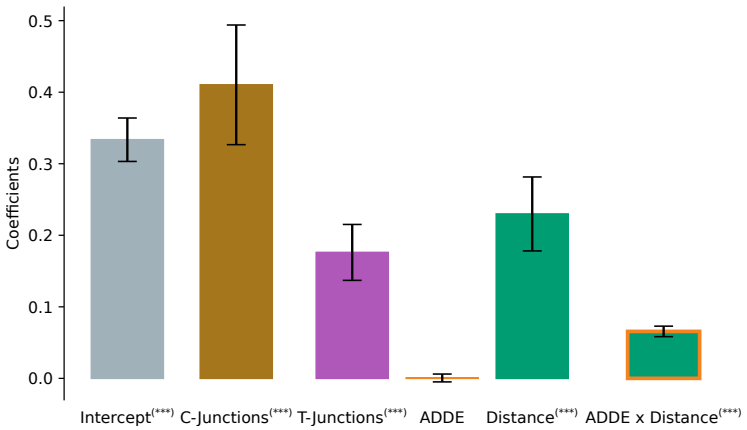


Fig. 6.4: Coefficients for the linear regression to explain different stopping times given the specific cell properties. The reference variable for the cell type is Turns/Straights with only two alternative decisions. Errorbars correspond to the standard error of the mean.

We differentiate between two different components in planning. Besides the regions the player looks at to find the solution path and the cognitive processes underlying it, it is first important to know at which points in the labyrinth this process started. Therefore, we first look for points where the player stops and continues to explore the maze using gaze. For this purpose, we calculated a linear regression over all subjects, looking for cell properties favoring a prolonged stop and, thus, a prolonged planning time. In addition to the cell types (baseline straight/turn with only two alternative paths), we also considered their position on the solution path (in distance cells to the target), their ADDE score, and the interaction effect of these. The fitted coefficients of the model ($F(5,1075)=157.8, p < 0.005$) are shown in Figure 6.4. We found longer waiting times the more alternative paths the current cell had (note that here the 2-alternative cell types Straights/Turns were taken as the baseline for the categorical variables). We also found a significant influence of the current position on the solution path. The further away the player is from the goal point, the longer he stays. This is consistent with our hypothesis that more planning and exploration is necessary at the beginning to find a possible solution path. The ADDE score alone has no significant influence but the interaction between the ADDE and the distance to the target does. This can be explained by the fact that longer and paths with increased branching have hardly any influence shortly before reaching the goal since the solution path has already been found. However, this has an enormous influence right at the beginning of solving a maze since possible solution path candidates have to be explored much longer during planning to see whether they are promising.

6.3.5 How to plan?

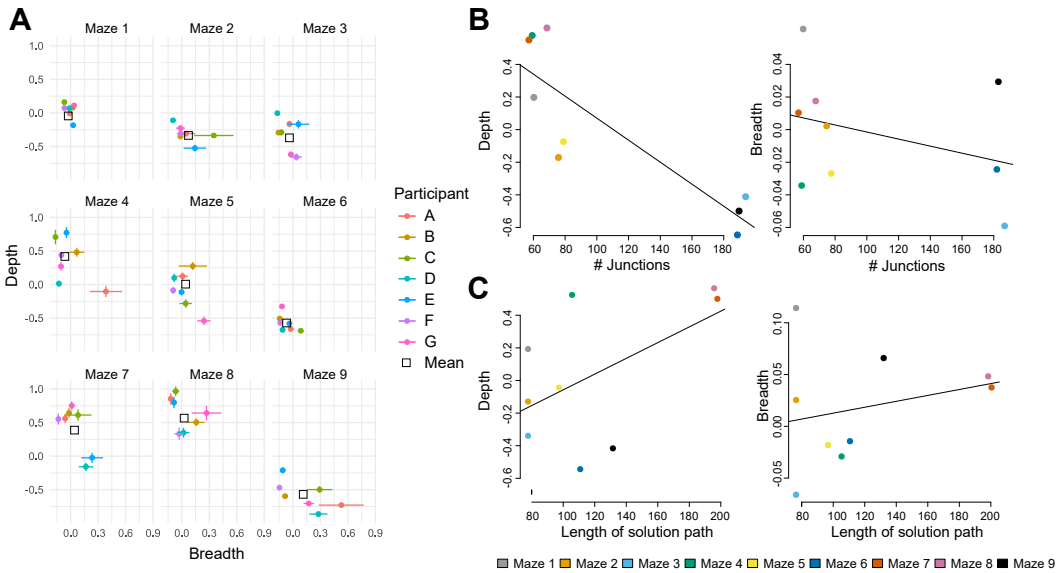


Fig. 6.5: Estimated breadth and depth values (z-scaled). (A) Mean breadth and width values for all participants split up into the nine mazes. Errorbars correspond to the standard error of the mean. All values are z-scaled. (B) Relationship between the number of junctions in a maze and the breadth and depth search values of participants (C) Relationship between the length of the solution path in a maze and the participants' depth and breadth search values.

To investigate people's planning strategy, we look at the breadth and depth values for their fixations paths as described and shown in Figure 6.2C. The calculated mean values for breadth and width for all participants in all nine mazes are displayed in 6.5A. These plots show that subjects show comparable planning strategies and despite some variability across participants, clusters can clearly be found for the depth and breadth values for each maze type.

The mazes were generated with different characteristics in the dimensions *number of junctions* and *length of the solution path* (see Figure 6.1). To investigate the influences of these two features on the depth and breadth values, we calculated linear regressions. Figure 6.5B visualises the influence of the number of junctions on breadth and depth. The plot demonstrates a significant impact of the amount of junctions in the maze on the depth of the search (linear regression, $R^2 = 0.74$,$p = 0.003$) but not on breadth value (linear regression, $R^2 = 0.12$,$p = 0.36$), i.e. with increasing number of junctions, participants maintained the breadth of their search, but reduced its depth.

For the influence of the length of the solution paths (Figure 6.5C) on breadth and depth, we see no influence, neither on the depth (linear regression, $R^2 = 0.28$,$p =$

0.14) nor the breadth values (linear regression, $R^2 = 0.06$,$p = 0.51$). Due to a longer solution path, the paths are less deep overall. People cannot anticipate these circumstances, though, because they cannot possibly know the length of the solution path in advance and thus don't know the depth of the alternatives. So, within the limits of their capacities, they keep both breadth and depth constant here.

6.4 Discussion

In this study, we investigated human planning strategies in maze-solving tasks. We generated different mazes according to fixed topological parameters and looked at both the performance and the planning behaviour by means of stopping times and measured eye movements of the subjects. We were able to reproduce and extend previously known results that suggested that performance decreases as mazes become more complex (that is, a higher amount and deeper branches within). At the cell level, we systematically found those locations where subjects needed more time to explore and plan their next steps. These cells were easily detectable, because participants spent significantly more time at these points in the maze without moving while gaze was moving along alternative paths.

To investigate the underlying planning process, we transformed the mazes into their equivalent decision trees to quantify the number of available alternative decisions at each point with the introduced ADDE measurement. The results show a strong effect of junctions on the stopping time. However, the depth of the possible search tree mainly had an effect at the beginning of the solution path, where subjects had to explore more extensively than towards the end when a solution path had been found with high probability. Evaluating the participant's eye movements within the decision tree, we were able to assign their internal planning to breadth or depth seeking. Subjects tended towards breadth planning, which suggests that they may be more likely to adopt a strategy that allows them to explore multiple options before committing to a specific solution. However, they are able to balance their strategy and change to more depth planning with decreasing number of junctions in a maze, e.g., adapting their strategy when they encounter situations with limited options. One explanation for this result is that the more alternative paths present themselves at junctions, i.e. the larger the branching in the equivalent search tree, the larger the memory requirement for evaluating all the different paths. Thus, by exploring each path to shallower degree, the burden on memory is reduced. These findings are consistent with Vidal et al., 2022, who designed the *BD apricot task*, an economic many-alternative task where subjects were asked to allocate finite search capacity to

sample the reward of the alternatives with the goal to choose the best one. Their results suggested that participants preferred deeper searches in environments where good outcomes where more likely. This is comparable to the increasing search depth of participants observed at maze locations with less alternative ways, i.e. deeper exploration of one of the few paths could lead to better results.

While the results of the current study indicate the utility of topological features in analyzing human planning strategies in maze solving tasks, there are a number of limitations which should be addressed in future studies. The introduced depth and breadth measures give a good indication of the quality of human planning behaviour, however, computational modeling of complete gaze sequences could give a more detailed illustration of the planning behaviour deployed. Since the subjects' did not explore each alternative in complete depth, they must use heuristics for pruning alternatives. An example of such a heuristic could be based on the angular direction from the current position to the target location, which could bias towards specific paths and alternatives in the planning process. To investigate the computational efficiency of the deployed trade-off between deeper and broader planning horizons, constrained breadth-first and depth-search searches could be compared with the strategic behaviour of the participants.

7

Increasing individuals' reading speed with an adaptive font model and Bayesian optimization

> *... optimum readability should always be foremost when developing a typeface.*
>
> — **Adrian Frutiger** (see Osterer and Stamm, 2014)

Language is arguably the most pervasive medium for exchanging knowledge between humans (Miller, 1991). But spoken language or abstract text need to be made visible in order to be read, e.g. in print or on screen. Traditionally, the transformation of text to the actually visible letters follows a complex process involving artistic and design considerations in the creation of typefaces, the selection of specific fonts, and many typographic decisions regarding the spatial placement and arrangement of letters and words (Carter et al., 2011; Zapf et al., 1991).

A fundamental question concerning the rendering of text is whether the way text is displayed affects how it is being read, processed, and understood. Scientifically, this question lies at the intersection of perceptual science and cognitive science. Numerous empirical investigations have measured how properties of written text influence its perception in reading both on paper (Bigelow, 2019; Landolt, 2019; Legge et al., 1985) and on electronic devices (Bernard et al., 2002; Bruijn et al., 1992; Dyson, 2004). Nevertheless, overall results of these empirical investigations on the relationship between reading speed or reading comprehension and font parameters are mixed and in part contradictory.

More recently, due to the pervasive use of electronic text that can be read on different devices under different viewing conditions, adaptive methods for font rendering and text display have been developed (André and Vatton, 1994; Microsoft, 2018; Sheedy et al., 2008). Moreover, fully parameterizable fonts intended to generate new fonts have been developed for multiple purposes (Arditi, 2004; Bragg et al., 2016; Devroye and McDougall, 1995; Hu, 1998). Some of these fonts are intended to help designing new fonts (Hu, 1998), some have been developed to enhance readability for individuals with low vision (Arditi, 2004), and others have been developed to increase readability on small electronic devices (Bragg et al., 2016).

Here, we close the loop by adaptively generating fonts and optimizing them progressively so as to increase individual's reading speed. The system is based on a generative font space, which is learnt in a data driven fashion using non-negative matrix factorization (NMF) of 25 classic fonts. We use this font space to generate new fonts and measure how a generated font affects reading speed on an individual user by user level. By utilizing Bayesian optimization, we sample new fonts from the generative font model to find those fonts that allow the reader to read texts faster. We demonstrate the feasibility of the approach and provide a user study showing that the found fonts indeed increase individual users' reading speed and that these fonts differ between individuals.

7.1 Related work

7.1.1 Typography, Font Design, and Reading

Traditionally, the development of a typeface is a complex creative process influenced by aesthetic, artistic, historic, economic, social, technological and other considerations (Carter et al., 2011). As such, designing of typefaces and typography underlie individual judgements, which may be highly variable between individuals and across time, see e.g. (Järlehed and Jaworski, 2015). Nevertheless, readability has often been a central goal in the creation of typefaces. What constitutes readability in this context and how to measure legibility and readability, has been debated extensively (Dale and Chall, 1949; York, 2008).

7.1.2 Typeface Features and Legibility Research

The scientific and empirical investigation of the relationship between how rendered text looks and how it is processed lies at the intersection of cognitive science and

perceptual science and is particularly relevant in psycholinguistics. The interested reader is refereed to reviews providing an overview of this broad body of work (Bigelow, 2019; Dyson, 2004; Landolt, 2019; Legge et al., 1985). Of particular relevance for the present study are investigations addressing how font features such as size, boldface, serifs, or different typefaces affect text readability. In this context, readability has been operationalized in different ways with subject's reading speed such as words per minute being the most common way. Note that within typography, the effect of a typeface's design and the glyph's shape is usually termed legibility and not readability.

Perceptual science has long investigated how features of typefaces and individual fonts affect letter perception e.g. with a series of experiments by Peterson and Tinker since the 1920ies (Paterson and Tinker, 1931). Since then, numerous features affecting readability of text have been investigated, including font size (Legge and Bigelow, 2011), serifs versus sans-serif typefaces (Arditi and Cho, 2005), lightness contrast (Blommaert and Timmers, 1987), color contrast (McLean, 1965) and many others. Perceptual science has also investigated how the spacing between adjacent letters influence letter recognition. Crowding describes the phenomenon that the perception of letters depends on their spatial separation as a function of foveal eccentricity (Bouma, 1970). It has been shown that crowding can account for the parafoveal reading rate (Pelli et al., 2007). Similarly, the effect of the vertical spacing of words on readability has been investigated (Chung, 2004). A number of studies have particularly focused on how readability of fonts changes with age (Darroch et al., 2005) and how to make fonts more readable for readers with medical conditions (Russell-Minda et al., 2007).

Because of the importance of text in electronic communication, the field of HCI has investigated the effect of fonts on readability particularly when rendering text to monitors and other electronic devices. Such studies include investigations of the effect of font size and type on glance reading (Dobres et al., 2016), measurement of user preferences and reading speed of different online fonts an sizes (Bernard et al., 2002; Boyarski et al., 1998), comparisons of the effect of font size and line spacing on online readability (Rello et al., 2016), quantifying the interaction of screen size and text layout (Bruijn et al., 1992), and investigating feature interaction such as color and typeface in on-screen displays (Garcia and Caldera, 1996).

Nevertheless, integrating all these results provides a mixed picture. Several studies concluded that increasing font size improves readability (Rello et al., 2016) while others reported that this holds only up to a critical print size (Chung et al., 1998). While early studies reported improvements in readability with increased font weight

(Luckiesh and Moss, 1940), other studies reported no significant effect (Bernard et al., 2013). While the presence of serifs did not significantly affect the speed of reading (Arditi and Cho, 2005), font type did but only modestly (Mansfield et al., 1996). Complicating the interpretation of some of these results is that most studies analyzed reading data across subjects, while some studies found more complex interactions of font features at the individual subject level (Korinth et al., 2020). Similarly, it remains unclear how reading text in print relates to reading on screen (Köpper et al., 2016). Finally, reading research has utilized many different reading tasks ranging from RSVP letter detection to comprehension of meaningful text (Salcedo et al., 1972).

7.1.3 Parametric, Adaptive, Generative, and Smart Fonts

The introduction of digital font has generated new techniques and challenges for adaptively rendering and setting text particularly on electronic devices, both from the point of view of the development of such algorithms (Knuth, 1999) and from the point of view of the typographer (Zapf et al., 1991). Parametric tools intended to help designers in developing new fonts have been introduced, e.g. (Shamir and Rappoport, 1998). Parametric techniques have been developed with the aim to increase readability, both in print (André and Vatton, 1994) as in rendering on screen, e.g. ClearType adjusts sub-pixel addressing in digital text to increase readability on small displays (Sheedy et al., 2008). Knuth's metafont goes a step further, as it can specify a font with one program per glyph, which in the case of the Computer Modern font family generates 72 fonts through variation of parameters. More recent developments include OpenType, which uses multiple glyph outline that are rendered to variable parametric fonts (Microsoft, 2018) as well as adaptive fonts, which have recently gained popularity in responsive design for the web (Latin, 2019). The ubiquity of hypertext has similarly lead to adaptive grid based layouts (Jacobs et al., 2003), which affects readability of text. Some of these systems are based on optimization algorithms that try to formulate and include font parameters (Hurst et al., 2009).

On the other hand, the development of computational and algorithmic art (Nake, 2009) and design (Fishwick, 2008) have extended the possibilities of creation in many fields. Parameterized generative font models allowing the parametric generation of new fonts have been created (Hu, 1998; McQueen III and Beausoleil, 1993). Other approaches aimed at describing characters as an assembly of structural elements resulting in a collection of parameterizable shape components (Hu and

Hersch, 2001). A prototype parametric font program called Font Tailor was specifically designed for users with visual impairments to customize a font to increase readability (Arditi, 2004). Dynamic fonts' (André and Borghi, 1989) shapes are instantiated every time a glyph is rendered allowing for parameterized variability, e.g. in emulating handwriting (Devroye and McDougall, 1995), and fluid typography (Brownie, 2007) changes glyphs' shape on screen thus blurring the lines between typography and animation. Some dynamic fonts have been design specifically to increase readability on small portable electronic devices (Bragg et al., 2016).

Our data driven approach to obtain a generative font space is most closely related to previous approaches of unsupervised learning of fonts (Campbell and Kautz, 2014) and recent systems based on Generative Adversarial Networks (GAN), e.g. (Azadi et al., 2018; Cha et al., 2020; Chen et al., 2019; Park et al., 2020; Roy et al., 2020). Differently from (Campbell and Kautz, 2014), which used a polyline representation of letters, we used a pixel based representation as the current study renders text only to a monitor up to a size of 40 pt.

7.1.4 Adaptive Design through Optimization in HCI

Recent work has explored the use of optimization methods from machine learning to evolve and adapt designs with explicit and implicit measures of users' preferences, which in general can be seen as a field belonging to interactive machine learning (Fails and Olsen, 2003). The difficulty lies in the availability or construction of an appropriate quantitative criterion that can be maximized and captures humans' explicit or implicit preferences. Note that this is different from the supervised learning of selecting fonts from many design examples (Zhao et al., 2018). One such area is the automatic design of interface layouts. Some approaches optimize layouts with respect to hand coded rules that are aimed at incorporating design criteria (Purvis et al., 2003). Off-line systems collect a large data set of user preferences or abilities and then approximate the user through a function, e.g. in ability-based user interfaces (Gajos et al., 2010; Gajos et al., 2008). More recent work trained a neural network to predict users' task performance from a previously collected data (Duan et al., 2020). Closed-loop adaptive systems, i.e. a system that parameterically changes to optimize some interaction criterion on-line while the user is actively interacting with it, are much rarer and have been used predominantly in the context of game design (Mahmud et al., 2014; Raffert et al., 2012; Zook et al., 2014).

7.1.5 Bayesian Optimization in HCI

One particularly attractive and powerful method for optimizing an unknown function is Bayesian optimization. It has a long tradition in many fields of research including optimal design of experiments (Shahriari et al., 2015) including closed-loop design of experiments (Lorenz et al., 2016). The power of Bayesian optimization lies in its ability to use statistical methods in modeling and efficiently optimizing "black-box" functions. Bayesian optimization has been used in several recent HCI systems both for open-loop optimization (Dudley et al., 2019) as well as in closed-loop systems, e.g. in the context of computer game design (Khajah et al., 2016). Current research also includes the application of Bayesian optimization in the field of computational design involving crowdsourcing (Koyama and Igarashi, 2018).

7.2 Methods

7.2.1 Preliminary User Study

In a preliminary user study we wanted to test whether reading speed was related to typefaces and font parameters at an individual subject's level. Seven subjects participated and read 250 texts each, whereby their reading speed was measured. All subjects had normal or corrected to normal vision and were German native speakers. They were seated at a distance of approximately 50 cm in front of a 24 inch monitor, which had a resolution of 1080 pixels horizontally. Participants were instructed to read the texts attentively and as quickly as possible, but only once in total.

The texts were *Tweets*, i.e. messages or status updates on the microblogging site *Twitter*, which are limited to 280 characters, written in German. The tweets were randomly downloaded from the Twitter pages of major news platforms and cleaned up from all mentions and hashtags. In order to ensure that subjects did understand

Tab. 7.1: Bayes Factors of the ANOVA relating font features and individuals' reading speed.

	1	2	3	4	5	6	7
Font Size	**4.34**	**20.22**	0.81	**6.08**	1.27	**12.69**	**40.80**
Font Name	**3.53**	0.46	1.15	0.54	0.16	0.55	0.01
Bold	1.20	0.69	**6.09**	0.14	**13.88**	0.10	0.03
Font Size + Bold	0.540	1.81	2.00	0.66	1.61	1.13	**6.17**
Font Size + Name	1.764	1.32	0.47	0.24	0.02	0.70	0.34
Font Name + Bold	0.472	0.08	2.92	0.08	0.18	0.08	0.00
Font Size + Name + Bold	0.244	0.21	1.26	0.03	0.03	0.09	0.13

the content of the texts, they were instructed to categorize each tweet after reading. To this end, eight categories were randomly displayed after subjects had completed reading a tweet. Participants were instructed to select as many categories applying to the content of the tweet as they thought appropriate. The tweets had been labelled previously independently by two of the authors of this study. A tweet was considered as labeled correctly by a participant if at least one correct label was selected for each text. A $5 \times 5 \times 2$ factorial design was chosen to investigate the influence of different fonts on reading speed. Each trial consisted of a combination of a different typeface (Arial, Cambria, Century Schoolbook, Helvetica and Times New Roman), a font size (10,20,30,40,50 pt) and whether the bold font was chosen. Each combination was presented exactly five times during the experiment, resulting in a total number of $5 \times 5 \times 2 \times 5 = 250$ trials. The measurement of reading time was started by subjects pressing a button, which also revealed a text, and stopped by a second button press, which also revealed the screen for the categorization. The time returned was used to compute the reading speed as the number of words read per minute. The distribution of all measured reading speeds can be found in Figure 7.7.

To analyze the influence of the various factors, a Bayesian ANOVA was carried out on a subject by subject basis using JASP (JASP Team, 2020). The Bayes Factors of the models of each subject are shown in Table 7.1. While there was substantial evidence in most subjects that font size influenced reading speed, font weight showed a substantial influence only in two subjects. For one subject, the analysis showed substantial evidence for an interaction between font size and bold face. These results provide evidence for individual differences in the magnitude and directions of the respective influencing factors. The results therefore provide a first indications that the factors influencing reading speed may differ on an individual subject basis.

7.2.2 Learning a Font Space

To generate fonts on a continuum, we first require a parametric generative font model. Here we used unsupervised learning to obtain a continuous font space.

Specifically, we chose non-negative matrix factorization (NMF) (Sra and Dhillon, 2006) as a method for dimensionality reduction of fonts. The idea behind this procedure is to approximate a matrix $\mathbf{X}$ with the product of two matrices $\mathbf{W}, \mathbf{H}$ so that the following relationship applies: $\mathbf{X} \simeq \mathbf{WH}$. This factorization is done under the constraint, that the approximation minimizes the Frobenius-Norm $||\mathbf{X} - \mathbf{WH}||_F$ and that all entries of $\mathbf{W}, \mathbf{H}$ are non-negative. The columns of $\mathbf{W}$ then represent

the dimensional features and **H** contains the weights to combine these features to reconstruct the rows in **X**. One advantage of the strictly positive components of NMF with respect to other methods such as Principal Component Analysis or Transformer Networks is that the basis functions can be thought of as printer's ink on paper and therefore as elements resembling actual glyphs. Another recent approach for font generation employs GANs, see e.g. (Azadi et al., 2018; Cha et al., 2020; Chen et al., 2019; Park et al., 2020; Roy et al., 2020). Even though this technique gives good results at the letter and word level, current implementations suffer from problems in alignment and kerning in continuous texts. Due to these difficulties, we were not able to employ GANs that could generate qualitatively satisfactory texts even though this approach will certainly be very promising in the future.

In order to cover a basic set of fonts including fonts with serifs and sans-serif fonts, we selected a list of classic and popular typefaces (Spooner, 2009), containing a total of 25 *classical* fonts. The typeface names are shown in their respective font in Figure 7.1. To perform the NMF, we generated a grayscale image of size 2375×51 pixels containing all 26 letters in upper and lower case, the numbers from zero to nine, the German umlauts, brackets, question and exclamation mark, dot, comma, hyphen, colon, semicolon, slash and quotation marks. Such an image was generated for each individual font using a font-size of 40 pt. Letters were arranged side by side. The image data was then concatenated together with information about the alignment of each glyph in the font, obtained from the respective information in the TrueType-file. The resulting font vectors form the rows of our matrix **X**, on which we performed the NMF.

To choose the appropriate number of dimensions, cross-validation was performed. As recommended by Kanagal & Sindhwani (Kanagal and Sindhwani, 2010), we

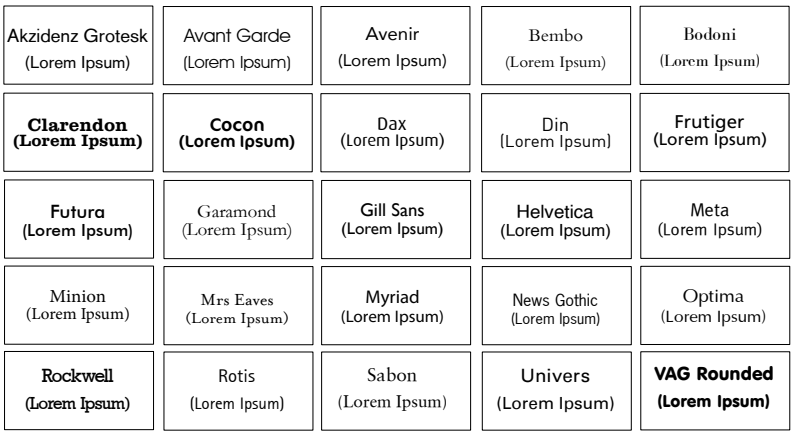


Fig. 7.1: The 25 baseline typefaces used in learning of a font space with NMF.

masked our data using Wold holdouts and fitted the NMF to the rest of the data using weighted NMF with a binary weight matrix. The Wold holdouts are created by splitting the matrix into 16 blocks and randomly selecting four blocks, one for each row, to hold out. Ten random Wold holdouts where created and for each we calculated the reconstruction error when choosing between one and five NMF components.

The mean reconstruction error showed that a low testing error can be obtained between one and three NMF components and that the model starts overfitting the data for four or more components. For our subsequent experiments we therefore decided to use three components to ensure a sufficiently rich font representation, while also avoiding overfitting to the data. The restriction to three dimensions had the additional advantage of allowing an individual font to be represented by a very low dimensional vector with only three entries, which reduces the computational burden for the subsequent Bayesian optimization process described in the following. The representation of the 25 fonts in the three-dimensional NMF space is shown in Figure 7.2.

Through this representation it is now possible to synthesize new fonts as a linear combinations of the three learnt font basis vectors. Accordingly, a font can be represented by a point in this three dimensional space. Figure 7.3 shows exemplary the influence of the three dimensions for the upper case letter K. Inspecting the letter renderings suggest that the first and third dimension are related to the scaling of letters in the vertical or horizontal directions, whereas the second dimension is related to the presence and strength of serifs. This generative model allows not only creating new fonts, but also the interpolation between actual fonts. Figure 7.4 shows the changes in three characters (capital *K*, lower case *y* and number *3*) resulting

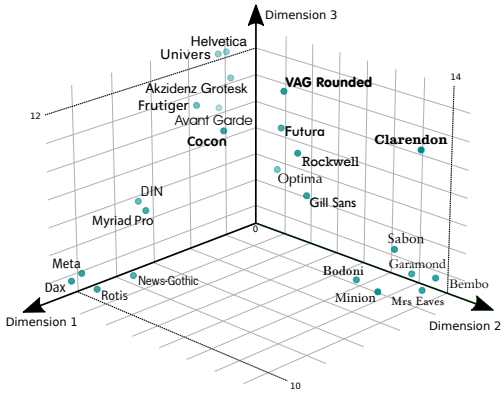


Fig. 7.2: Three-dimensional representation of the original fonts in the learned NMF font space.

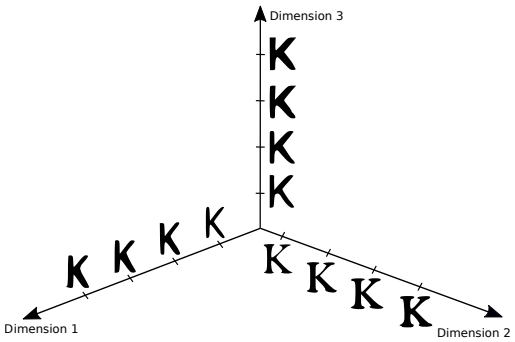


Fig. 7.3: Influence of the three font dimensions exemplarily demonstrated for one capital letter K.

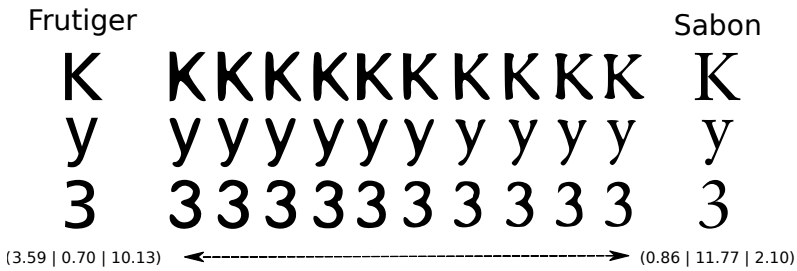


Fig. 7.4: Interpolating between the two fonts Frutiger and Sabon. At the edges are the original fonts, right next to them the NMF approximations and in between the linear interpolation through the font space. The coordinates for both fonts in our three-dimensional space is given.

from linearly moving in euclidean space between the points corresponding to the fonts Frutiger and Sabon. Note that the changes are gradual and smooth.

To be able to generate texts with a synthesized font we generate TrueType-Font (TTF) files on the fly. A linear combination of the basis vectors obtained through NMF yields a vector that contains a new greyscale bitmap image of all the glyphs in the new synthesized font as well as alignment information for each individual glyph. The bitmap images of the glyphs are traced individually to Scalable Vector Graphics (SVG) using Potrace (Selinger, 2001) and then stitched together into an SVG font. The SVG font then was converted into the TTF format using FontForge (Williams and FontForge Project contributors, 2000). At this point, the alignment information is directly inserted into a TTF file using the fontTools python package (Rossum, 2017). The entire process of font generation works automatically and in real time.

7.2.3 Bayesian Optimization

To generate new fonts that increase participants' reading speed, we need to select an optimization technique that finds corresponding regions in the three dimensional font space. The underlying assumption is that fonts change smoothly within the generative font space and that individual's reading speed is also changing smoothly along similar fonts. The objective function to be maximized is an individual's reading speed as function of a specific font, which in this case is represented by a point in our three dimensional font space. Since the font space is infinitely large and only a limited number of texts can be represented, we decided to use Bayesian Optimization, because it is a well-known optimization technique to find extrema of complex cost functions when the number of samples that can be drawn is limited (Brochu et al., 2010). Bayesian optimization starts with an a priori uncertainty

across the three-dimensional font volume and selects successive points within that volume, for which the reading speed is evaluated through a reading experiment. Since the choice for this prior distribution is hard in general, a common choice is the Gaussian process prior due to its flexibility and tractability (Snoek et al., 2012). The central assumption of the Gaussian Process is that any finite number of points can be expressed by an Multivariate Gaussian Distribution. The idea now is to take one of these points and assume that it is the value of the underlying unknown function. With the marginalization properties of the Multivariate Gaussian distribution it is now possible to compute marginal and conditional distributions given the observed data point. By obtaining the reading speed for this font through an experiment the algorithm reduces the uncertainty at that location in font space. To select the next best font for testing, a balance has to be struck between exploration, i.e. selecting a region in font space for which the uncertainty about reading speed is large, and exploitation, i.e. selecting a region in font space where the reading speed is expected to be high, given previous reading experiments. This selection process is handled by the so called acquisition function and its corresponding parameters. Thus, Bayesian Optimization proceeds by sampling more densely those region, where reading speed is high, and more evenly in those regions, where uncertainty is high.

7.2.4 Experimental Design and Data

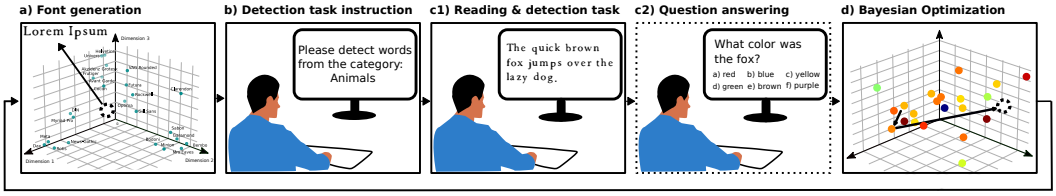


Fig. 7.5: Schematic of the closed-loop algorithm for generating and optimizing fonts to increase individuals' reading speed.

The overall closed-loop logic of the experiment (see Figure 7.5) was to have the Bayesian optimization algorithm generate a font in the generative font space and use the reading speed of individual participants to generate new fonts to increase the reading speed.

Participants

Eleven subjects (5 females, 6 males; age $M = 24$, $SD = 2.64$) participated in the experiment. All participants were German native speaker and had normal or corrected to normal vision. Due to the COVID-19 pandemic, it was not possible to

invite subjects to the laboratory, so they all performed the experiment on their private computers at home. Participants were acquired from graduate and undergraduate students of the research group and received the necessary materials for participation by e-mail. All subjects received a detailed, multi-page instruction to keep influences like sitting position, viewing distances, room lighting, etc. as similar as possible. In addition to detailed instructions and information on the experiment, the subjects also received an executable file containing the experiment. Subjects were naive with respect to the mechanics of font generation and selection.

Stimuli

Subjects read a total of 95 texts, which were taken from a German Wiki for children's encyclopedia texts. These texts were chosen because they are easily understandable for adult native speakers, so that the content of the texts had minimal effect on reading speed. Furthermore, the texts were chosen so that they were comparable in length, i.e. the number of words for the first 94 texts was on average 99.7 with a minimum of 91 and a maximum of 109 words. Additionally, a single text with only 51 words was selected to check whether the reading speed deviated significantly depending on text length. The texts were presented in a random order for each participant.

Procedure

Subjects were instructed to read the texts attentively and as quickly as possible but only once in total. To check that the texts were read and processed in terms of content, the subjects had the task of detecting words from previously specified categories while reading. Each individual trial started with the introduction of the category for the next text, e.g. before reading the text, it was indicated that words from the category *animals* should be detected subsequently. The category only referred to the next text, and a new category was selected for each trial. Each text's category had been previously independently labelled by two of the authors of this study and each subject was given the same category for a text.

For each text there were between one and ten words belonging to the instructed categories (mean 3.07, SD 2.05). Once a subject had read the category for the next trial, they could use the space bar to display the text and begin reading. When the text was displayed, the time measurement for the reading speed also started. To avoid an accidental start, the task could only be initiated a few seconds after the

category had been displayed (this was indicated to the participants by a red/green signal). Each time a term matching the previously specified category was detected, the participants could press the space bar to indicate this. As an example, in the short sentence "The quick brown fox jumps over the lazy dog.", when reading the words *fox* and *dog*, the user had to press the space bar to indicate that the objects were recognized for the category *animals*. Once they had finished reading the text they could press the Enter key to stop reading and thus also stop the timing.

In addition, a multiple choice question was asked after having read eleven texts at random trials throughout the experiment, in order to additionally test subjects' text comprehension. The question always referred to the last text read and six possible answers were given, of which exactly one was correct. After solving one of the multiple choice questions, participants received feedback on their answer. This feedback consisted of the average reading speed, the number of correct detections, and the correctness of the multiple choice question. These three components were combined into a score to further motivate participants to read quickly, but also correctly and attentively.

In order to investigate whether regions of the generative font space exist which allow higher reading speeds, subjects read the 95 texts in different synthesized fonts.

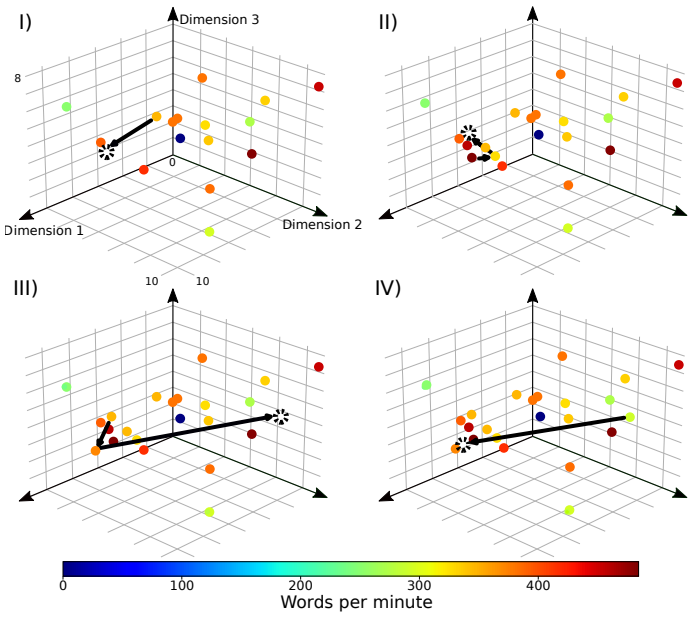


Fig. 7.6: Illustration of the Bayesian optimization probing process, after a new maximum was found. In I) and II) new points in a high reading speed cluster are sampled in an exploitation phase. After a slow new sample in III) was found in the exploration phase, the process is searching again in the region of the faster speeds in IV). Data from subject 7 in trials (I) 19 (II) 23 & 24 (III) 25 & 26 (IV) 27 respectively.

Since the three-dimensional space is infinitely large and only a limited number of texts can be presented, the Bayesian global optimization with Gaussian process was used in order to sample only 95 fonts. The target function to be optimized is the reading speed as parameterized by the three dimensions. The process was implemented with (Nogueira, 2014) and the following configuration was selected: We choose the Upper Confidence Bound (UCB) as acquisition function for balancing exploration and exploitation of new fonts (Srinivas et al., 2009). The exploration parameter was set to $\kappa = 5$. We started with ten random initialization points and chose the noise handling parameter as $\alpha = 0.001$ for the exploration strategy. We chose a Matern Kernel with parameter $\nu = 2.5$ for the covariance function and the L-BGFS-B algorithm for the optimization of the kernel parameters. Figure 7.6 shows an example for the sampling process through the Bayesian optimization algorithm.

As combinations of NMF components that have too high or too low magnitudes would for sure lead to unreadable fonts, we constrained the exploration space. If the sum of the three components is too low, the font is only faintly visible. If instead it is too high, the font is too bold and not readable either. Therefore, we empirically decided to constrain the magnitude of every dimension to be between 0 and 13 and the sum of all magnitudes to be between 7 and 20 units. It could happen that, in spite of these constraints, generated fonts were not readable. In such cases, subjects had been instructed to reset the corresponding trial via a previously defined key. Afterwards, the same text was presented in a new font, ensuring that all participants had read all 95 texts. This happened on 94 trials in total across all subjects (mean 8.54, SD 3.79). The Bayesian optimization algorithm nevertheless received the data that this linear combination returned a reading speed of zero words per minute.

7.3 Results

7.3.1 Ruling out Confounders

In order to rule out the possibility that the texts, despite the selected constraints, led to differences in reading speed in terms of content, the average reading speed in words per minute was determined for each text over all participants. A Bayesian ANOVA was used to check whether there was a significant difference in mean value and thus a dependence of the reading speed on the respective text. The Bayes Factor $B_{01} = 360.36$ gives decisive evidence for the null hypothesis that there was no significant differences between the texts.

A second possible confounder relates to the length of the texts, which participants read. By including a text with 51 words, we were able to check that the reading speed as measured in words per minute did not significantly depend on the chosen length of texts. Indeed, the reading speed for the shorter text (mean: 257.37 words per minute) was within the 95% confidence interval of reading speeds of the 94 remaining texts for all individual participants (mean: 265.02, SD: 89.82).

Another confounder refers to the detection task and the number of key presses required during a trial. In order to investigate whether the detections and motor behavior have an influence on the measured time, the Bayesian Pearson Correlation Coefficient between the number of words belonging to the instructed category and the corresponding reading time was computed. For this purpose the reading speeds were averaged over the number of expected detections to exclude the influences of all other factors influencing reading speed. No significant correlation was found ($r = -0.223; BF_{10} = 0.47$) thereby confirming that the number of words belonging to the instructed category did not have a significant effect on reading speed.

It might be argued that although reading speeds with AdaptiFont may be improved, reading speeds are still larger for traditional fonts. To exclude this possibility, we compared the reading speeds over all texts and subjects between the preliminary study, in which classic typefaces were used, and the main study using AdaptiFont. A Kolmogorov-Smirnov test indicates that the two distributions are indeed significantly different ($D = 0.367; p < .001$) and a Bayesian Two-Samples t-Test showed that the mean reading speeds for AdaptiFont were significantly higher than for the traditional typefaces ($BF_{+0} = 6.782$). Figure 7.7 shows the corresponding distributions.

Finally, we asked whether the optimization throughout the experiment actually reduced uncertainty about reading speed within the font space. For this, the point to be sampled next by the optimization process was considered. At this point, the

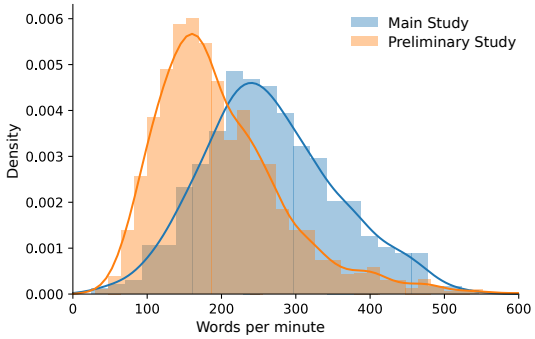


Fig. 7.7: Histograms of measured reading speeds for the preliminary (orange) and the main study (blue). A kernel density estimator was fitted to both histograms.

variance of the Gaussian Process was integrated within the neighborhood of a font with radius $r = 0.225$ before and after reading a text in the corresponding font. This radius is exactly half the distance between the two closest classic baseline fonts. For each iteration, the variance change in this region was now examined. Bayesian optimization indeed resulted in an average variance reduction of 943.65 units per iteration, the corresponding distribution can be found in Figure 7.8.

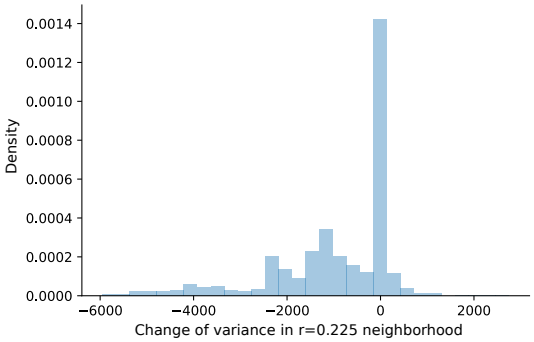


Fig. 7.8: Distribution of the per-sample uncertainty reduction. The histogram shows the distribution of all changes in uncertainty about the individual user's reading speed in the neighborhood of a sampled font after a reading experiment with a single font.

7.3.2 Clustering

To detect regions in the generative font space with higher reading speeds a clustering method was used, which includes the reading speed as fourth dimension in addition to the three font dimensions for each individual participant. We chose the OPTICS algorithm (Ankerst et al., 1999) to cluster the data points, which is a density based clustering procedure, which orders the points of a data set linearly, so spatially nearest neighbours can be clustered together. This has the advantage over other methods, such as k-Means, because the number of clusters does not have to be specified in advance and allows detecting outliers that are not assigned to any cluster. The latter feature was useful in the analysis of our data, as it excluded trials with deviations in reading speed due to lapses in attention.

As a free parameter of the algorithm, the minimum number of data points that must be present in a cluster must be defined. Following the recommendations of (Sander et al., 1998), this parameter can be determined using the heuristic of using twice the number of data dimensions minus one, i.e. $(2 \times \text{Dimensions}) - 1$, so that in our case we decided to set the parameter to $n = 5$. For visualization purposes, Figure 7.9 shows the best clusters, i.e. the clusters with the highest mean average reading speed for each of the 11 subjects in the generative font space. The clusters are

represented by an ellipsoid, with the center of all associated data points and the standard error in all three dimensions as main axes.

In order to better compare the optimized synthesized fonts with the traditional fonts, we additionally considered their position in font space with respect to each other. The optimized fonts for all subjects lay within a region bounded by the traditional fonts Rockwell, Myriad, Optima and News Gothic. To obtain an indication of the similarity and difference between individuals' optimized font, we computed the mean pairwise distance between individuals' optimized fonts. This distance is 3.99 units in the font space, which is comparable to the distance between Avantgarde and Universe. By comparison, the distance between slowest and fastest fonts for each subject was on average 11.3 units. In order to compare these distances within the font space, we computed the distances between classic fonts, with the smallest distance (0.45) between Helvetica and Univers, the largest distance (18.8) between Univers and Bembo, and the average distance (9.96), which was comparable to the distance between Optima and Minion.

7.3.3 Reading Speed Improvements with AdaptiFont

To check whether the font clusters correspond to significantly different reading speeds, a Bayesian ANOVA was calculated. Figure 7.10 shows the mean reading speeds for each of the found font clusters together with the corresponding 95% credible interval for each subject. The Bayes Factors are reported in Table 7.2, where every Bayes Factor gives decisive evidence against the null hypothesis, so

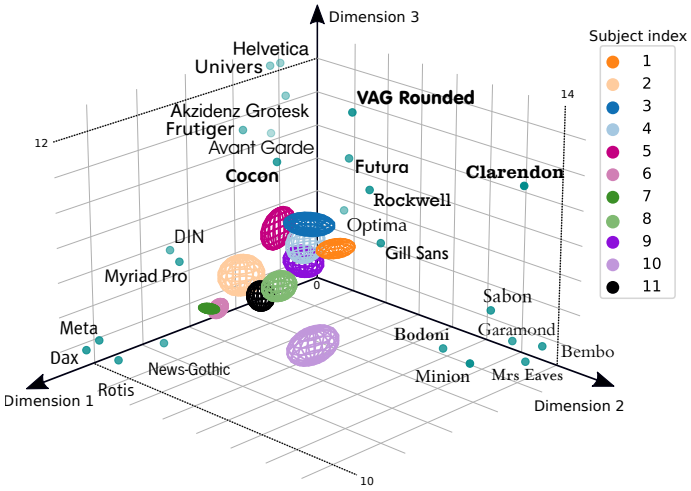


Fig. 7.9: Ellipsoids with centroids and standard errors in the three font space dimensions of clusters corresponding to highest reading speed.

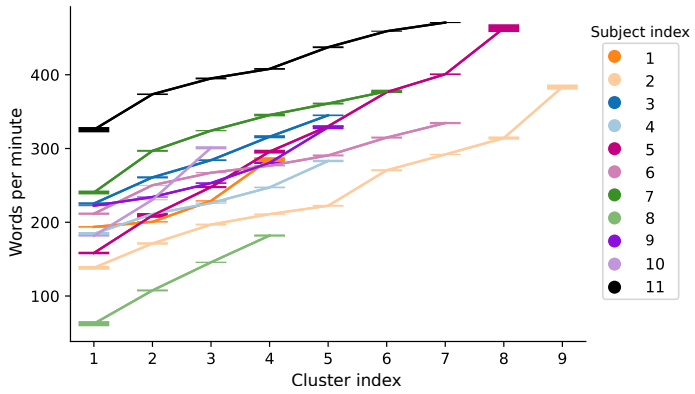


Fig. 7.10: Found clusters for every subject based on the OPTICS algorithms and the corresponding mean reading speed together with their standard errors. The mean reading speeds of all cluster were rank ordered for each subject. Note that the standard errors of reading speed were very small compared to the mean reading speeds for each cluster and subject.

that the clusters differ significantly in their average reading speed. A MANOVA established statistically significant differences in reading speed between clusters of highest reading speed of the subjects, $F(10, 75) = 3.24$, $p < .001$; Pillai Trace $= 0.91$. The corresponding ANOVA for every single dimension was performed resulting in three significant differences $D_1 : F(10, 75) = 2.01, p < .05$; $D_2 : F(10, 75) = 3.63, p < .05$; $D_3 : F(10, 75) = 4.02, p < .05$. Thus, the differences between clusters of individuals' highest reading speed visible in figure 7.9 are statistically significant. To summarize together with the results from section 7.3.1 and Figure 7.7: The distribution generated with AdaptiFont is not only statistically significantly different, the reading speed with Adaptifont is also statistically significantly faster on average. All subjects showed improved reading speed.

Tab. 7.2: Bayes Factors of the ANOVAs relating individuals' reading speed between the found clusters.

Subject	BF_{10}
1	4.458e+23
2	5.651e+56
3	5.890e+35
4	3.592e+30
5	5.752e+45
6	6.091e+41
7	1.804e+37
8	5.974e+31
9	4.683e+29
10	5.956e+14
11	2.008e+48

Finally, figure 7.11 shows a pangram written in the respective font of the center of the cluster of highest reading speed for three exemplary subjects. The fonts are actual TrueType-fonts and can be used as such.

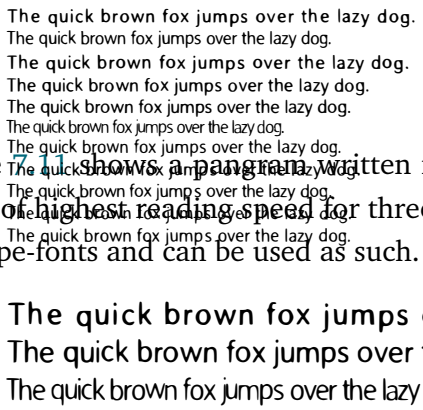


Fig. 7.11: Fonts generated from the centroids of the best clusters of subjects 1, 5, and 6.

7.3.4 Eye movement behavior

In another study, we tested the AdaptiFont system on eight additional naïve subjects. We could find the same effects and results as in the main study. In addition, we measured eye movements. This could be a possible extension to later measure reading speed without manual indication of the readers. We can see an exponential relationship between the two metrics in Figure 7.12. The more fixations made while reading a text, the slower the subjects read. By measuring just this relationship, reading speed can be estimated in the future by simply counting fixations while reading.

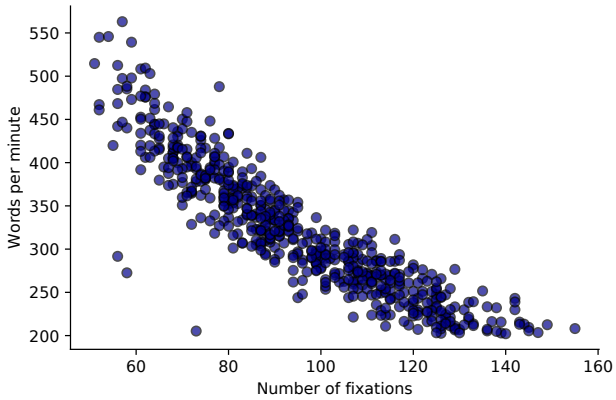


Fig. 7.12: Relationship between the number of fixations and the measured reading speed.

7.4 Discussion

We presented AdaptiFont, a human-in-the-loop system to increase readability of digitally rendered text. The system uses a learned generative font space and Bayesian optimization to generate successive fonts that are evaluated in terms of individual

user's reading speed. A preliminary study gave first indications that individual differences in readability exist in terms different font features. With AdaptiFont, variation in font features are captured in a specific font space learned through NMF. Using this representation, we explore the reading speed of individual users with the help of Bayesian optimization. The results of the user study show the feasibility of the approach, both in terms of the required data to find fonts that increase individual's reading speed as well as the statistically significant magnitude of the improvement in individual's reading speed. Finally, significant differences between subjects exist in the font regions that contain fonts that are associated with high legibility. The generated fonts are actual TrueType-font files, which can be installed and used.

Although reading speed maximizing fonts found in this study were significantly different across individual subjects, it cannot be concluded that our system finds one single font that maximizes reading speed unchangeably for a single subject or across all subjects. Rather, our system can be understood as dynamically and continuously creating fonts for an individual, which maximizes the reading speed at the time of use. This may depend on the content of the text, whether you are exhausted, or perhaps are using different display devices. The empirical data we obtained in the experiments provides clear evidence that Adaptifont increased all participants' reading speeds. How these optimized fonts are related across subjects or within subjects across days or viewing contexts are very interesting further empirical questions, which can be systematically addressed by utilizing Adaptifont in future research.

While the current study demonstrates the feasibility of interactively optimizing fonts through Bayesian optimization, it has a number of limitations, which should be addressed in further research. First, while reading speed is a natural target for optimizing fonts, other criteria are conceivable, such as text comprehension, aesthetic appeal, or memorability either of texts or fonts. Second, ample research has demonstrated cognitive and linguistic influences on reading speed, which were not taken into account here. Third, future generative font spaces, e.g. based on GANs, may provide a more expressive font synthesis with smaller approximation errors to classic fonts and better alignment and kerning. Fourth, while the current study maximized reading speed at the individual subject level, it is straightforward to use the system to maximize optimization criteria across multiple subjects. Fifth, the generative font space based on NMF may result in synthetic fonts that violate typographic rules. Including typographic constraints or interactively including a typography expert in the font synthesis process may yield an additional avenue for the development of new fonts.

While constant, variable, and parametric fonts have been designed and developed in the past, Adaptifont is an interactive system, which generates new fonts based on a user's interaction with text. In this chapter, we used a generative font model and Bayesian optimization to generate fonts, which progressively increased users' reading speeds. Adaptifont can be understood as a system that dynamically and continuously creates fonts for an individual, which in this study maximized the reading speed at the time of use. Evaluation of the system in a user study showed the feasibility of the approach. Adaptifonts can be utilized to address further scientific questions such as how the optimized fonts depend on sensory, linguistic, or contextual factors or how optimized fonts are related across individuals. The system does in no way detract from other work on generating fonts, both from the point of view of developing such algorithms or from the point of view of the typographer, as developing typefaces can be guided by different motivations and with different goals.

8 Conclusion

In this thesis, we investigated the framework of active vision as a probabilistic decision-making process under uncertainty. For this purpose, we have developed several experimental tasks to investigate different aspects that make planning necessary in visual behavior: sensory uncertainty, action variability, behavioral costs and internal model uncertainty. We have used the Partially Observable Markov Decision Processes framework, which allows us to model these components normatively and infer individuals' uncertainties and internal costs.

8.1 Individual uncertainties and costs

First, we reasoned that vision is a highly variable process between individuals driven by sensory and cognitive factors (Hayhoe and Ballard, 2005; Kowler, 2011; Land and Tatler, 2009; Tatler et al., 2011). Perception is subject to uncertainty, as is the execution of our actions and their results. In addition, there is a variance in the internal objectives of each individual and the associated costs. We could show these differences in human behavior in various chapters and related experiments. We observe sensory uncertainty, for example, in the exact position of stimuli (degree on the circular path in Chapter 3, the position of the random walk in Chapter 5, or the position within the maze in Chapter 6) or even in the visual perception of reward (noise pattern in Chapter 5). Time perception is also subject to uncertainty and can be modeled using Weber's law (see, e.g., Hoppe and Rothkopf, 2016; Mauk, Buonomano, et al., 2004 and Chapter 5). Similarly, when reading, we perceive fonts better or worse based on typographic factors (see, e.g., Bouma, 1970; Paterson and Tinker, 1931 and Chapter 7). Our actions are also subject to uncertainty in their execution; for example, the subsequent fixation may land in different places due to motor variability or cognitive costs such as deciding where to move next and when (Chapter 4). In the same way, we may also follow different strategies in solving problems, so our actions may also depend on how we view our environment (Chapter 6). But also daily circumstances can influence our actions, such as fatigue, cognitive load, age, or medical conditions Leigh and Zee, 2015. An individual's preferred font may also vary based on this so that we can observe variability between individuals

and within them (Chapter 7). Each individual has internal goals which are not visible to the outside world. These significantly influence our behavior and are associated with behavioral costs. For example, a task may be more or less important to us, directly affecting our performance (Chapter 3). Another well-known example is the free-viewing paradigm. Subjects are not instructed to look at a visual scene. In this case, the internal objective is not detectable, so free-viewing can be understood as integration across all possible tasks (Chapter 4). But also, the preferences in the expression of fonts are highly variable and depend on the individual (Chapter 7). We also introduced the concept of internal model uncertainty. We have seen this, for example, in chapter 5, where the model forms a belief about where the random walk is at the moment. In addition to the uncertainty of the sensory percepts, the internal model uncertainty grows over time. An example is the uncertainty about the generative model in Chapter 3, where we introduced belief parameters to represent the uncertainty about the actual underlying event distribution.

8.2 Visual planning

Planning is a series of cognitive processes that help bridge the discrepancy between a current state and a desired end state (Miller et al., 2017). For our visual apparatus, this means that we must actively perform sequential actions to obtain our necessary information from the visual environment. The most straightforward strategy here would be that, given our task, we select the visual location that maximizes our immediate reward. This strategy is not feasible in practical scenarios since tasks often require a series of interconnected actions. Considering sequential eye movements, rewards accruing after a sequence of actions can become significant. Thus, the challenge for the visual system is to determine an optimal series of actions, that is, to plan the visual actions. We found evidence for this in the experiments we conducted. In Chapter 3, we saw that subjects do not simply blink to get the immediate reward of eye moistening. Instead, they develop a trade-off of the neurophysiological and task-related costs. Chapter 4 showed us that one-step prediction is insufficient to predict an entire sequence of fixations. In Chapter 5, subjects performed the monitoring task differently given the position of the random walk and planned their action rather than simply briefly updating the necessary information. When solving the mazes in Chapter 6, we even saw and investigated different planning strategies similar to depth or breadth search. All these observations support the hypothesis that vision is a sequence of visual actions and that people plan this over a longer time horizon. Therefore, models of human visual behavior must also model this aspect appropriately.

8.3 Partially Observable Markov Decision Processes

We gathered empirical evidence that human visual behavior can be highly individualized and identified the components of sensory uncertainty, action variability, and behavioral cost. We have also seen that visual behavior consists of active, sequential decisions that people plan. Therefore, we looked at the Partially Observable Markov Decision Processes framework as a unifying modeling approach for modeling purposes. POMDPs allow for incorporating different sources of uncertainties and costs associated with human behavior. We have seen this in the various parts of the thesis. Thus, in Chapter 3, we used a POMDP to model normatively the previously approximate cost model of (Hoppe et al., 2018). In this computational planning framework, we demonstrate that individual behavioral costs are linked to blink suppression and task motivation. We have successfully simulated the interrelation between the cost-trade-off parameter and task performance, unveiling a logistic relationship. Moreover, we devised a normative model characterizing human behavior, avoiding the presumption or fitting of any specific functional form of the cost functions. Given that many models in the field adopt this approach (see, for example, Petitet et al., 2021), our framework is more flexible and natural. We also modeled the planning behavior in Chapter 5 using a POMDP. Our experimental design considers dynamic uncertainties and reward rates. This lets us examine adaptive responses to these dynamics, investigate switching behaviors, and determine participants' intrinsic costs. Our findings indicate that humans can manage temporal uncertainties in their observations and navigate uncertain rewards to distinguish trade-offs. Contrary to earlier studies, we can also present empirical data suggesting that the subjective internal costs of saccadic gaze shifts are significantly higher. These findings underscore the significance of planning for our visual system.

8.3.1 Human visual behavior as probabilistic planning under uncertainty

We have seen that seeing is a highly individual process. In addition to uncertainties in perception and action, internal and external costs also play a role. We have seen that we can find empirical evidence for this and deduced that seeing is an action of active sequential decisions. This corresponds computationally to the concept of planning, which can manifest itself in different strategies. We have shown that these are needed for successful completion and again found evidence that people implement them. For modeling, we then looked at the framework of Partially Observable Markov Decision Processes, a normative and promising way to represent these sequential decisions under uncertainty and including costs. We have

shown the results of these models and presented their advantages. Finally, we can conclude that human visual behavior corresponds to sequential decision behavior under uncertainty. Active vision theory models this and emphasizes the need for modeling using individual components.

8.4 Future Work

Future work should further advance the introduced normative framework of Partially Observable Markov Decision Processes for studying visual behavior, extending it to implement suitable models for various everyday situations and experimental paradigms. For example, the developed method for studying temporal event statistics in blinks can be used to predict behavior in continuous natural stimuli, such as videos, utilizing so-called activity forecasting methods to estimate the underlying generative model of event statistics. Along with the insights gained by exploring the paradigm with simultaneous dynamic uncertainties and rewards, the future study of human cost functions can also be advanced. A distinct advantage of our methods lies in their normative nature, which does not necessitate the prior specification of functional relationships in the model. Moreover, the framework of active decisions under uncertainty can enhance the one-step predictions in the salience model framework, describing the entire fixation train in future instances. However, our findings also present practical applications beyond academic settings. A salient example already presented is the maximization of adaptive reading speed using AdaptiFont. This human-in-the-loop process can conceivably be extended to numerous other areas, providing customized individual solutions.

Bibliography

Achcar, Jorge Alberto and Ricardo Puziol de Oliveira (2022). “Climate change: use of non-homogeneous Poisson processes for climate data in presence of a change-point”. In: *Environmental Modeling & Assessment*, pp. 1–14 (cit. on p. 30).

Adebayo, Julius, Justin Gilmer, Michael Muelly, et al. (2018). “Sanity checks for saliency maps”. In: *Advances in Neural Information Processing Systems*, pp. 9505–9515 (cit. on p. 47).

Adelson, Edward H and James R Bergen (1985). “Spatiotemporal energy models for the perception of motion”. In: *Josa a* 2.2, pp. 284–299 (cit. on p. 2).

Al-Dousari, Ahmad, Asad Ellahi, and Ijaz Hussain (2021). “Use of non-homogeneous Poisson process for the analysis of new cases, deaths, and recoveries of COVID-19 patients: A case study of Kuwait”. In: *Journal of King Saud University-Science* 33.8, p. 101614 (cit. on p. 30).

Aloimonos, John, Isaac Weiss, and Amit Bandyopadhyay (1988). “Active vision”. In: *International journal of computer vision* 1, pp. 333–356 (cit. on p. 6).

Alonso, Alejandra, Jacqueline van der Meij, Dorothy Tse, and Lisa Genzel (2020). “Naïve to expert: Considering the role of previous knowledge in memory”. In: *Brain and Neuroscience Advances* 4 (cit. on p. 87).

Anderson, John R (1991). “Is human cognition adaptive?” In: *Behavioral and brain sciences* 14.3, pp. 471–485 (cit. on p. 67).

André, Jacques and Bruno Borghi (1989). “Dynamic fonts”. In: *Raster imaging and digital typography*. Cambridge: Cambridge University Press, pp. 198–204 (cit. on p. 105).

André, Jacques and Irène Vatton (1994). “Dynamic optical scaling and variable-sized characters”. In: *Electronic Publishing* 7.4, pp. 231–250 (cit. on pp. 102, 104).

Andreu-Sanchez, Celia, Miguel Angel Martin-Pascual, Agnes Gruart, and Jose Maria Delgado-Garcia (2021). “Viewers change eye-blink rate by predicting narrative content”. In: *Brain sciences* 11.4, p. 422 (cit. on p. 30).

Ankerst, Mihael, Markus M. Breunig, Hans-Peter Kriegel, and Jörg Sander (June 1999). “OPTICS: Ordering Points to Identify the Clustering Structure”. In: *SIGMOD Rec.* 28.2, 49–60 (cit. on p. 116).

Aqel, Mohammad OA, Ahmed Issa, Mohammed Khdair, et al. (2017). “Intelligent maze solving robot based on image processing and graph theory algorithms”. In: *2017 International Conference on Promising Electronic Technologies (ICPET)*. IEEE, pp. 48–53 (cit. on p. 87).

Arditi, Aries (Apr. 2004). “Adjustable typography: an approach to enhancing low vision text accessibility”. In: *Ergonomics* 47.5, pp. 469–482 (cit. on pp. 102, 105).

Arditi, Aries and Jianna Cho (Nov. 2005). “Serifs and font legibility”. In: *Vision Research* 45.23, pp. 2926–2933 (cit. on pp. 7, 103, 104).

Assens, Marc, Kevin McGuinness, Xavier Giro-i Nieto, and Noel E. O’Connor (Aug. 2017). “SaltiNet: Scan-path Prediction on 360 Degree Images using Saliency Volumes”. en. In: *arXiv:1707.03123 [cs]*. arXiv: 1707.03123 (cit. on pp. 46, 48).

Assens, Marc, Xavier Giro-i Nieto, Kevin McGuinness, and Noel E. O’Connor (2019). “PathGAN: Visual Scanpath Prediction with Generative Adversarial Networks”. en. In: *Computer Vision – ECCV 2018 Workshops*. Ed. by Laura Leal-Taixé and Stefan Roth. Vol. 11133. Cham: Springer International Publishing, pp. 406–422 (cit. on pp. 46–48).

Åström, Karl Johan (1965). “Optimal control of Markov processes with incomplete state information”. In: *Journal of mathematical analysis and applications* 10.1, pp. 174–205 (cit. on p. 24).

Azadi, Samaneh, Matthew Fisher, Vladimir Kim, et al. (2018). “Multi-content gan for few-shot font style transfer”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Vol. 11. Salt Lake City: IEEE, p. 13 (cit. on pp. 105, 108).

Bajcsy, Ruzena (1988). “Active perception”. In: *Proceedings of the IEEE* 76.8, pp. 966–1005 (cit. on p. 6).

Ballard, Dana H. (1991). “Animate vision”. In: *Artificial Intelligence* 48.1, pp. 57–86 (cit. on pp. 1, 6).

Ballard, Dana H, Mary M Hayhoe, Polly K Pook, and Rajesh PN Rao (1997). “Deictic codes for the embodiment of cognition”. In: *Behavioral and brain sciences* 20.4, pp. 723–742 (cit. on p. 6).

Baloh, R. W., A. W. Sills, W. E. Kumley, and V. Honrubia (1975). “Quantitative measurement of saccade amplitude, duration, and velocity”. In: *Neurology* 25.11, pp. 1065–1070 (cit. on p. 79).

Bar, Moshe (2004). “Visual objects in context”. In: *Nature Reviews Neuroscience* 5.8, pp. 617–629 (cit. on p. 2).

Bayes, Thomas (1763). “LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S”. In: *Philosophical transactions of the Royal Society of London* 53, pp. 370–418 (cit. on p. 14).

Bellmann, Richard (1957). “A Markovian Decision Process”. In: *Journal of Mathematics and Mechanics* 6.5, pp. 679–684 (cit. on pp. 21, 22).

Bellot, V., M. Cautrès, J-M. Favreau, et al. (2021). “How to generate perfect mazes?” In: *Information Sciences* 572, pp. 444–459 (cit. on p. 88).

Belousov, Boris, Gerhard Neumann, Constantin A Rothkopf, and Jan R Peters (2016). “Catching heuristics are optimal control policies”. In: *Advances in neural information processing systems* 29 (cit. on p. 27).

Berman, Brian D, Silvina G Horovitz, Brent Morel, and Mark Hallett (2012). "Neural correlates of blink suppression and the buildup of a natural bodily urge". In: *Neuroimage* 59.2, pp. 1441–1450 (cit. on p. 41).

Bernard, Jean-Baptiste, Girish Kumar, Jasmine Junge, and Susana TL Chung (2013). "The effect of letter-stroke boldness on reading speed in central and peripheral vision". In: *Vision research* 84, pp. 33–42 (cit. on p. 104).

Bernard, Michael, Bonnie Lida, Shannon Riley, Telia Hackler, and Karen Janzen (2002). "A comparison of popular online fonts: Which size and type is best". In: *Usability news* 4.1, p. 2002 (cit. on pp. 101, 103).

Bezanson, Jeff, Alan Edelman, Stefan Karpinski, and Viral B Shah (2017). "Julia: A fresh approach to numerical computing". In: *SIAM Review* 59.1, pp. 65–98 (cit. on pp. 23, 35, 78).

Biederman, Irving (2017). "On the semantics of a glance at a scene". In: *Perceptual organization*. Routledge, pp. 213–253 (cit. on p. 2).

Bigelow, Charles (Nov. 2019). "Typeface features and legibility research". In: *Vision Research* 165 (cit. on pp. 101, 103).

Bingham, Eli, Jonathan P Chen, Martin Jankowiak, et al. (2019). "Pyro: Deep universal probabilistic programming". In: *The Journal of Machine Learning Research* 20.1, pp. 973–978 (cit. on p. 17).

Bishop, Christopher M and Nasser M Nasrabadi (2006). *Pattern recognition and machine learning*. Vol. 4. 4. Springer (cit. on p. 14).

Bitzer, Sebastian, Hame Park, Felix Blankenburg, and Stefan J Kiebel (2014). "Perceptual decision making: drift-diffusion model is equivalent to a Bayesian model". In: *Frontiers in human neuroscience* 8, p. 102 (cit. on p. 19).

Bjørklund, Roald A (1991). "Reaction time and movement time measured in a key-press and a key-release condition". In: *Perceptual and Motor skills* 72.2, pp. 663–673 (cit. on p. 79).

Blommaert, Frans JJ and Han Timmers (1987). "Letter recognition at low contrast levels: effects of letter size". In: *Perception* 16.4, pp. 421–432 (cit. on pp. 7, 103).

Boccignone, Giuseppe, Vittorio Cuculo, and Alessandro D'Amelio (Aug. 2020). "How to Look Next? A Data-Driven Approach for Scanpath Prediction". In: Springer, pp. 131–145 (cit. on p. 48).

Boccignone, Giuseppe and Mario Ferraro (Jan. 2004). "Modelling gaze shift as a constrained random walk". en. In: *Physica A: Statistical Mechanics and its Applications* 331.1-2, pp. 207–218 (cit. on pp. 46, 48, 63).

Borges, Felipe Placeres, Denny Marcos Garcia, and Antonio Augusto Velasco Cruz (2010). "Distribution of spontaneous inter-blink interval in repeated measurements with and without topical ocular anesthesia". In: *Arquivos brasileiros de oftalmologia* 73, pp. 329–332 (cit. on p. 30).

Borji, Ali (May 2019). “Saliency Prediction in the Deep Learning Era: Successes, Limitations, and Future Challenges”. en. In: *arXiv:1810.03716 [cs]*. arXiv: 1810.03716 (cit. on pp. 46, 48).

Borji, Ali and Laurent Itti (2013). “State-of-the-art in visual attention modeling”. In: *IEEE transactions on pattern analysis and machine intelligence* 35.1, pp. 185–207 (cit. on pp. 46, 48).

Borji, Ali, Dicky N Sihite, and Laurent Itti (2012). “Probabilistic learning of task-specific visual attention”. In: *2012 IEEE Conference on computer vision and pattern recognition*. IEEE, pp. 470–477 (cit. on p. 47).

– (2013). “What stands out in a scene? A study of human explicit saliency judgment”. In: *Vision research* 91, pp. 62–77 (cit. on p. 67).

Bouma, Herman (1970). “Interaction effects in parafoveal letter recognition”. In: *Nature* 226.5241, pp. 177–178 (cit. on pp. 103, 123).

Boyarski, Dan, Christine Neuwirth, Jodi Forlizzi, and Susan Harkness Regli (1998). “A Study of Fonts Designed for Screen Display”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’98. Los Angeles, California, USA: ACM Press/Addison-Wesley Publishing Co., 87–94 (cit. on p. 103).

Bragg, Danielle, Shiri Azenkot, and Adam Tauman Kalai (2016). “Reading and Learning Smartfonts”. In: *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. UIST ’16. Tokyo, Japan: Association for Computing Machinery, 391–402 (cit. on pp. 102, 105).

Brautigan, Richard (2014). *Trout fishing in America*. Vol. 30. Canongate Books (cit. on p. 29).

Breese, B. B. (1899). “On inhibition”. In: *The Psychological Review: Monograph Supplements* 3.1, pp. i–65 (cit. on p. 2).

Bristow, Davina, Chris Frith, and Geraint Rees (2005). “Two distinct neural effects of blinking on human visual processing”. In: *Neuroimage* 27.1, pp. 136–145 (cit. on p. 29).

Brochu, Eric, Vlad M. Cora, and Nando de Freitas (2010). “A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning”. In: *CoRR* abs/1012.2599. arXiv: 1012.2599 (cit. on p. 110).

Bron, AJ, JM Tiffany, SM Gouveia, N Yokoi, and LW Voon (2004). “Functional aspects of the tear film lipid layer”. In: *Experimental eye research* 78.3, pp. 347–360 (cit. on p. 29).

Brownie, Barbara (Dec. 2007). “One Form, Many Letters: Fluid and transient letterforms in screen-based typographic artefacts”. In: *Networking Knowledge: Journal of the MeCCSA Postgraduate Network* 1 (cit. on p. 105).

Bruce, Neil and John Tsotsos (June 2007). “Attention based on information maximization”. In: *Journal of Vision* 7.9, pp. 950–950 (cit. on p. 57).

Bruce, Vicki and Andy Young (1986). “Understanding face recognition”. In: *British journal of psychology* 77.3, pp. 305–327 (cit. on p. 2).

Bruijn, David de, Sjaak de Mul, and Herre van Oostendorp (1992). “The influence of screen size and text layout on the study of text”. In: *Behaviour & information technology* 11.2, pp. 71–78 (cit. on pp. 101, 103).

Buecher, Simon J, Christoph Holscher, and Jan Wiener (2009). “Search strategies and their success in a virtual maze”. In: *Proceedings of the annual meeting of the cognitive science society*. Vol. 31. 31 (cit. on p. 87).

Buswell, Guy Thomas (1935). “How people look at pictures: a study of the psychology and perception in art.” In: (cit. on p. 2).

Butko, Nicholas J. and Javier R. Movellan (2008). “I-POMDP: An infomax model of eye movement”. In: *2008 7th IEEE International Conference on Development and Learning*, pp. 139–144 (cit. on pp. 27, 68).

– (2009). “Optimal scanning for faster object detection”. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2751–2758 (cit. on pp. 27, 68).

Bylinskii, Zoya, Tilke Judd, Aude Oliva, Antonio Torralba, and Fredo Durand (Mar. 2019). “What Do Different Evaluation Metrics Tell Us About Saliency Models?” en. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41.3, pp. 740–757 (cit. on pp. 47, 55, 58).

Byrd, Richard H., Peihuang Lu, Jorge Nocedal, and Ciyou Zhu (1995). “A Limited Memory Algorithm for Bound Constrained Optimization”. In: *SIAM Journal on Scientific Computing* 16.5, pp. 1190–1208. eprint: https://doi.org/10.1137/0916069 (cit. on p. 56).

Cai, Rick H., Alexandre Pouget, Madeleine Schlag-Rey, and John Schlag (Apr. 1997). “Perceived geometrical relationships affected by eye-movement signals”. In: *Nature* 386.6625, pp. 601–604 (cit. on p. 6).

Callaway, Frederick, Bas van Opheusden, Sayan Gul, et al. (2022). “Rational use of cognitive resources in human planning”. In: *Nature Human Behaviour* 6.8, pp. 1112–1125 (cit. on p. 66).

Campbell, Neill DF and Jan Kautz (2014). “Learning a manifold of fonts”. In: *ACM Transactions on Graphics (TOG)* 33.4, pp. 1–11 (cit. on p. 105).

Carpenter, Bob, Andrew Gelman, Matthew D Hoffman, et al. (2017). “Stan: A probabilistic programming language”. In: *Journal of statistical software* 76.1 (cit. on p. 17).

Carter, Rob, Philip B Meggs, and Ben Day (2011). *Typographic design: Form and communication*. Hoboken, New Jersey: John Wiley & Sons (cit. on pp. 101, 102).

Carton, Daniel, Verena Nitsch, Dominik Meinzer, and Dirk Wollherr (2016). “Towards assessing the human trajectory planning horizon”. In: *Plos one* 11.12, e0167021 (cit. on p. 88).

Cassandra, A. R., L. P. Kaelbling, and M. L. Littman (1994). “Acting optimally in partially observable stochastic domains”. In: *Proceedings of the 12th National Conference on Artificial Intelligence*. Vol. 2, 1023–1028 (cit. on pp. 26, 31, 35, 78).

Cerf, Moran, Jonathan Harel, Wolfgang Einhäuser, and Christof Koch (2008). "Predicting human gaze using low-level saliency combined with face detection". In: *Advances in neural information processing systems*, pp. 241–248 (cit. on p. 47).

Cha, Junbum, Sanghyuk Chun, Gayoung Lee, et al. (2020). "Few-shot Compositional Font Generation with Dual Memory". In: *European Conference on Computer Vision (ECCV)*. Online: Springer (cit. on pp. 105, 108).

Chaspari, Theodora, Matthew Goodwin, Oliver Wilder-Smith, et al. (2014). "A non-homogeneous Poisson process model of skin conductance responses integrated with observed regulatory behaviors for Autism intervention". In: *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, pp. 1611–1615 (cit. on p. 31).

Chebolu, Sahiti, Peter Dayan, and Kevin Lloyd (Oct. 2022). "Vigilance, arousal, and acetylcholine: Optimal control of attention in a simple detection task". In: *PLOS Computational Biology* 18.10, pp. 1–36 (cit. on pp. 27, 68).

Chen, Chenyi, Ari Seff, Alain Kornhauser, and Jianxiong Xiao (2015). "Deepdriving: Learning affordance for direct perception in autonomous driving". In: *Proceedings of the IEEE international conference on computer vision*, pp. 2722–2730 (cit. on p. 8).

Chen, Tianlang, Zhaowen Wang, Ning Xu, Hailin Jin, and Jiebo Luo (2019). "Large-scale Tag-based Font Retrieval with Generative Feature Learning". In: *CoRR* abs/1909.02072. arXiv: 1909.02072 (cit. on pp. 105, 108).

Chung, Susana TL (2004). "Reading speed benefits from increased vertical word spacing in normal peripheral vision". In: *Optometry and vision science: official publication of the American Academy of Optometry* 81.7, p. 525 (cit. on p. 103).

Chung, Susana TL, J Stephen Mansfield, and Gordon E Legge (1998). "Psychophysics of reading. XVIII. The effect of print size on reading speed in normal peripheral vision". In: *Vision research* 38.19, pp. 2949–2962 (cit. on p. 103).

Churchland, Mark M., Afsheen Afshar, and Krishna V. Shenoy (2006). "A Central Source of Movement Variability". In: *Neuron* 52.6, pp. 1085–1096 (cit. on p. 4).

Cifuentes-Amado, María Victoria and Edilberto Cepeda-Cuervo (2015). "Non-homogeneous Poisson process to model seasonal events: Application to the health diseases". In: *International Journal of Statistics in Medical Research* 4.4, p. 337 (cit. on p. 31).

Clark, Andy (2013). "Whatever next? Predictive brains, situated agents, and the future of cognitive science". In: *Behavioral and brain sciences* 36.3, pp. 181–204 (cit. on p. 5).

Cornia, Marcella, Lorenzo Baraldi, Giuseppe Serra, and Rita Cucchiara (Oct. 2018). "Predicting Human Eye Fixations via an LSTM-based Saliency Attentive Model". en. In: *IEEE Transactions on Image Processing* 27.10. arXiv: 1611.09571, pp. 5142–5154 (cit. on pp. 46–48, 54).

Cristino, Filipe, Sebastiaan Mathôt, Jan Theeuwes, and Iain D Gilchrist (2010). "ScanMatch: A novel method for comparing fixation sequences". In: *Behavior research methods* 42.3, pp. 692–700 (cit. on p. 48).

Crowe, David A., Bruno B. Averbeck, Matthew V. Chafee, John H. Anderson, and Apostolos P. Georgopoulos (2000). "Mental Maze Solving". In: *Journal of Cognitive Neuroscience* 12.5, pp. 813–827 (cit. on pp. 88, 90).

Curcio, Christine A, Kenneth R Sloan, Robert E Kalina, and Anita E Hendrickson (1990). "Human photoreceptor topography". In: *Journal of comparative neurology* 292.4, pp. 497–523 (cit. on p. 4).

Dale, Edgar and Jeanne S Chall (1949). "The concept of readability". In: *Elementary English* 26.1, pp. 19–26 (cit. on p. 102).

Dam, Loes CJ van and Raymond van Ee (2006). "The role of saccades in exerting voluntary control in perceptual and binocular rivalry". In: *Vision research* 46.6-7, pp. 787–799 (cit. on p. 2).

Dang, Hongshe, Jinguo Song, and Qin Guo (2010). "An efficient algorithm for robot maze-solving". In: *2010 Second International Conference on Intelligent Human-Machine Systems and Cybernetics*. Vol. 2. IEEE, pp. 79–82 (cit. on p. 87).

Darroch, Iain, Joy Goodman, Stephen Brewster, and Phil Gray (2005). "The effect of age and font size on reading text on handheld computers". In: *IFIP conference on human-computer interaction*. Springer. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 253–266 (cit. on p. 103).

Davies, Joseph, Huy Truong-Ba, Michael E Cholette, and Geoffrey Will (2021). "Optimal inspections and maintenance planning for anti-corrosion coating failure on ships using non-homogeneous Poisson Processes". In: *Ocean Engineering* 238, p. 109695 (cit. on p. 30).

Daw, Nathaniel D, Yael Niv, and Peter Dayan (2005). "Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control". In: *Nature neuroscience* 8.12, pp. 1704–1711 (cit. on p. 23).

Dayan, Peter and Nathaniel D Daw (2008). "Decision theory, reinforcement learning, and the brain". In: *Cognitive, Affective, & Behavioral Neuroscience* 8.4, pp. 429–453 (cit. on p. 20).

De Haas, Benjamin, Alexios L Iakovidis, D Samuel Schwarzkopf, and Karl R Gegenfurtner (2019). "Individual differences in visual salience vary along semantic dimensions". In: *Proceedings of the National Academy of Sciences* 116.24, pp. 11687–11692 (cit. on p. 88).

Devroye, Luc and Michael McDougall (1995). "Random fonts for the simulation of handwriting". In: *Electronic Publishing (EP—ODD)* 8, pp. 281–294 (cit. on pp. 102, 105).

Dhakal, Vivek, Anna Maria Feit, Per Ola Kristensson, and Antti Oulasvirta (2018). "Observations on Typing from 136 Million Keystrokes". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 1–12 (cit. on p. 79).

Diaz, Gabriel, Joseph Cooper, Constantin Rothkopf, and Mary Hayhoe (2013). "Saccades to future ball location reveal memory-based prediction in a virtual-reality interception task". In: *Journal of vision* 13.1, pp. 20–20 (cit. on p. 65).

Doane, Marshall G (1980). “Interactions of eyelids and tears in corneal wetting and the dynamics of the normal human eyeblink.” In: *American journal of ophthalmology* 89.4, pp. 507–516 (cit. on p. 29).

Dobres, Jonathan, Nadine Chahine, Bryan Reimer, et al. (2016). “Utilising psychophysical techniques to investigate the effects of age, typeface design, size and display polarity on glance legibility”. In: *Ergonomics* 59.10, pp. 1377–1391 (cit. on p. 103).

Dolan, Ray J and Peter Dayan (2013). “Goals and habits in the brain”. In: *Neuron* 80.2, pp. 312–325 (cit. on p. 66).

Doll, Bradley B, Katherine D Duncan, Dylan A Simon, Daphna Shohamy, and Nathaniel D Daw (2015). “Model-based choices involve prospective neural activity”. In: *Nature neuroscience* 18.5, pp. 767–772 (cit. on p. 85).

Dowling, John E (1987). *The retina: an approachable part of the brain*. Harvard University Press (cit. on p. 3).

Droste, Richard, Jianbo Jiao, and J. Alison Noble (2020). “Unified Image and Video Saliency Modeling”. In: *Proceedings of the 16th European Conference on Computer Vision (ECCV)* (cit. on pp. 48, 57).

Duan, Peitong, Casimir Wierzynski, and Lama Nachman (2020). “Optimizing User Interface Layouts via Gradient Descent”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI ’20. Honolulu, HI, USA: Association for Computing Machinery, 1–12 (cit. on p. 105).

Dudley, John J., Jason T. Jacques, and Per Ola Kristensson (2019). “Crowdsourcing Interface Feature Design with Bayesian Optimization”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI ’19. Glasgow, Scotland Uk: Association for Computing Machinery, 1–12 (cit. on p. 106).

Dyson, Mary C (2004). “How physical text layout affects reading from screen”. In: *Behaviour & information technology* 23.6, pp. 377–393 (cit. on pp. 101, 103).

D’Angelo, Nicoletta, Giada Adelfio, Antonino Abbruzzo, and Jorge Mateu (2022). “Inhomogeneous spatio-temporal point processes on linear networks for visitors’ stops data”. In: *The Annals of Applied Statistics* 16.2, pp. 791–815 (cit. on p. 31).

Eckstein, Miguel P, Barbara A Drescher, and Steven S Shimozaki (2006). “Attentional cues in real scenes, saccadic targeting, and Bayesian priors”. In: *Psychological science* 17.11, pp. 973–980 (cit. on p. 19).

Eco, Umberto (1980). *The name of the Rose*. Vintage Classics (cit. on p. 87).

Egorov, Maxim, Zachary N. Sunberg, Edward Balaban, et al. (2017). “POMDPs.jl: A Framework for Sequential Decision Making under Uncertainty”. In: *Journal of Machine Learning Research* 18.26, pp. 1–5 (cit. on pp. 23, 35, 78).

Ernst, Marc O and Heinrich H Bülthoff (2004). “Merging the senses into a robust percept”. In: *Trends in cognitive sciences* 8.4, pp. 162–169 (cit. on p. 18).

Fails, Jerry Alan and Dan R. Olsen (2003). “Interactive Machine Learning”. In: *Proceedings of the 8th International Conference on Intelligent User Interfaces*. IUI ’03. Miami, Florida, USA: Association for Computing Machinery, 39–45 (cit. on p. 105).

Faisal, A. Aldo, Luc P. J. Selen, and Daniel M. Wolpert (2008). “Noise in the nervous system”. In: *Nature Reviews Neuroscience* 9.4, pp. 292–303 (cit. on pp. 3, 4).

Fan, Shaojing, Zhiqi Shen, Ming Jiang, et al. (June 2018). “Emotional Attention: A Study of Image Sentiment and Visual Attention”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (cit. on pp. 46, 48, 54).

Fard, Pouyan R, Hame Park, Andrej Warkentin, Stefan J Kiebel, and Sebastian Bitzer (2017). “A Bayesian reformulation of the extended drift-diffusion model in perceptual decision making”. In: *Frontiers in computational neuroscience* 11, p. 29 (cit. on p. 19).

Farrow, Damian and Bruce Abernethy (2003). “Do expertise and the degree of perception—action coupling affect natural anticipatory performance?” In: *Perception* 32.9, pp. 1127–1139 (cit. on p. 7).

Findlay, John M and Iain D Gilchrist (2003). *Active vision: The psychology of looking and seeing*. 37. Oxford University Press (cit. on pp. 5, 6, 45, 65, 66).

Fischer, B. and E. Ramsperger (1984). “Human express saccades: extremely short reaction times of goal directed eye movements”. In: *Experimental Brain Research* 57.1, pp. 191–195 (cit. on p. 79).

Fishwick, Paul A (2008). *Aesthetic computing*. Cambridge, Massachusetts: Mit Press (cit. on p. 104).

Ford, CC, Guido Bugmann, and Phil Culverhouse (2013). “Modeling the human blink: A computational model for use within human–robot interaction”. In: *International Journal of Humanoid Robotics* 10.01, p. 1350006 (cit. on p. 41).

Foster, David H. (2011). “Color constancy”. In: *Vision Research* 51.7, pp. 674–700 (cit. on p. 2).

Foulsham, Tom, Alan Kingstone, and Geoffrey Underwood (2008). “Turning the world around: Patterns in saccade direction vary with picture orientation”. In: *Vision research* 48.17, pp. 1777–1790 (cit. on pp. 49, 63).

Gajos, Krzysztof Z, Daniel S Weld, and Jacob O Wobbrock (2010). “Automatically generating personalized user interfaces with Supple”. In: *Artificial Intelligence* 174.12-13, pp. 910–950 (cit. on p. 105).

Gajos, Krzysztof Z., Jacob O. Wobbrock, and Daniel S. Weld (2008). “Improving the Performance of Motor-Impaired Users with Automatically-Generated, Ability-Based Interfaces”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’08. Florence, Italy: Association for Computing Machinery, 1257–1266 (cit. on p. 105).

Garcia, Marelys L and Cesar I Caldera (1996). “The effect of color and typeface on the readability of on-line text”. In: *Computers & industrial engineering* 31.1-2, pp. 519–524 (cit. on p. 103).

Ge, Hong, Kai Xu, and Zoubin Ghahramani (2018). "Turing: a language for flexible probabilistic inference". In: *International conference on artificial intelligence and statistics*. PMLR, pp. 1682–1690 (cit. on p. 17).

Geisler, Wilson S (2011). "Contributions of ideal observer theory to vision research". In: *Vision research* 51.7, pp. 771–781 (cit. on p. 4).

Gershman, Samuel J, Eric J Horvitz, and Joshua B Tenenbaum (2015). "Computational rationality: A converging paradigm for intelligence in brains, minds, and machines". In: *Science* 349.6245, pp. 273–278 (cit. on p. 67).

Gezeck, Stefan, Burkhart Fischer, and Jens Timmer (1997). "Saccadic reaction times: a statistical analysis of multimodal distributions". In: *Vision research* 37.15, pp. 2119–2131 (cit. on p. 79).

Gibson, James J (2014). *The ecological approach to visual perception: classic edition*. Psychology press (cit. on pp. 1, 7).

Glimcher, Paul W and Aldo Rustichini (2004). "Neuroeconomics: the consilience of brain and decision". In: *Science* 306.5695, pp. 447–452 (cit. on p. 20).

Gold, Joshua I, Michael N Shadlen, et al. (2007). "The neural basis of decision making". In: *Annual review of neuroscience* 30.1, pp. 535–574 (cit. on p. 67).

Gottlieb, Gilbert (1971). "Ontogenesis of sensory function in birds and mammals". In: *The biopsychology of development* 81 (cit. on p. 1).

Gottlieb, Jacqueline (2012). "Attention, Learning, and the Value of Information". In: *Neuron* 76.2, pp. 281–295 (cit. on p. 5).

Gou, Qifeng, Sunnan Li, and Runping Wang (2022). "Study on eye movement characteristics and intervention of basketball shooting skill". In: *PeerJ* 10, e14301 (cit. on pp. 1, 65).

Gregory, Richard L (1966). *Eye and brain: The psychology of seeing*. Vol. 80. Princeton university press (cit. on p. 65).

Griffiths, Thomas L, Frederick Callaway, Michael B Chang, et al. (2019). "Doing more with less: meta-reasoning and meta-learning in humans and machines". In: *Current Opinion in Behavioral Sciences* 29, pp. 24–30 (cit. on p. 66).

Griffiths, Thomas L and Joshua B Tenenbaum (2009). "Theory-based causal induction." In: *Psychological review* 116.4, p. 661 (cit. on p. 5).

Groen, Yvonne, NA Börger, J Koerts, J Thome, and O Tucha (2017). "Blink rate and blink timing in children with ADHD and the influence of stimulant medication". In: *Journal of Neural Transmission* 124, pp. 27–38 (cit. on p. 30).

Guarnaccia, Claudio, Joseph Quartieri, Carmine Tepedino, and Eliane R Rodrigues (2016). "A time series analysis and a non-homogeneous Poisson model with multiple change-points applied to acoustic data". In: *Applied acoustics* 114, pp. 203–212 (cit. on p. 30).

– (2015). "An analysis of airport noise data using a non-homogeneous Poisson model with a change-point". In: *Applied Acoustics* 91, pp. 33–39 (cit. on p. 30).

Guler Dincer, Nevin, Serdar Demir, and Muhammet Oğuzhan Yalçin (2022). "Forecasting COVID19 reliability of the countries by using non-homogeneous poisson process models". In: *New Generation Computing*, pp. 1–22 (cit. on p. 31).

Harris, Christopher M and Daniel M Wolpert (1998). "Signal-dependent noise determines motor planning". In: *Nature* 394.6695, pp. 780–784 (cit. on pp. 4, 5).

Hayhoe, Mary and Dana Ballard (2005). "Eye movements in natural behavior". In: *Trends in cognitive sciences* 9.4, pp. 188–194 (cit. on pp. 1, 2, 4, 47, 65, 66, 88, 123).

Hayhoe, Mary M and Constantin A Rothkopf (2011). "Vision in the natural world". In: *Wiley Interdisciplinary Reviews: Cognitive Science* 2.2, pp. 158–166 (cit. on pp. 1, 65).

Heesy, Christopher P (2009). "Seeing in stereo: the ecology and evolution of primate binocular vision and stereopsis". In: *Evolutionary Anthropology: Issues, News, and Reviews* 18.1, pp. 21–35 (cit. on p. 1).

Henderson, John M., Taylor R. Hayes, Candace E. Peacock, and Gwendolyn Rehrig (June 2019). "Meaning and Attentional Guidance in Scenes: A Review of the Meaning Map Approach". eng. In: *Vision (Basel, Switzerland)* 3.2. 31735820[pmid], p. 19 (cit. on p. 47).

Henderson, John M and Andrew Hollingworth (1998). "Eye movements during scene viewing: An overview". In: *Eye guidance in reading and scene perception*. Elsevier, pp. 269–293 (cit. on p. 47).

Henderson, John M. and Andrew Hollingworth (1999). "High-level scene perception". In: *Annual review of psychology* 50.1, pp. 243–271 (cit. on p. 2).

Heydt, Rüdiger Von der, Esther Peterhans, and Gunter Baumgartner (1984). "Illusory contours and cortical neuron responses". In: *Science* 224.4654, pp. 1260–1262 (cit. on p. 2).

Hirst, Rebecca J, Lucy Cragg, and Harriet A Allen (2018). "Vision dominates audition in adults but not children: A meta-analysis of the Colavita effect". In: *Neuroscience & Biobehavioral Reviews* 94, pp. 286–301 (cit. on p. 1).

Hoffman, James E and Baskaran Subramaniam (1995). "The role of visual attention in saccadic eye movements". In: *Perception & psychophysics* 57.6, pp. 787–795 (cit. on p. 76).

Holmqvist, Kenneth, Marcus Nyström, Richard Andersson, et al. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford (cit. on p. 76).

Holtmaat, Anthony and Karel Svoboda (2009). "Experience-dependent structural synaptic plasticity in the mammalian brain". In: *Nature Reviews Neuroscience* 10.9, pp. 647–658 (cit. on p. 5).

Hoppe, David, Stefan Helfmann, and Constantin A Rothkopf (2018). "Humans quickly learn to blink strategically in response to environmental task demands". In: *Proceedings of the National Academy of Sciences* 115.9, pp. 2246–2251 (cit. on pp. 4, 29–31, 37–41, 125).

Hoppe, David and Constantin A Rothkopf (2016). "Learning rational temporal eye movement strategies". In: *Proceedings of the National Academy of Sciences* 113.29, pp. 8332–8337 (cit. on pp. 49, 52, 67, 81, 123).

Hoppe, David and Constantin A. Rothkopf (2019). “Multi-step planning of eye movements in visual search”. In: *Scientific reports* 9.1, pp. 1–12 (cit. on pp. 1, 5, 46, 49, 50, 52, 62, 66, 67, 88).

Hu, Changyuan (1998). *Synthesis of parametrisable fonts by shape components*. Tech. rep. EPFL (cit. on pp. 102, 104).

Hu, Changyuan and Roger D Hersch (2001). “Parameterizable fonts based on shape components”. In: *IEEE Computer Graphics and Applications* 21.3, pp. 70–85 (cit. on p. 104).

Huang, Xun, Chengyao Shen, Xavier Boix, and Qi Zhao (Dec. 2015). “SALICON: Reducing the Semantic Gap in Saliency Prediction by Adapting Deep Neural Networks”. en. In: *2015 IEEE International Conference on Computer Vision (ICCV)*. Santiago, Chile: IEEE, pp. 262–270 (cit. on pp. 46, 47).

Hubel, David H and Torsten N Wiesel (1962). “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex”. In: *The Journal of physiology* 160.1, p. 106 (cit. on p. 3).

Hurst, Nathan, Wilmot Li, and Kim Marriott (2009). “Review of Automatic Document Formatting”. In: *Proceedings of the 9th ACM Symposium on Document Engineering*. DocEng ’09. Munich, Germany: Association for Computing Machinery, 99–108 (cit. on p. 104).

Hutmacher, Fabian (2019). “Why Is There So Much More Research on Vision Than on Any Other Sensory Modality?” In: *Frontiers in Psychology* 10 (cit. on p. 1).

Huys, Quentin JM, Neir Eshel, Elizabeth O’Nions, et al. (2012). “Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees”. In: *PLoS computational biology* 8.3, e1002410 (cit. on pp. 23, 66, 88).

Itti, L., C. Koch, and E. Niebur (Nov. 1998). “A model of saliency-based visual attention for rapid scene analysis”. en. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20.11, pp. 1254–1259 (cit. on pp. 8, 46–48).

Itti, Laurent and Christof Koch (2000). “A saliency-based search mechanism for overt and covert shifts of visual attention”. In: *Vision research* 40.10-12, pp. 1489–1506 (cit. on pp. 2, 67).

Jacobs, Charles, Wilmot Li, Evan Schrier, David Bargeron, and David Salesin (2003). “Adaptive grid-based document layout”. In: *ACM transactions on graphics (TOG)* 22.3, pp. 838–847 (cit. on p. 104).

Järlehed, Johan and Adam Jaworski (2015). *Typographic landscaping: Creativity, ideology, movement* (cit. on p. 102).

Jarodzka, Halszka, Kenneth Holmqvist, and Marcus Nyström (2010). “A vector-based, multidimensional scanpath similarity measure”. In: *Proceedings of the 2010 symposium on eye-tracking research & applications*, pp. 211–218 (cit. on p. 48).

JASP Team (2020). *JASP (Version 0.13.1)[Computer software]* (cit. on p. 107).

Jia, Sen and Neil D.B. Bruce (2020). “EML-NET: An Expandable Multi-Layer NETwork for saliency prediction”. In: *Image and Vision Computing* 95, p. 103887 (cit. on pp. 46, 48, 54).

Jiang, Ming, Xavier Boix, Gemma Roig, et al. (June 2016). “Learning to Predict Sequences of Human Visual Fixations”. en. In: *IEEE Transactions on Neural Networks and Learning Systems* 27.6, pp. 1241–1252 (cit. on pp. 46, 49, 62).

Jiang, Ming, Shengsheng Huang, Juanyong Duan, and Qi Zhao (June 2015). “SALICON: Saliency in Context”. en. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA, USA: IEEE, pp. 1072–1080 (cit. on p. 46).

Johansson, Roland S, Göran Westling, Anders Bäckström, and J Randall Flanagan (2001). “Eye–hand coordination in object manipulation”. In: *Journal of neuroscience* 21.17, pp. 6917–6932 (cit. on p. 65).

Jokinen, Jussi P. P., Tuomo Kujala, and Antti Oulasvirta (2021). “Multitasking in Driving as Optimal Adaptation Under Uncertainty”. In: *Human Factors* 63.8. PMID: 32731763, pp. 1324–1341 (cit. on p. 7).

Jones, Joshua L, Guillem R Esber, Michael A McDannald, et al. (2012). “Orbitofrontal cortex supports behavior and learning using inferred but not cached values”. In: *Science* 338.6109, pp. 953–956 (cit. on p. 85).

Jonsdottir, Kristin, Mattias Lindman, Roland Roberts, Björn Lund, and Reynir Bödvarsson (2006). “Modelling fundamental waiting time distributions for earthquake sequences”. In: *Tectonophysics* 424.3-4, pp. 195–208 (cit. on p. 31).

Judd, T., K. Ehinger, F. Durand, and A. Torralba (2009). “Learning to predict where humans look”. In: *2009 IEEE 12th International Conference on Computer Vision*, pp. 2106–2113 (cit. on pp. 56, 57).

Judd, Tilke, Frédo Durand, and Antonio Torralba (2012). “A Benchmark of Computational Models of Saliency to Predict Human Fixations”. In: *MIT Technical Report* (cit. on pp. 55, 58).

Jütte, Robert (2005). *A History of the Senses: from Antiquity to Cyberspace*. Polity (cit. on p. 1).

Kadner, Florian, Yannik Keller, and Constantin A. Rothkopf (2021). “AdaptiFont: Increasing Individuals’ Reading Speed with a Generative Font Model and Bayesian Optimization”. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI ’21. Yokohama, Japan: Association for Computing Machinery (cit. on p. 11).

Kadner, Florian, Tobias Thomas, David Hoppe, and Constantin A. Rothkopf (Jan. 2023a). “Improving Saliency Models’ Predictions of the Next Fixation With Humans’ Intrinsic Cost of Gaze Shifts”. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 2104–2114 (cit. on pp. 10, 31, 67, 88).

Kadner, Florian, Tabea A. Wilke, Thi DK Vo, David Hoppe, and Constantin A. Rothkopf (2022). “Trade-off between uncertainty reduction and reward collection reveals intrinsic cost of gaze switches”. In: *Journal of Vision* 22.14, pp. 3400–3400 (cit. on pp. 11, 31).

Kadner, Florian, Hannah Willkomm, Inga Ibs, and Constantin Rothkopf (2023b). “Finding your Way Out: Planning Strategies in Human Maze-Solving Behavior”. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 45. 45 (cit. on p. 11).

Kaelbling, Leslie Pack, Michael L. Littman, and Anthony R. Cassandra (1998). “Planning and acting in partially observable stochastic domains”. In: *Artificial Intelligence* 101.1, pp. 99–134 (cit. on pp. 21, 24, 31, 35, 66, 78).

Kanagal, Bhargav and Vikas Sindhwani (Jan. 2010). “Rank Selection in Low-rank Matrix Approximations: A Study of Cross-Validation for NMFs”. In: *Proc Conf Adv Neural Inf Process* 1 (cit. on p. 108).

Kaplan, Raphael and Karl J. Friston (Aug. 2018). “Planning and navigation as active inference”. In: *Biological Cybernetics* 112.4, pp. 323–343 (cit. on p. 88).

Kersten, Daniel, Pascal Mamassian, and Alan Yuille (2004). “Object perception as Bayesian inference”. In: *Annu. Rev. Psychol.* 55, pp. 271–304 (cit. on p. 18).

Kessler, Fabian, Julia Frankenstein, and Constantin A. Rothkopf (2022). “A Dynamic Bayesian Actor Model explains Endpoint Variability in Homing Tasks”. In: *bioRxiv* (cit. on pp. 5, 27, 87).

Khajah, Mohammad M., Brett D. Roads, Robert V. Lindsey, Yun-En Liu, and Michael C. Mozer (2016). “Designing Engaging Games Using Bayesian Optimization”. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI ’16. San Jose, California, USA: Association for Computing Machinery, 5571–5582 (cit. on p. 106).

Kieras, David E and Anthony J Hornof (2014). “Towards accurate and practical predictive models of active-vision-based visual search”. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 3875–3884 (cit. on p. 7).

Kim, Paul Hyunjin and Roger Crawfis (2018). “Intelligent maze generation based on topological constraints”. In: *2018 7th International Congress on Advanced Applied Informatics (IIAI-AAI)*. IEEE, pp. 867–872 (cit. on pp. 87, 90, 92).

– (2015). “The quest for the perfect perfect-maze”. In: *2015 Computer Games: AI, Animation, Mobile, Multimedia, Educational and Serious Games (CGAMES)*. IEEE, pp. 65–72 (cit. on p. 88).

Knill, David C and Alexandre Pouget (2004). “The Bayesian brain: the role of uncertainty in neural coding and computation”. In: *TRENDS in Neurosciences* 27.12, pp. 712–719 (cit. on pp. 3, 18).

Knuth, Donald Ervin (1999). *Digital typography*. Vol. 78. Stanford, California: Csli Publications (cit. on p. 104).

Kochenderfer, Mykel J, Tim A Wheeler, and Kyle H Wray (2022). *Algorithms for decision making*. MIT press (cit. on p. 13).

König, Peter, Niklas Wilming, Tim C Kietzmann, et al. (2016). “Eye movements as a window to cognitive processes”. In: *Journal of Eye Movement Research* 9.5, pp. 1–16 (cit. on p. 88).

Köpper, Maja, Susanne Mayr, and Axel Buchner (2016). “Reading from computer screen versus reading from paper: does it still make a difference?” In: *Ergonomics* 59.5, pp. 615–632 (cit. on p. 104).

Körding, Konrad P and Daniel M Wolpert (2006). “Bayesian decision theory in sensorimotor control”. In: *Trends in cognitive sciences* 10.7, pp. 319–326 (cit. on p. 6).

– (2004). “Bayesian integration in sensorimotor learning”. In: *Nature* 427.6971, pp. 244–247 (cit. on pp. 5, 18).

Korinth, Sebastian P., Kerstin Gerstenberger, and Christian J. Fiebach (2020). “Wider Letter-Spacing Facilitates Word Processing but Impairs Reading Rates of Fast Readers”. In: *Frontiers in Psychology* 11, p. 444 (cit. on p. 104).

Kowler, Eileen (2011). “Eye movements: The past 25 years”. In: *Vision research* 51.13, pp. 1457–1483 (cit. on pp. 2, 123).

Koyama, Yuki and Takeo Igarashi (2018). “Computational Design with Crowds”. eng. In: Computational Interaction. Oxford: Oxford University Press. Chap. 7, pp. 153–184 (cit. on p. 106).

Kredel, Ralf, Christian Vater, André Klostermann, and Ernst-Joachim Hossner (2017). “Eye-tracking technology and the dynamics of natural gaze behavior in sports: A systematic review of 40 years of research”. In: *Frontiers in psychology* 8, p. 1845 (cit. on pp. 1, 65).

Kruthiventi, Srinivas S. S., Kumar Ayush, and R. Venkatesh Babu (Oct. 2015). “DeepFix: A Fully Convolutional Neural Network for predicting Human Eye Fixations”. en. In: *arXiv:1510.02927 [cs]*. arXiv: 1510.02927 (cit. on pp. 46, 47).

Kryven, Marta, Max Kleiman-Weiner, Joshua Tenenbaum, and Suhyoun Yu (2022). “Planning ahead in spatial search”. In: *PsyarXiv preprint* (cit. on p. 87).

Kümmerer, Matthias and Matthias Bethge (2021). “State-of-the-Art in Human Scanpath Prediction”. In: *arXiv preprint arXiv:2102.12239* (cit. on pp. 46, 48, 58, 62).

Kümmerer, Matthias, Thomas S. A. Wallis, and Matthias Bethge (2018). “Saliency Benchmarking Made Easy: Separating Models, Maps and Metrics”. In: *Computer Vision – ECCV 2018*. Ed. by Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss. Lecture Notes in Computer Science. Springer International Publishing, pp. 798–814 (cit. on pp. 55, 58).

Kümmerer, M., T. S. A. Wallis, L. A. Gatys, and M. Bethge (2017). “Understanding Low- and High-Level Contributions to Fixation Prediction”. In: *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 4799–4808 (cit. on pp. 46–48, 54).

Kümmerer, Matthias, Zoya Bylinskii, Tilke Judd, et al. (n.d.). *MIT/Tübingen Saliency Benchmark* (cit. on p. 54).

Land, Michael, Neil Mennie, and Jennifer Rusted (1999). “The Roles of Vision and Eye Movements in the Control of Activities of Daily Living”. In: *Perception* 28.11. PMID: 10755142, pp. 1311–1328 (cit. on pp. 1, 6, 65).

Land, Michael and Benjamin Tatler (2009). *Looking and acting: vision and eye movements in natural behaviour*. Oxford University Press (cit. on pp. 2, 7, 65, 123).

Land, Michael F (1995). “The functions of eye movements in animals remote from man”. In: *Studies in visual information processing*. Vol. 6. Elsevier, pp. 63–76 (cit. on p. 6).

Land, Michael F. (2009). “Vision, eye movements, and natural behavior”. In: *Visual Neuroscience* 26.1, 51–62 (cit. on p. 1).

Land, Michael F and David N Lee (1994). “Where we look when we steer”. In: *Nature* 369.6483, pp. 742–744 (cit. on pp. 1, 7, 65).

Land, Michael F and Peter McLeod (2000). “From eye movements to actions: how batsmen hit the ball”. In: *Nature neuroscience* 3.12, pp. 1340–1345 (cit. on p. 65).

Land, Michael F and Dan-Eric Nilsson (2012). *Animal eyes*. OUP Oxford (cit. on p. 6).

Landolt, C (2019). “Visual factors in reading”. In: *Vision Research* 161, pp. 60–62 (cit. on pp. 101, 103).

Lappi, Otto (2014). “Future path and tangent point models in the visual control of locomotion in curve driving”. In: *Journal of vision* 14.12, pp. 21–21 (cit. on p. 7).

Latin, Matej (2019). *Better web typography for a better web*. San Francisco: BLURB (cit. on p. 104).

Le Meur, Olivier and Thierry Baccino (Mar. 2013). “Methods for comparing scanpaths and saliency maps: strengths and weaknesses”. en. In: *Behavior Research Methods* 45.1, pp. 251–266 (cit. on pp. 55, 58).

Le Meur, Olivier, Antoine Coutrot, Adrien Le Roch, et al. (2017). “Age-dependent saccadic models for predicting eye movements”. In: *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, pp. 3740–3744 (cit. on p. 63).

Legge, Gordon E. and Charles A. Bigelow (Aug. 2011). “Does print size matter for reading? A review of findings from vision science and typography”. In: *Journal of Vision* 11.5, pp. 8–8 (cit. on pp. 7, 103).

Legge, Gordon E, Denis G Pelli, Gar S Rubin, and Mary M Schleske (1985). “Psychophysics of reading—I. Normal vision”. In: *Vision research* 25.2, pp. 239–252 (cit. on pp. 101, 103).

Leigh, R John and David S Zee (2015). *The neurology of eye movements*. Contemporary Neurology (cit. on pp. 4, 123).

Lennie, Peter (2003). “The cost of cortical computation”. In: *Current biology* 13.6, pp. 493–497 (cit. on p. 6).

Lenskiy, Artem and Rafal Paprocki (2016). “Blink rate variability during resting and reading sessions”. In: *2016 IEEE Conference on Norbert Wiener in the 21st Century (21CW)*. IEEE, pp. 1–6 (cit. on p. 30).

Li, Jason, Nicholas Watters, Wang Yingting, Hansem Sohn, and Mehrdad Jazayeri (2022). *Modeling Human Eye Movements with Neural Networks in a Maze-Solving Task* (cit. on p. 88).

Lim, Kar Wai, W Wang, H Nguyen, et al. (2016). “Traffic flow modelling with point processes”. In: *Proceedings of the 23rd World Congress on Intelligent Transport Systems*, pp. 1–12 (cit. on p. 31).

Lisi, Matteo, Joshua A. Solomon, and Michael J. Morgan (Aug. 2019). “Gain control of saccadic eye movements is probabilistic”. en. In: *Proceedings of the National Academy of Sciences* 116.32, pp. 16137–16142 (cit. on p. 67).

Litjens, Geert, Thijs Kooi, Babak Ehteshami Bejnordi, et al. (2017). “A survey on deep learning in medical image analysis”. In: *Medical image analysis* 42, pp. 60–88 (cit. on p. 8).

Littman, Michael L., Anthony R. Cassandra, and Leslie Pack Kaelbling (1995). “Learning policies for partially observable environments: Scaling up”. In: *Proceedings of the Twelfth International Conference on Machine Learning*. Morgan Kaufmann, pp. 362–370 (cit. on pp. 25, 35, 79).

Liu, Huiying, Dong Xu, Qingming Huang, et al. (Dec. 2013). “Semantically-Based Human Scanpath Estimation with HMMs”. en. In: *2013 IEEE International Conference on Computer Vision*. Sydney, Australia: IEEE, pp. 3232–3239 (cit. on pp. 46, 48).

Liversedge, Simon P and John M Findlay (2000). “Saccadic eye movements and cognition”. In: *Trends in cognitive sciences* 4.1, pp. 6–14 (cit. on p. 76).

Lorenz, Romy, Ricardo Pio Monti, Inês R Violante, et al. (2016). “The Automatic Neuroscientist: A framework for optimizing experimental design with closed-loop real-time fMRI”. In: *NeuroImage* 129, pp. 320–334 (cit. on p. 106).

Lu, Yi and José Garrido (2005). “Doubly periodic non-homogeneous Poisson models for hurricane data”. In: *Statistical Methodology* 2.1, pp. 17–35 (cit. on p. 30).

Luckiesh, Matthew and Fbank K Moss (1940). “Boldness as a factor in type-design and typography.” In: *Journal of Applied Psychology* 24.2, p. 170 (cit. on p. 104).

Ma, Ili, Wei Ji Ma, and Todd M Gureckis (2021). “Information sampling for contingency planning.” In: *Proceedings of the 43rd Annual Conference of the Cognitive Science Society* (cit. on pp. 6, 88).

Ma, Wei Ji, Konrad Paul Kording, and Daniel Goldreich (2023). *Bayesian Models of Perception and Action: An Introduction*. MIT press (cit. on p. 14).

MacIver, Malcolm A, Lars Schmitz, Ugurcan Mugan, Todd D Murphey, and Curtis D Mobley (2017). “Massive increase in visual range preceded the origin of terrestrial vertebrates”. In: *Proceedings of the National Academy of Sciences* 114.12, E2375–E2384 (cit. on p. 85).

Mahmud, M. M. Hassan, Benjamin Rosman, Subramanian Ramamoorthy, and Pushmeet Kohli (July 2014). “Adapting Interaction Environments to Diverse Users through Online Action Set Selection”. In: *Proceedings of the AAAI 2014 Workshop on Machine Learning for Interactive Systems*. .: Association for the Advancement of ArtificialIntelligence (cit. on p. 105).

Maloney, Laurence T and Pascal Mamassian (2009). “Bayesian decision theory as a model of human visual perception: Testing Bayesian transfer”. In: *Visual neuroscience* 26.1, pp. 147–155 (cit. on p. 20).

Mann, Derek TY, A Mark Williams, Paul Ward, and Christopher M Janelle (2007). “Perceptual-cognitive expertise in sport: A meta-analysis”. In: *Journal of sport and exercise psychology* 29.4, pp. 457–478 (cit. on p. 7).

Mansfield, J Stephen, Gordon E Legge, and Mark C Bane (1996). “Psychophysics of reading. XV: Font effects in normal and low vision.” In: *Investigative Ophthalmology & Visual Science* 37.8, pp. 1492–1501 (cit. on p. 104).

Marr, David (2010). *Vision: A computational investigation into the human representation and processing of visual information*. MIT press (cit. on p. 6).

Mathe, Stefan and Cristian Sminchisescu (2013). "Action from still image dataset and inverse optimal control to learn task specific visual scanpaths". In: *Advances in neural information processing systems*, pp. 1923–1931 (cit. on pp. 46, 49, 62).

Mattar, Marcelo G. and Máté Lengyel (2022). "Planning in the brain". In: *Neuron* 110.6, pp. 914–934 (cit. on p. 66).

Mauk, Michael D, Dean V Buonomano, et al. (2004). "The neural basis of temporal processing". In: *Annual review of neuroscience* 27.1, pp. 307–340 (cit. on pp. 81, 123).

McLean, Michael V (1965). "Brightness contrast, color contrast, and legibility". In: *Human factors* 7.6, pp. 521–527 (cit. on pp. 7, 103).

McQueen III, Clyde D and Raymond G Beausoleil (1993). "Infinifont: a parametric font generation system". In: *ELECTRONIC PUBLISHING-CHICHESTER-* 6, pp. 117–117 (cit. on p. 104).

Microsoft (Mar. 2018). *OpenType® specification* (cit. on pp. 102, 104).

Miller, George A, Galanter Eugene, and Karl H Pribram (2017). "Plans and the Structure of Behaviour". In: *Systems Research for Behavioral Science*. Routledge, pp. 369–382 (cit. on p. 124).

Miller, George Armitage (1991). *The science of words*. .: Scientific American Library (cit. on p. 101).

Mnih, Volodymyr, Nicolas Heess, Alex Graves, and Koray Kavukcuoglu (2014). "Recurrent Models of Visual Attention". In: *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*. NIPS'14. Montreal, Canada: MIT Press, 2204–2212 (cit. on pp. 47, 49).

Moraitis, Timoleon and Arko Ghosh (2014). "Withdrawal of voluntary inhibition unravels the off state of the spontaneous blink generator". In: *Neuropsychologia* 65, pp. 279–286 (cit. on p. 41).

Mugan, Ugurcan and Malcolm A MacIver (2020). "Spatial planning with long visual range benefits escape from visual predators in complex naturalistic environments". In: *Nature communications* 11.1, pp. 1–14 (cit. on p. 85).

Mugumaarhahama, Yannick, Adande Belarmain Fandohan, Arsene Ciza Mushagalusa, Idelphonse Akoeugnigan Sode, and Romain L Glele Kakai (2022). "Inhomogeneous Poisson point process for species distribution modelling: relative performance of methods accounting for sampling bias and imperfect detection". In: *Modeling Earth Systems and Environment* 8.4, pp. 5419–5432 (cit. on p. 30).

Murphy, Kevin P. (2022). *Probabilistic Machine Learning: An introduction*. MIT Press (cit. on p. 14).

Naase, Taher, Michael J Doughty, and Norman F Button (2005). “An assessment of the pattern of spontaneous eyeblink activity under the influence of topical ocular anaesthesia”. In: *Graefe’s Archive for Clinical and Experimental Ophthalmology* 243, pp. 306–312 (cit. on p. 30).

Najemnik, Jiri and Wilson S. Geisler (Mar. 2008). “Eye movement statistics in humans are consistent with an optimal search strategy”. In: *Journal of Vision* 8.3, pp. 4–4 (cit. on pp. 2, 19).

Najemnik, Jiri and Wilson S Geisler (2005). “Optimal eye movement strategies in visual search”. In: *Nature* 434.7031, pp. 387–391 (cit. on pp. 50, 67).

Nakano, Tamami and Yuta Miyazaki (2019). “Blink synchronization is an indicator of interest while viewing videos”. In: *International Journal of Psychophysiology* 135, pp. 1–11 (cit. on p. 30).

Nakano, Tamami, Atsuya Sakata, and Akihiro Kishimoto (2020). “Estimating blink probability for highlight detection in figure skating videos”. In: *arXiv preprint arXiv:2007.01089* (cit. on pp. 30, 41).

Nakano, Tamami, Yoshiharu Yamamoto, Keiichi Kitajo, Toshimitsu Takahashi, and Shigeru Kitazawa (2009). “Synchronization of spontaneous eyeblinks while viewing video stories”. In: *Proceedings of the Royal Society B: Biological Sciences* 276.1673, pp. 3635–3644 (cit. on p. 30).

Nakayama, Ken (1990). “The iconic bottleneck and the tenuous link between early visual processing and perception”. In: *Vision: Coding and efficiency* 411422 (cit. on p. 6).

Nake, Frieder (2009). “The semiotic engine: notes on the history of algorithmic images in Europe”. In: *Art Journal* 68.1, pp. 76–89 (cit. on p. 104).

Navalpakkam, Vidhya and Laurent Itti (2005). “Modeling the influence of task on attention”. In: *Vision research* 45.2, pp. 205–231 (cit. on p. 47).

Neupärtl, Nils and Constantin A Rothkopf (2021). “Inferring perceptual decision making parameters from behavior in production and reproduction tasks”. In: *arXiv preprint arXiv:2112.15521* (cit. on p. 5).

Ngailo, Triphonia, Nyimvua Shaban, Joachim Reuder, Edwin Rutalebwa, and Isaac Mugume (2016). “Non homogeneous poisson process modelling of seasonal extreme rainfall events in Tanzania”. In: *International journal of science and research (IJSR)* 5.10, pp. 1858–1868 (cit. on p. 30).

Nin, A. (1961). *Seduction of the Minotaur*. A Swallow paperbook. Swallow Press (cit. on p. 13).

Nishizono, Ryota, Naoki Saijo, and Makio Kashino (2023). “Highly reproducible eyeblink timing during formula car driving”. In: *iScience* (cit. on p. 30).

Nogueira, Fernando (2014). *Bayesian Optimization: Open source constrained global optimization tool for Python* (cit. on p. 114).

Noy, Y Ian, Tracy L Lemoine, Christopher Klachan, and Peter C Burns (2004). “Task interruptability and duration as measures of visual distraction”. In: *Applied Ergonomics* 35.3, pp. 207–213 (cit. on p. 7).

Ognibene, Dimitri and Yiannis Demiris (2013). “Towards Active Event Recognition.” In: *IJCAI*, pp. 2495–2501 (cit. on p. 7).

Opheusden, Bas van, Ionatan Kuperwajs, Gianni Galbiati, et al. (2023). “Expertise increases planning depth in human gameplay”. In: *Nature*, pp. 1–6 (cit. on p. 66).

O’Regan, J. Kevin and Alva Noë (2001). “A sensorimotor account of vision and visual consciousness”. In: *Behavioral and Brain Sciences* 24.5, pp. 939–973 (cit. on p. 3).

Osterer, Heidrun and Philipp Stamm (2014). *Adrian Frutiger–Typefaces: The Complete Works*. Basel: Walter de Gruyter (cit. on p. 101).

Padoa-Schioppa, Camillo and John A Assad (2006). “Neurons in the orbitofrontal cortex encode economic value”. In: *Nature* 441.7090, pp. 223–226 (cit. on p. 67).

Palmer, Stephen E (1999). *Vision science: Photons to phenomenology*. MIT press (cit. on p. 3).

Pan, Junting, Elisa Sayrol, Xavier Giro-I-Nieto, Kevin McGuinness, and Noel E. OConnor (June 2016). “Shallow and Deep Convolutional Networks for Saliency Prediction”. en. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE, pp. 598–606 (cit. on pp. 46, 47).

Park, Song, Sanghyuk Chun, Junbum Cha, Bado Lee, and Hyunjung Shim (2020). *Few-shot Font Generation with Localized Style Representations and Factorization*. arXiv: 2009.11042 [cs.CV] (cit. on pp. 105, 108).

Paterson, Donald G and Miles A Tinker (1931). “Studies of typographical factors influencing speed of reading.” In: *Journal of Applied Psychology* 15.3, p. 241 (cit. on pp. 103, 123).

Pedziwiatr, Marek A, Thomas SA Wallis, Matthias Kümmerer, and Christoph Teufel (2019). “Meaning maps and deep neural networks are insensitive to meaning when predicting human fixations”. In: *Journal of Vision* 19.10. Publisher: The Association for Research in Vision and Ophthalmology, pp. 253c–253c (cit. on p. 47).

Peeling, Paul, Chung-fai Li, and Simon Godsill (2007). “Poisson point process modeling for polyphonic music transcription”. In: *The Journal of the Acoustical Society of America* 121.4, EL168–EL175 (cit. on p. 30).

Pelli, Denis G and Bart Farell (1999). “Why use noise?” In: *JOSA A* 16.3, pp. 647–653 (cit. on p. 4).

Pelli, Denis G., Katharine A. Tillman, Jeremy Freeman, et al. (Oct. 2007). “Crowding and eccentricity determine reading rate”. In: *Journal of Vision* 7.2, pp. 20–20 (cit. on p. 103).

Peters, Robert J., Asha Iyer, Laurent Itti, and Christof Koch (2005). “Components of bottom-up gaze allocation in natural images”. In: *Vision Research* 45.18, pp. 2397–2416 (cit. on p. 55).

Peterson, Matthew F and Miguel P Eckstein (2012). “Looking just below the eyes is optimal across face recognition tasks”. In: *Proceedings of the National Academy of Sciences* 109.48, E3314–E3323 (cit. on pp. 50, 67).

Petitet, Pierre, Bahaaeddin Attaallah, Sanjay G. Manohar, and Masud Husain (2021). “The computational cost of active information sampling before decision-making under uncertainty”. In: *Nature Human Behaviour* 5.7, pp. 935–946 (cit. on pp. 41, 67, 125).

Pfeifer, Rolf and Josh Bongard (2006). *How the body shapes the way we think: a new view of intelligence*. MIT press (cit. on p. 8).

Pineau, Joelle, Geoff Gordon, Sebastian Thrun, et al. (2003). “Point-based value iteration: An anytime algorithm for POMDPs”. In: *Ijcai*. Vol. 3, pp. 1025–1032 (cit. on p. 25).

Ponder, Eric and WP Kennedy (1927). “On the act of blinking”. In: *Quarterly journal of experimental physiology: Translation and integration* 18.2, pp. 89–110 (cit. on pp. 30, 38).

Posner, Michael I, Yoav Cohen, et al. (1984). “Components of visual orienting”. In: *Attention and performance X: Control of language processes* 32, pp. 531–556 (cit. on p. 2).

Posner, Michael I, Mary J Nissen, and Raymond M Klein (1976). “Visual dominance: an information-processing account of its origins and significance.” In: *Psychological review* 83.2, p. 157 (cit. on p. 1).

Purvis, Lisa, Steven Harrington, Barry O’Sullivan, and Eugene C. Freuder (2003). “Creating Personalized Documents: An Optimization Approach”. In: *Proceedings of the 2003 ACM Symposium on Document Engineering*. DocEng ’03. Grenoble, France: Association for Computing Machinery, 68–77 (cit. on p. 105).

Raffert, Anna, Matei Zaharia, and Thomas Griffiths (2012). “Optimally designing games for cognitive science research”. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. 34. Sapporo, Japan: Cognitive Science Society, pp. 280–287 (cit. on p. 105).

Ranti, Carolyn, Warren Jones, Ami Klin, and Sarah Shultz (2020). “Blink rate patterns provide a reliable measure of individual engagement with scene content”. In: *Scientific reports* 10.1, p. 8267 (cit. on p. 30).

Rayner, Keith (1998). “Eye movements in reading and information processing: 20 years of research.” In: *Psychological bulletin* 124.3, p. 372 (cit. on pp. 7, 8, 76).

Redish, A David (2016). “Vicarious trial and error”. In: *Nature Reviews Neuroscience* 17.3, pp. 147–159 (cit. on p. 66).

Reise, Steven P and Niels G Waller (2009). “Item response theory and clinical measurement”. In: *Annual review of clinical psychology* 5, pp. 27–48 (cit. on p. 4).

Rello, Luz, Martin Pielot, and Mari-Carmen Marcos (2016). “Make It Big! The Effect of Font Size and Line Spacing on Online Readability”. In: CHI ’16. San Jose, California, USA: Association for Computing Machinery, 3637–3648 (cit. on p. 103).

Reynolds, AM (2010). “Maze-solving by chemotaxis”. In: *Physical Review E* 81.6, p. 062901 (cit. on p. 91).

Riggs, Lorrin A, Frances C Volkmann, and Robert K Moore (1981). “Suppression of the blackout due to blinks”. In: *Vision Research* 21.7, pp. 1075–1079 (cit. on p. 29).

Robinson, D. A. (1964). “The mechanics of human saccadic eye movement”. eng. In: *The Journal of physiology* 174.2, pp. 245–264 (cit. on p. 79).

Rosenberg, Matthew, Tony Zhang, Pietro Perona, and Markus Meister (June 2021). “Mice in a labyrinth show rapid learning, sudden insight, and efficient exploration”. In: *eLife* 10. Ed. by Mackenzie W Mathis, Catherine Dulac, and Gordon J Berman, e66175 (cit. on p. 87).

Rossum, Just van (2017). *FontTools* (cit. on p. 110).

Rothkopf, Constantin A., Dana H. Ballard, and Mary M. Hayhoe (2016). “Task and context determine where you look”. In: *Journal of Vision* 7.14, pp. 16–16 (cit. on p. 7).

Roy, Prasun, Saumik Bhattacharya, Subhankar Ghosh, and Umapada Pal (June 2020). “STEFANN: Scene Text Editor using Font Adaptive Neural Network”. In: *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Virtual: IEEE (cit. on pp. 105, 108).

Russell, Stuart and Peter Norvig (2021). “Artificial Intelligence: a modern approach, 4th US ed”. In: *University of California, Berkeley* (cit. on p. 66).

Russell-Minda, Elizabeth, Jeffrey W Jutai, J Graham Strong, et al. (2007). “The legibility of typefaces for readers with low vision: A research review”. In: *Journal of Visual Impairment & Blindness* 101.7, pp. 402–415 (cit. on p. 103).

Ryali, Chaitanya K, Stanny Goffin, Piotr Winkielman, and Angela J Yu (2020). “From likely to likable: The role of statistical typicality in human social assessment of faces”. In: *Proceedings of the National Academy of Sciences* 117.47, pp. 29371–29380 (cit. on p. 2).

Salcedo, Rodolfo N, Hadley Read, James F Evans, and Ana C Kong (1972). “A broader look at legibility”. In: *Journalism Quarterly* 49.2, pp. 285–295 (cit. on p. 104).

Saldanha, Pedro LC, Elaine A De Simone, and PF Frutuoso e Melo (2001). “An application of non-homogeneous Poisson point processes to the reliability analysis of service water pumps”. In: *Nuclear engineering and design* 210.1-3, pp. 125–133 (cit. on p. 30).

Salvucci, Dario D and Joseph H Goldberg (2000). “Identifying fixations and saccades in eye-tracking protocols”. In: *Proceedings of the 2000 symposium on Eye tracking research & applications*, pp. 71–78 (cit. on p. 76).

Sander, Jörg, Martin Ester, Hans-Peter Kriegel, and Xiaowei Xu (June 1998). “Density-Based Clustering in Spatial Databases: The Algorithm GDBSCAN and Its Applications”. In: *Data Min. Knowl. Discov.* 2.2, 169–194 (cit. on p. 116).

Schmidt, Brandy, Anneke A Duin, and A David Redish (2019). “Disrupting the medial prefrontal cortex alters hippocampal sequences during deliberative decision making”. In: *Journal of neurophysiology* 121.6, pp. 1981–2000 (cit. on p. 85).

Schrittwieser, Julian, Ioannis Antonoglou, Thomas Hubert, et al. (2020). “Mastering atari, go, chess and shogi by planning with a learned model”. In: *Nature* 588.7839, pp. 604–609 (cit. on p. 67).

Selinger, Peter (2001). *Potrace*. http://potrace.sourceforge.net/ (cit. on p. 110).

Shahriari, Bobak, Kevin Swersky, Ziyu Wang, Ryan P Adams, and Nando De Freitas (2015). "Taking the human out of the loop: A review of Bayesian optimization". In: *Proceedings of the IEEE* 104.1, pp. 148–175 (cit. on p. 106).

Shamir, Ariel and Ari Rappoport (1998). "Feature-based design of fonts using constraints". In: *Electronic Publishing, Artistic Imaging, and Digital Typography*. Ed. by Roger D. Hersch, Jacques André, and Heather Brown. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 93–108 (cit. on p. 104).

Shcherbakov, Robert, Gleb Yakovlev, Donald L Turcotte, and John B Rundle (2005). "Model for the distribution of aftershock interoccurrence times". In: *Physical review letters* 95.21, p. 218501 (cit. on p. 31).

Sheedy, Jim, Yu-Chi Tai, Manoj Subbaram, Sowjanya Gowrisankaran, and John Hayes (2008). "ClearType sub-pixel text rendering: Preference, legibility and reading performance". In: *Displays* 29.2, pp. 138–151 (cit. on pp. 102, 104).

Shibata, Kazuya, Koichiro Rinsaka, and Tadashi Dohi (2006). "Metrics-based software reliability models using non-homogeneous Poisson processes". In: *2006 17th International Symposium on Software Reliability Engineering*. IEEE, pp. 52–61 (cit. on p. 30).

Shin, Young Seok, Won-du Chang, Jinsick Park, et al. (2015). "Correlation between interblink interval and episodic encoding during movie watching". In: *PloS one* 10.11, e0141242 (cit. on p. 30).

Siegrist, Kyle (2021). *Probability, Mathematical Statistics, Stochastic Processes*. LibreTexts Statistics (cit. on p. 35).

Silver, David, Satinder Singh, Doina Precup, and Richard S Sutton (2021). "Reward is enough". In: *Artificial Intelligence* 299, p. 103535 (cit. on p. 81).

Silver, David and Joel Veness (2010). "Monte-Carlo planning in large POMDPs". In: *Advances in neural information processing systems* 23 (cit. on p. 25).

Silverman, Irwin (2010). "Simple reaction time: It is not what it used to be". In: *The American journal of psychology* 123, pp. 39–50 (cit. on p. 79).

Simon, Herbert A (1955). "A behavioral model of rational choice". In: *The quarterly journal of economics* 69, pp. 99–118 (cit. on p. 67).

Simons, Daniel J and Daniel T Levin (1997). "Change blindness". In: *Trends in cognitive sciences* 1.7, pp. 261–267 (cit. on p. 2).

Simonyan, Karen and Andrew Zisserman (Apr. 2015). "Very Deep Convolutional Networks for Large-Scale Image Recognition". en. In: *arXiv:1409.1556 [cs]*. arXiv: 1409.1556 (cit. on p. 46).

Snoek, Jasper, Hugo Larochelle, and Ryan P. Adams (2012). "Practical Bayesian Optimization of Machine Learning Algorithms". In: *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 2*. NIPS'12. Lake Tahoe, Nevada: Curran Associates Inc., 2951–2959 (cit. on p. 111).

Solway, Alec and Matthew M Botvinick (2012). "Goal-directed decision making as probabilistic inference: a computational framework and potential neural correlates." In: *Psychological review* 119.1, p. 120 (cit. on p. 23).

Somani, Adhiraj, Nan Ye, David Hsu, and Wee Sun Lee (2013). "DESPOT: Online POMDP planning with regularization". In: *Advances in neural information processing systems* 26 (cit. on p. 25).

Spering, Miriam (2022). "Eye Movements as a Window into Decision-Making". In: *Annual Review of Vision Science* 8.1. PMID: 35676097, pp. 427–448 (cit. on p. 88).

Spooner, Chris (2009). *25 Classic Fonts That Will Last a Whole Design Career*. https://blog.spoongraphics.co.uk/articles/25-classic-fonts-that-will-last-a-whole-design-career (cit. on p. 108).

Sra, Suvrit and Inderjit S. Dhillon (2006). "Generalized Nonnegative Matrix Approximations with Bregman Divergences". In: *Advances in Neural Information Processing Systems 18*. Ed. by Y. Weiss, B. Schölkopf, and J. C. Platt. Cambridge, MA, USA: MIT Press, pp. 283–290 (cit. on p. 107).

Srinivas, Niranjan, Andreas Krause, Sham M. Kakade, and Matthias W. Seeger (2009). "Gaussian Process Bandits without Regret: An Experimental Design Approach". In: *CoRR* abs/0912.3995. arXiv: 0912.3995 (cit. on p. 114).

Stankiewicz, Brian J., Gordon E. Legge, J. Stephen Mansfield, and Erik J. Schlicht (June 2006). "Lost in virtual space: Studies in human and ideal spatial navigation." In: *Journal of Experimental Psychology: Human Perception and Performance* 32.3, pp. 688–704 (cit. on pp. 27, 68).

Straub, Dominik and Constantin A Rothkopf (2022). "Putting perception into action with inverse optimal control for continuous psychophysics". In: *Elife* 11, e76635 (cit. on p. 5).

Sullivan, Brian T., Leif Johnson, Constantin A. Rothkopf, Dana Ballard, and Mary Hayhoe (Dec. 2012). "The role of uncertainty and reward on eye movements in a virtual driving task". In: *Journal of Vision* 12.13, pp. 19–19 (cit. on pp. 1, 7, 65).

Sun, Xiaoshuai, Yao, Hongxun, and Ji, Rongrong (June 2012). "What are we looking for: Towards statistical modeling of saccadic eye movements and visual saliency". en. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. Providence, RI: IEEE, pp. 1552–1559 (cit. on pp. 8, 46, 48).

Sutton, Richard S (1990). "Integrated architectures for learning, planning, and reacting based on approximating dynamic programming". In: *Machine learning proceedings 1990*. Elsevier, pp. 216–224 (cit. on p. 51).

Sutton, Richard S and Andrew G Barto (2018). *Reinforcement learning: An introduction*. MIT press (cit. on pp. 23, 66).

Sweeney, Deborah F, Thomas J Millar, and Shiwani R Raju (2013). "Tear film stability: a review". In: *Experimental eye research* 117, pp. 28–38 (cit. on p. 29).

Tatler, Benjamin W, Mary M Hayhoe, Michael F Land, and Dana H Ballard (2011). "Eye guidance in natural vision: Reinterpreting salience". In: *Journal of vision* 11.5, pp. 5–5 (cit. on pp. 2, 45, 47, 123).

Tavakoli, Hamed Rezazadegan, Esa Rahtu, and Janne Heikkilä (2013). "Stochastic bottom–up fixation prediction and saccade generation". In: *Image and Vision Computing* 31.9, pp. 686–693 (cit. on p. 63).

Thomas, Tobias, David Hoppe, and Constantin A. Rothkopf (2022). "The neuroeconomics of individual differences in saccadic decisions". In: *bioRxiv*. eprint: https://www.biorxiv.org/content/early/2022/06/04/2022.06.03.494508.full.pdf (cit. on pp. 46, 47, 49, 52, 54, 62, 63).

Thoreau, Henry David (2006). *Walden*. Yale University Press (cit. on p. 45).

Todorov, Emanuel (2004). "Optimality principles in sensorimotor control". In: *Nature neuroscience* 7.9, pp. 907–915 (cit. on p. 6).

Todorov, Emanuel and Michael I Jordan (2002). "Optimal feedback control as a theory of motor coordination". In: *Nature neuroscience* 5.11, pp. 1226–1235 (cit. on p. 5).

Torralba, Antonio, Aude Oliva, Monica S Castelhano, and John M Henderson (2006). "Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search." In: *Psychological review* 113.4, p. 766 (cit. on p. 2).

Train, Kenneth E (2009). *Discrete choice methods with simulation*. Cambridge university press (cit. on p. 54).

Treisman, Anne M and Garry Gelade (1980). "A feature-integration theory of attention". In: *Cognitive psychology* 12.1, pp. 97–136 (cit. on pp. 2, 45).

Treue, Stefan (2003). "Visual attention: the where, what, how and why of saliency". In: *Current opinion in neurobiology* 13.4, pp. 428–432 (cit. on p. 47).

Trommershäuser, Julia, Laurence T. Maloney, and Michael S. Landy (Aug. 2008). "Decision making, movement planning and statistical decision theory". In: *Trends in Cognitive Sciences* 12.8, pp. 291–297 (cit. on pp. 4, 20).

Trommershäuser, Julia, Laurence T Maloney, and Michael S Landy (2003). "Statistical decision theory and the selection of rapid, goal-directed movements". In: *JOSA A* 20.7, pp. 1419–1433 (cit. on p. 20).

Tseng, Po-He, Ran Carmi, Ian G. M. Cameron, Douglas P. Munoz, and Laurent Itti (July 2009). "Quantifying center bias of observers in free viewing of dynamic natural scenes". In: *Journal of Vision* 9.7, pp. 4–4 (cit. on p. 2).

Tsividis, Pedro A, Joao Loula, Jake Burga, et al. (2021). "Human-level reinforcement learning through theory-based modeling, exploration, and planning". In: *arXiv preprint arXiv:2107.12544* (cit. on p. 88).

Tsotsos, John K. (May 2011). *A Computational Perspective on Visual Attention*. The MIT Press (cit. on p. 8).

Turcott, Robert G, Steven B Lowen, Eric Li, et al. (1994). "A nonstationary Poisson point process describes the sequence of action potentials over long time scales in lateral-superior-olive auditory neurons". In: *Biological cybernetics* 70, pp. 209–217 (cit. on p. 31).

Tutt, Ron, Arthur Bradley, Carolyn Begley, and Larry N Thibos (2000). “Optical and visual impact of tear break-up in human eyes”. In: *Investigative ophthalmology & visual science* 41.13, pp. 4117–4123 (cit. on p. 29).

Underwood, Geoffrey, Peter Chapman, Neil Brocklehurst, Jean Underwood, and David Crundall (2003). “Visual attention while driving: sequences of eye fixations made by experienced and novice drivers”. In: *Ergonomics* 46.6, pp. 629–646 (cit. on p. 7).

Vales-Alonso, Javier, Francisco J Parrado-Garcia, Pablo Lopez-Matencio, Juan J Alcaraz, and Francisco J Gonzalez-Castano (2013). “On the optimal random deployment of wireless sensor networks in non-homogeneous scenarios”. In: *Ad Hoc Networks* 11.3, pp. 846–860 (cit. on p. 31).

Vidal, Alice, Salvador Soto-Faraco, and Rubén Moreno-Bote (Sept. 2022). “Balance between breadth and depth in human many-alternative decisions”. In: *eLife* 11. Ed. by Valentin Wyart, Michael J Frank, and Konstantinos Tsetsos, e76985 (cit. on pp. 88, 98).

Wang, Wei, Cheng Chen, Yizhou Wang, et al. (2011). “Simulating human saccadic scanpaths on natural images”. In: *CVPR 2011*. IEEE, pp. 441–448 (cit. on pp. 8, 46, 48, 63).

Wang, Wenguan, Jianbing Shen, Fang Guo, Ming-Ming Cheng, and Ali Borji (June 2018). “Revisiting Video Saliency: A Large-Scale Benchmark and a New Model”. en. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT: IEEE, pp. 4894–4903 (cit. on pp. 46, 47).

Warton, David I and Leah C Shepherd (2010). “Poisson point process models solve the pseudo-absence problem for presence-only data in ecology”. In: *The Annals of Applied Statistics*, pp. 1383–1402 (cit. on p. 30).

Wässle, Heinz (2004). “Parallel processing in the mammalian retina”. In: *Nature Reviews Neuroscience* 5.10, pp. 747–757 (cit. on p. 3).

Watkins, Christopher JCH and Peter Dayan (1992). “Q-learning”. In: *Machine learning* 8, pp. 279–292 (cit. on p. 22).

Welleck, Sean, Jialin Mao, Kyunghyun Cho, and Zheng Zhang (Jan. 2017). “Saliency-based sequential image attention with multiset prediction”. English (US). In: *Advances in Neural Information Processing Systems* 2017-December. 31st Annual Conference on Neural Information Processing Systems, NIPS 2017 ; Conference date: 04-12-2017 Through 09-12-2017, pp. 5174–5184 (cit. on p. 47).

Wertheimer, Max (1912). “Experimentelle studien uber das sehen von bewegung”. In: *Zeitschrift fur psychologie* 61, pp. 161–165 (cit. on p. 2).

Williams, George and the FontForge Project contributors (2000). *FontForge*. `https://fontforge.org/en-US/` (cit. on p. 110).

Wilming, Niklas, Torsten Betz, Tim C. Kietzmann, and Peter König (Sept. 2011). “Measures and Limits of Models of Fixation Selection”. In: *PLOS ONE* 6.9, pp. 1–19 (cit. on p. 58).

Wilson, Simon P and Mark J Costello (2005). “Predicting future discoveries of European marine species by using a non-homogeneous renewal process”. In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54.5, pp. 897–918 (cit. on p. 30).

Wiseman, Richard J and Tamami Nakano (2016). “Blink and you’ll miss it: the role of blinking in the perception of magic tricks”. In: *PeerJ* 4, e1873 (cit. on p. 29).

Wolfe, Jeremy M (1994). “Guided search 2.0 a revised model of visual search”. In: *Psychonomic bulletin & review* 1, pp. 202–238 (cit. on p. 2).

– (1998). “What can 1 million trials tell us about visual search?” In: *Psychological Science* 9.1, pp. 33–39 (cit. on p. 7).

Wu, Zhaohui, Nenggan Zheng, Shaowu Zhang, et al. (Sept. 2016). “Maze learning by a hybrid brain-computer system”. In: *Scientific Reports* 6.1, p. 31746 (cit. on p. 87).

Xia, Chen, Junwei Han, Fei Qi, and Guangming Shi (2019). “Predicting human saccadic scanpaths based on iterative representation learning”. In: *IEEE Transactions on Image Processing* 28.7, pp. 3502–3515 (cit. on pp. 46, 48).

Xu, Juan, Ming Jiang, Shuo Wang, Mohan S. Kankanhalli, and Qi Zhao (Jan. 2014). “Predicting human gaze beyond pixels”. In: *Journal of Vision* 14.1, pp. 28–28 (cit. on p. 57).

Yakovlev, Gleb, John B Rundle, Robert Shcherbakov, and Donald L Turcotte (2005). “Inter-arrival time distribution for the non-homogeneous Poisson process”. In: *arXiv preprint cond-mat/0507657* (cit. on p. 34).

Yang, Scott Cheng-Hsin, Mate Lengyel, and Daniel M Wolpert (2016). “Active sensing in the categorization of visual patterns”. In: *Elife* 5, e12215 (cit. on p. 67).

Yarbus, Alfred L (2013). *Eye movements and vision*. Springer (cit. on p. 65).

– (1967). “Eye movements during perception of complex objects”. In: *Eye movements and vision*, pp. 171–211 (cit. on p. 2).

Yeung, Nick and Christopher Summerfield (2012). “Metacognition in human decision-making: confidence and error monitoring”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 367.1594, pp. 1310–1321 (cit. on p. 5).

York, Jeremy (2008). “Legibility and Large-Scale Digitization”. In: *Hathi Trust Digital Library* . (Cit. on p. 102).

Zanca, Dario, Stefano Melacci, and Marco Gori (2019). “Gravitational laws of focus of attention”. In: *IEEE transactions on pattern analysis and machine intelligence* 42.12, pp. 2983–2995 (cit. on pp. 8, 48, 63).

Zapf, Hermann, William Andrews Clark Memorial Library, and John Dreyfus (1991). *Classical typography in the computer age*. California: U. of California. (cit. on pp. 101, 104).

Zhang, Shen, Xin Liu, Jinjun Tang, Shaowu Cheng, and Yinhai Wang (2019). “Urban spatial structure and travel patterns: Analysis of workday and holiday travel using inhomogeneous Poisson point process models”. In: *Computers, Environment and Urban Systems* 73, pp. 68–84 (cit. on p. 31).

Zhao, Min and Andre G Marquez (2013). “Understanding Humans’ Strategies in Maze Solving”. In: *arXiv preprint arXiv:1307.5713* (cit. on p. 88).

Zhao, Nanxuan, Ying Cao, and Rynson W.H. Lau (2018). “Modeling Fonts in Context: Font Prediction on Web Designs”. In: *Computer Graphics Forum (Proc. Pacific Graphics 2018)* 37 (7) (cit. on p. 105).

Zhong, Yatao, Bicheng Xu, Guang-Tong Zhou, Luke Bornn, and Greg Mori (2018). “Time perception machine: Temporal point processes for the when, where and what of activity prediction”. In: *arXiv preprint arXiv:1808.04063* (cit. on p. 30).

Zhu, Seren, Kaushik J Lakshminarasimhan, Nastaran Arfaei, and Dora E Angelaki (May 2022). “Eye movements reveal spatiotemporal dynamics of visually-informed planning in navigation”. In: *eLife* 11. Ed. by Hang Zhang, Joshua I Gold, and Hugo J Spiers, e73097 (cit. on p. 88).

Zook, Alexander, Eric Fruchter, and Mark O. Riedl (2014). “Automatic playtesting for game parameter tuning via active learning”. In: *Proceedings of the 9th International Conference on the Foundations of Digital Games, FDG 2014, Liberty of the Seas, Caribbean, April 3-7, 2014*. Ed. by Michael Mateas, Tiffany Barnes, and Ian Bogost. Liberty of the Seas, Caribbean: Society for the Advancement of the Science of Digital Games (cit. on p. 105).

List of Figures

List of Tables

Colophon

This thesis was typeset with LaTeX 2_ε. It uses the *Clean Thesis* style developed by Ricardo Langner. The design of the *Clean Thesis* style is inspired by user guide documents from Apple Inc.

Download the *Clean Thesis* style at `http://cleanthesis.der-ric.de/`.

The illustrations at the beginning of each chapter, as well as Figures 1.1 and 2.10, were generated by myself via Midjourney.

Declaration

Erklärungen laut Promotionsordnung

§ 8 Abs. 1 lit. c PromO Ich versichere hiermit, dass die elektronische Version meiner Dissertation mit der schriftlichen Version übereinstimmt.

§ 8 Abs. 1 lit. d PromO Ich versichere hiermit, dass zu einem vorherigen Zeitpunkt noch keine Promotion versucht wurde. In diesem Fall sind nähere Angaben über Zeitpunkt, Hochschule, Dissertationsthema und Ergebnis dieses Versuchs mitzuteilen.

§ 9 Abs. 1 PromO Ich versichere hiermit, dass die vorliegende Dissertation selbstständig und nur unter Verwendung der angegebenen Quellen verfasst wurde.

§ 9 Abs. 2 PromO Die Arbeit hat bisher noch nicht zu Prüfungszwecken gedient.

Ort, Datum	Florian Kadner